NBER WORKING PAPER SERIES

HOW IMPORTANT ARE CULTURAL FRICTIONS FOR INTERNAL MIGRATION? EVIDENCE FROM THE NINETEENTH CENTURY UNITED STATES

Taylor Jaworski
Erik O. Kimbrough
Nicole Saito

Working Paper 33192
http://www.nber.org/papers/w33192

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

The authors would like to thank the Kay Family Foundation for financial support. We thank Allison Ma, Pranit Nanda, Manav Chakravarthy, Emma Kochenderfer, and Anna Pasnau for excellent research assistance. We would also like to thank Brian Cadena, Ann Carlos, Maggie Jones, Cindy Rivas, Jared Rubin, Hagop Shahbazian, and participants in seminars at Caltech, Santa Clara, UC-Irvine, Yale, and at the 2022 meetings of the Association for the Study of Religion, Economics and Culture for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Important Are Cultural Frictions for Internal Migration? Evidence from the Nineteenth Century United States
Taylor Jaworski, Erik O. Kimbrough, and Nicole Saito
NBER Working Paper No. 33192
November 2024
JEL No. N0, N71, N91

ABSTRACT

We propose a new measure of cultural distance based on differences in the composition of first names and church denominations between locations. We use a gravity equation to estimate the elasticity of migration flows with respect to the two components of cultural distance as well as a standard measure of travel costs via the transportation network in the United States between 1850 and 1870. Our findings indicate a modest role for cultural distance relative to travel costs in explaining migration flows. We construct migration costs that reflect the distinct contributions of cultural distance and travel costs, and use an economic geography model of migration to quantify their effects. Travel costs are substantially more important than cultural distance for aggregate welfare. Nevertheless, we provide evidence that the components of cultural distance play a role in shaping of how many people move and their final destinations.

Taylor Jaworski
Department of Economics
University of Colorado Boulder
256 UCB
Boulder, CO 80309
and NBER
Taylor.Jaworski@Colorado.edu

Erik O. Kimbrough
Smith Institute for Political Economy
and Philosophy
Chapman University
One University Drive
Orange, CA 92866
ekimbrou@chapman.edu

Nicole Saito
Northwestern University
nicolesaito2028@u.northwestern.edu

1 Introduction

In the middle of the 19th century, domestic mobility in the United States reached its historic peak (Thernstrom and Knights, 1970; Hall and Ruggles, 2004; Rosenbloom and Sundstrom, 2004). This period of internal migration was an important precursor to the era of mass immigration and coincided with US westward expansion. Following the expulsion of native peoples, white native-born Americans primarily drove population inflows to the newly incorporated states of the lower South and Midwest. The extent of mobility they enjoyed at this time naturally suggests an economic interpretation of their motives for relocation; namely, individuals weighed the benefits and costs of migrating and chose their location given their preferences and constraints (e.g. Ravenstein, 1885, 1889).

The benefits of migration depend on comparisons between an individual's origin and potential destinations. Economists distinguish between "push" and "pull" factors as the basis of such comparisons. Push factors impel individuals to evaluate exit options, and pull factors attract prospective migrants to particular destinations. For example, migrants may be pushed out by a mismatch between their preferences for and a region's actual level of taxation and public goods provision (Tiebout, 1956; Rhode and Strumpf, 2003; Cebula, 2009; Clark and Hunter, 1992; Egger and Radulescu, 2009). Alternatively, migrants may be pulled to a new region that offers amenities unavailable at their origin. Similarly, migration may depend on differences in employment conditions, land prices, or wages between the origin and destination (Harris and Todaro, 1970; Fields, 1976; Herzog Jr and Schlottmann, 1984; Stark and Bloom, 1985).

In a standard economic model, given the costs and benefits of migration, optimizing behavior by individuals should produce aggregate outflows and inflows of migrants and generate an equilibrium spatial distribution of population. In the long run, migration should arbitrage away the differentials that motivated it in the first place. Yet, sizable economic disparities persist across regions (Partridge and Rickman, 1995, 1997), and contemporary evidence suggests that people move less than would be predicted by pecuniary factors alone (Putnam, 2000; Molloy, Smith and Wozniak, 2011; Cooke, 2013; Wilson, 2021). This seems to have also been the case in the mid-19th century; for example, Bleakley and Rhode (2022) find evidence of differences in wages between free-versus slave-states and argue that such differences persisted because the slave system "repulsed potential settlers and migrants" (p. 25). This suggests the existence of frictions due to cultural mismatch between origin and destination locations.

Cultural mismatch may shape a migrant's decision through effects on both consumption and production. On the consumption side, there are social and psychological benefits of being in a familiar cultural environment. Identity matters, and people value social networks and living amongst those who are culturally familiar (Bailey, Cao, Kuchler, Stroebel and Wong, 2018). While kinship and direct social ties have been shown to be important in supporting networks of reciprocity, mutual insurance, etc. (Munshi, 2014), cultural familiarity may also ease integration into such networks among strangers. Moreover, while cultural familiarity can yield benefits, cultural differences may diminish social capital and trust in civil society, generating social and economic costs (Putnam, 2000). On the production side, cultural similarity can facilitate exchange by reducing transactions costs. Alternatively, cultural mismatch may increase communication costs, introducing frictions into attempts to reap gains from specialization and exchange (see e.g. Deming, 2017, for a model of how communication costs reduce efficient team production).

Cultural "fit" provides consumption value and facilitates exchange; cultural frictions impede both. Thus, all else equal, if a migrant anticipates these consequences, they would prefer to live in a place where they understand the norms and practices and share the beliefs and values of their neighbors.[1] These arguments all suggest that it is sensible to think of cultural differences as part of a measure between capturing the "distance" between locations, and thus that the effects of cultural distance on migration decisions can be integrated into standard models of migration.

In this paper, we estimate a gravity model of migration flows in which agents' decisions depend both on economic conditions at their origin and prospective destinations and on bilateral costs. We capture one component of bilateral migration costs with the travel costs associated with using the existing transportation network (e.g., railroads, waterways, and footpaths). In addition, we introduce a second component of migration costs based on two new measures of cultural distance that reflect the degree of cultural

[1] Of course, this overlooks that some heterogeneity (e.g. in preferences or abilities) is itself a spur to specialization and exchange. The relative importance of these offsetting effects is unclear *ex ante*.

mismatch between an origin and destination. Specifically, we represent the culture of each location in the United States in two ways. Our first approach uses the distribution of given names following research in cultural geography (Zelinsky, 1970, 1992), history (Stewart, 1948; Smith, 1985, 1994; Main, 1996), and more recently in economic history (Abramitzky, Boustan and Eriksson, 2020a; Knudsen, 2019; Bazzi, Fiszbein and Gebresilasse, 2020). Second, using the distribution of religious institutions (Baltzell, 1979; Fischer, 1989; Phillips, 1999). We then calculate the cultural distance along each of these dimensions between each location pair using a dissimilarity-measure in the spirit of Hoberg and Phillips (2016).

To evaluate the impact of both components of migration costs we first estimate a gravity equation that yields the elasticity of migration flows with respect to each component. The results confirm a role of cultural distance in explaining migration flows that survives including alternative measures of pecuniary travel costs as well as origin and destination fixed effects that capture location-specific push and pull factors. In particular, between 1850 and 1870, we estimate a semi-elasticity of migration flows with respect to first name dissimilarity of -0.23 and church dissimilarity of -0.84. These semi-elasticities decrease to -0.18 and -0.66, respectively, but remain statistically significant and qualitatively important after controlling for travel costs measured via the transportation network. This raises the question of how important cultural distance – in our case, first name and church dissimilarity – is, relative to standard economic factors such as transportation costs for understanding aggregate migration flows.

We then use a simple economic geography of migration to account for the potential general equilibrium responses to counterfactual changes in migration costs and quantify their implications for aggregate welfare and the spatial distribution of population. Though cultural distance matters, we find that travel costs are considerably more important than cultural distance for aggregate welfare: counterfactually eliminating the travel costs as a component of migration costs increases welfare by 44.3 percent in our baseline specification (between 34 and 54 percent in alternatives), while removing cultural distance increases welfare by 5.5 percent (between 2.5 and 9.3 percent in alternative specifications). However, travel costs and cultural distance have a distinctive impact on the spatial distribution of the population in 1870: in the absence of travel costs Appalachia and the west would have populated more rapidly, while the absence of cultural distance would have pulled some people farther north and others farther south (and east).

That real people responded to the forces reflected in our results is consistent with histories of the period, with migrants expressing desire to escape poor economic conditions at home or capitalize on opportunities elsewhere. Westward migration was driven by the desire to take advantage of perceived economic opportunities. For example, one of Thomas Jefferson's grandchildren, Francis Eppes, lamented the decline of their native Virginia: "I see no ties which should bind any descendants of our grandfather to this state [...] the soil is exhausted, the staple reduced almost to the prime cost of the materials — a level to which it is fast progressing. What inducement is there to remain!" (Fischer and Kelly, 2000, p. 204). Another Virginian, John Finley, wrote a poem extolling the prospects of moving to Indiana: "Blest Indiana! in thy soil / Are found the sure rewards of toil, / Where honest poverty and worth / May make a Paradise on earth" (Finley, 1866).

In addition, migrants were constrained by travel costs. Although by 1850 the era of steamboats was well underway and the railroad network was beginning to expand, access depended on proximity to navigable waterways and railroads. Migrants' other option was overland transit via stagecoach on "scarcely maintained and often impassable roads and trails" (Sellers, 1991, p. 5). Stagecoach transit was expensive compared to rail and steamboat travel (Larson, 2009; Fogel, 1962), so much so that taking one's belongings "more than thirty or forty miles cost more than they were worth" (Sellers, 1991, p. 5). This rendered some moves far more affordable than others.

There is good reason to believe that cultural frictions were important as well (de Tocqueville, 1835; Zelinsky, 1970, 1992; Fischer, 1989; Woodard, 2011). Founding populations from different regions of Britain established colonies where they reproduced unique cultural patterns imported from their homelands. Visitors to early America regularly noted the cultural differences between the strictly religious townspeople of New England, the faux-aristocratic planters of the Tidewater region, the rough-and-tumble settlers of the Appalachians, and the moderate Quakers of Pennsylvania (Fischer, 1989). Indeed, colonists seem to have been well aware of these differences, on the basis of which they often expressed interregional prejudices (Phillips, 1999). These sentiments found expression even under existential threat. When fighting for national unity during the Revolutionary War, George Washington himself succumbed to regional chauvinism, disparaging New Englanders as "an exceeding[ly] dirty & nasty people" (Taylor, 2021, pp. 19).

Mutual suspicion persisted beyond the country's founding. Representatives regularly met in the new capital at Washington, DC, where they disagreed vehemently about matters of policy, such as taxation (Sellers, 1991, p. 73). Yet their disagreements extended beyond the business of the legislature into matters of daily life. "Most of the residents came from the South, and southern pastimes — horse racing, gambling, cockfighting, and dueling — prevailed to the dismay of pious New Englanders" (Taylor, 2021, pp. 3). Cultural differences like these preserved fault lines along which interregional tensions simmered, fueling political (and eventually military) conflict (Faust, 1989; Eli, Salisbury and Shertzer, 2018).

Culture competed with economic factors to shape individual migration decisions. Fischer and Kelly (2000) cite a French observer of a Virginia planter named James Meade, who in 1796 noted "[Meade] is even so disgusted with a culture wherein it is necessary to make use of slaves that he is tempted to sell his possessions in Virginia and remove to New England." Yet, they report that Meade was ultimately unwilling "to take so radical a step" and instead settled in Kentucky, attracted by low land prices (pp. 159-162). Meade's story illustrates how culture might factor into location decisions. Although Meade detested slavery, cultural differences between Virginia and New England attracted him instead to Kentucky, which was more familiar.

When migrants did cross cultural boundaries, it was often because they already had adopted, or intended to adopt, new cultural practices. For example, Fischer and Kelly document the movement of Virginian, Robert Carter, and describe him as motivated by a mixture of the "material and moral" (pp. 198-199). Struggling to support the expenses of his plantation and morally troubled by slavery, he enacted the largest emancipation prior to 1860 — freeing his 500 slaves — and moved to Baltimore. There, he joined a religious community that supported abolition. In the absence of such commitment, however, migrants were less likely to cross cultural boundaries.

Background on Measuring Culture

As economists have embraced culture as a factor explaining economic outcomes, they have increasingly incorporated it into theory. Nevertheless, empirical tests for the relevance of culture are limited by the difficulty of measuring cultural differences. Social scientists have proposed several approaches. Many contemporary analyses tend to rely on survey measures of beliefs and practices, such as those recorded in the World Values Survey. However, surveys typically do not extend back beyond several decades. As a result, historical empirical studies of culture often employ more indirect approaches. One approach is to assume that culture is coterminous with observable markers of group identity, such as nationality, ethnicity, language, race, religion, etc. This allows researchers to link contemporary populations to ethnographic records of the cultural practices of their ancestors (e.g. Bahrami-Rad, Becker and Henrich, 2021). However, this is a strong assumption since observable group markers need not coincide with cultural differences. Indeed, Desmet, Ortuño-Ortín and Wacziarg (2017) use recent survey data to show that in the contemporary United States, there is more cultural variation within ethnic groups than between ethnic groups.

As such, there is a need for alternative measures of culture. For this paper, we adopt two approaches. First, we draw inspiration from cultural geography. In particular, we follow Zelinsky (1970), who made a seminal contribution to the measurement of culture by linking onomastics — the study of the history and origins of proper names — to cultural geography. He argues that the ideal measure of culture should fulfill six criteria: sensitivity, ubiquity, durability, simplicity, purity, and accessibility. He concludes that first names come closer to fulfilling these criteria than any other observable characteristic (Zelinsky, 1970, p. 744).

Unlike surnames, first names are actively chosen by an individual's parents, guardians, or caretakers. Therefore, they are 'sensitive' indicators of the expressed cultural preferences of one's family network. First names are 'ubiquitous' and 'accessible.' Every person is given a first name at birth, and names are regularly recorded in birth and other official records encompassing the entire population. First names are also 'pure' and 'durable' in the sense that they generally remain stable across an individual's lifetime.[2]

[2]The exceptions to this tend to prove the rule, as changes to one's forename typically result from a decision to adopt a new cultural identity, e.g. adopting a religious name post-conversion.

Once they are given, they do not change as a result of economic or political shocks, and thus do not generate endogeneity problems. In addition, the informal rules governing naming practices tend to be conceptually 'simple', as is the intuition which underlies their relevance to culture: first names reflect the beliefs, values, and practices of the community in which one is reared. For example, consider the high prevalence of the girls' names "Chastity" and "Purity" in Puritan settlements in the late 18th century. Zelinsky (1970) validates his approach by showing that county-level name distributions in 16 US counties reveal both persistent cultural differences and evidence of cultural convergence when sampled in 1790 and 1968.

To our knowledge, Zelinsky's contributions have, for the most part, only recently begun to influence other disciplines. One early application comes from Fischer (1989), who includes naming practices among his list of folkways that differentiate regional cultures in colonial America. More recently, in economics, Jurajda and Kovač (2021) and Assouad (2020) use naming decisions as measures of nationalism in Croatia and Turkey, respectively. Jurajda and Kovač (2021) strikingly show that having the same first name as a historical political leader is associated with a higher likelihood of volunteering (and dying) in the Croatian War of Independence. Research like this provides supporting evidence for one of Zelinsky's key insights: first names reflect the political, cultural, and social values of their time period and place.

Most relevant for this paper are studies conducted by Abramitzky, Boustan and Eriksson (2020a), Knudsen (2019), and Bazzi, Fiszbein and Gebresilasse (2020) on culture and migration. These papers share a common time period, namely the eras of mass migration and westward expansion in the United States and use first names to measure cultural attributes. Directly linking onomastics and culture, Abramitzky, Boustan and Eriksson (2020a) use convergence of naming practices to measure assimilation among first-generation immigrants in the United States. They argue that during the Age of Mass Migration, immigrants faced a trade-off when choosing a first name for their child: they may either give their child a 'foreign' name, i.e., one indicative of their original cultural values, or give their child an 'American' name in hopes that they will be accepted by their new host culture. In choosing an American name, immigrants make a symbolic choice to assimilate. This analysis reinforces the view that naming practices reflect cultural commitments and provides additional evidence connecting culture and migration decisions.

Taking a different approach, Knudsen (2019) and Bazzi, Fiszbein and Gebresilasse (2020) use first name distinctiveness as a measure of 'individualism' and show that this measure is positively related to individuals' migration propensities. Knudsen (2019) shows that people with unique first names are overrepresented among Scandinavian immigrants to the US, while Bazzi, Fiszbein and Gebresilasse (2020) show that those with unique first names were more likely to move to western frontier settlements. Both results imply that migration may be driven by personal attributes, such as individualism. Bazzi, Fiszbein and Gebresilasse (2020) particularly emphasize that individuals, when choosing whether and where to migrate, consider the "fit" between their own personalities and those of the destination population, and this coheres with psychological evidence that people have higher self-esteem when surrounded by others with more similar personalities (Bleidorn, Schönbrodt, Gebauer, Rentfrow, Potter and Gosling, 2016).

Knudsen (2019), Bazzi, Fiszbein and Gebresilasse (2020) and Abramitzky, Boustan and Eriksson (2020a) assign individuals to particular cultural groups based on their first names. These approaches depend on researcher-defined mappings between first names and cultures: i.e., what constitutes an American versus foreign name, or how unique a name must be in order to qualify as individualistic. Our approach sidesteps the need to link specific names to particular cultures. We instead create a bilateral measure of cultural distance between location pairs based on the dissimilarity of their first name distributions. This is our first measure of cultural distance.

Our second approach to capturing cultural distance is to use measures of the distribution of religious institutions in each location. Fischer (1989) also emphasizes religious practices as a key folkway distinguishing regional cultures in the early US. According to him, the early Puritan settlers of New England were distinct from the primarily Anglican settlers of the Tidewater region, and both were distinct from the Quakers of Pennsylvania and the Presbyterians who settled the Appalachian region. Similarly, Phillips (1999) highlights religious distrust between Catholics and Protestants as a key factor in American political conflict in the lead-up to the Civil War. Given the salience of religious differences for Americans during this period and the potential for tensions experienced by a migrant to a religiously incompatible destination, it is plausible that religious institutional dissimilarity constituted another friction to migration. As such, we use data on the mix of religious institutions present in each county as our second measure of culture. As above, we don't use religious institutions to assign a specific religious label to each county. Instead, we create a bilateral measure of cultural distance between location pairs based on the dissimilarity of their religious institutions.

In our empirical analysis, migration flows depend on these two measures of cultural distance between origin and destination as well as more standard travel costs, and origin and destination fixed effects. This approach allows us to remain agnostic regarding to which cultural group any particular region ought to be assigned and instead allows us to test whether cultural difference, per se, shapes migration flows.

2 Data

We study internal migration in the United States between 1850 and 1870. At that time, settlement was expanding rapidly westward, with migrants emanating from eastern regions that historians and cultural geographers have identified as having distinctive cultures. We construct measures of migration, cultural distance and travel costs to estimate the extent to which migration flows were sensitive to cultural "fit" between migrants' origins and destinations, controlling for the pecuniary costs of migration.

In the main empirical analysis, the dependent variable is a measure of migration constructed from the decennial censuses and information linking individuals over time. The key variables of interest are our measures of cultural distance. One is based on the distribution of first names which we construct from the complete count of the US Census in 1850. The second is the distribution of churches, by denomination, at the county level, as tabulated in the 1850 Census. We control for travel costs by estimating the least cost path between location pairs using detailed information about the available transportation network at the time. In the rest of this section, we explain the construction of these variables in more detail and provide an illustrative example to build intuition.

2.1 Geographic Unit

The goal of our empirical analysis is to quantify the determinants of migration flows between locations within the United States in the middle of the 19th century. First, we start with the sample of US counties in each year, restricted to territory that was incorporated within states since 1820 (including the areas that became Maine and West Virginia). Second, due to land reapportionments, we aggregate counties that share more than 0.5 percent of their land mass across the sample period to ensure that our analysis are conducted on constant geographic units. This allows us to exploit county-level data without fractionalizing county populations due to land reapportionments.[3] The key consequence of this is that we cannot simply rely on previous county-level travel cost estimates in our analysis; we discuss how we compute travel costs below. Our procedure leaves us with 927 locations ("supercounties") with complete data for a total of 824,464 migration pathways.

2.2 Internal Migration Flows

We construct measures of internal migration between 1850 and 1870. Specifically, we consider three periods: 1850 to 1860, 1860 to 1870, and 1850 to 1870. For our baseline migration sample, we use the complete count US censuses for 1850, 1860, and 1870 (Ruggles, Nelson, Sobek, Fitch, Goeken, Hacker, Robers and Robert, 2024) together with links for our three time periods from the Census Tree (Price, Buckles, Van Leeuwen and Riley, 2021; Buckles, Haws, Price and Wilbert, 2023). We restrict the analysis to whites since enslaved blacks were not enumerated with identifying information in the 1850 and 1860 censuses. From each census, we know each individual's county of residence and thus can identify those who have moved as those who reside in a county different from the one in which they resided during the previous census. We then aggregate this information to the supercounty level to construct bilateral migration flows. Using links from the Census Tree allows us to include women in our baseline samples. To assess the robustness of our findings to alternative ways of linking individuals across Census years, we also construct migration using links from Multigenerational Longitudinal Panel (Ruggles, Fitch, Goeken, Hacker, Helgertz, Roberts, Sobek, Thompson, Warren and Wellington, 2020) and the Census Linking Project (Abramitzky, Boustan, Eriksson, Pérez and Rashid, 2020b).

2.3 Measuring Cultural Distance

We construct a measure of cultural distance between each location pair from the complete count data from the 1850 Census so that our measure captures cultural dis-

[3]Section A.1 of the Appendix provides additional details of the procedure used to construct the constant geographic boundaries used in this paper.

tance prior to the migration flows that we observe. Our measure of cultural distance uses two components, which can be generalized to more depending on the availability of suitable data.

First, starting with the complete count census in 1850 (Ruggles, Nelson, Sobek, Fitch, Goeken, Hacker, Robers and Robert, 2024), we extract the first name of every enumerated person. We then construct first name distributions as one of the cultural components of migration costs, drawing on previous work that hypothesizes a link between naming practices and cultural distinctiveness (Stewart, 1948; Zelinsky, 1970; Smith, 1985; Fischer, 1989; Zelinsky, 1992; Smith, 1994; Main, 1996). In particular, we interpret dissimilarity in first name distributions as indicative of cultural distance.

We restrict attention to whites and focus on the native-born population since these individuals are arguably more representative of the longstanding distinctive regional cultural groups in the United States.[4] We use a reference set of names into which we classify the digitized Census records. Enumerator errors or errors in OCR transcription may result in misleading name distributions if we simply include the universe of all digitized names in our sample of given names. For example, some of the digitized "names" are not recognizable as names (e.g. "Aaa"). To address this concern, each enumerated name is matched to a list constructed by Jaworski, Kimbrough, Nanda and Chakravarthy (2020) that includes more than 50,000 forenames that either appear in Social Security birth records at least five times between 1880 and 2010 or appear at least 40 times in the 1850 complete count data. After making adjustments for transcription errors and removing honorifics such as "Rev." and "Capt.", we match over 95 percent of individuals in the complete count data to a forename in this list, and we compute a location-level name distribution as the number of occurrences of each forename in each county. Given our sample and geographic restrictions, our first name distributions are based on just over 13 million people.

We transform the first name data from counts to shares, and then compute a measure of the distance between the first name distributions of each location pair. Let $s_i(n)$ denote the share of first name n in location i so that the distance between i and j is given

[4] Including foreign born individuals does not substantially alter our results since major foreign migrations to the US mostly occur after 1850, but this may also reflect the fact that receptiveness to foreign migrants is cultural to some extent.

by[5]:

$$\text{Name Dissimilarity}_{ij} = 1 - \frac{\sum_{n=1}^{N} s_i(n) \times s_j(n)}{\sqrt{\sum_{n=1}^{N} s_i(n)^2} \times \sqrt{\sum_{n=1}^{N} s_j(n)^2}} \tag{1}$$

Drawing on information from Manson, Schroeder, Van Riper, Kugler and Ruggles (2021), we construct a second component of cultural distance based on the difference in the distribution of church denominations across location pairs. From the 1850 Census, we have a count of the number of churches from each of 23 denominations in each county. The denominations included are Baptist, Christian, Congregational-Orthodox, Congregational, Dutch Reformed, Episcopal, Free Churches, Friends or Quaker, German Reformed, Jewish, Lutheran, Mennonite, Methodist, Moravian, Presbyterian, Protestant, Roman Catholic, Swedenborgian, Tunker, Union, Unitarian, Universalist, and a category for other minor sects. As noted by Finke and Stark (2005), while counts of religious institutions are available from 1850 in the Census, membership counts are not available until 1880. As above, we aggregate to supercounties, convert from counts to shares for church denominations in each location, and calculate Church Dissimilarity$_{ij}$ based on equation (1).[6]

The raw correlation coefficient between name and church dissimilarity is 0.0894. Figure C1 illustrates the positive relationship between these two measures of cultural distance after controlling for a "stayer" indicator and origin and destination fixed effects. As we show below, both measures of cultural distance remain significant predictors of reduced migration flows.

2.4 Calculating Travel Costs

Our first measure of travel costs is the physical distance between location centroids, calculated using the Haversine formula. This measure can be thought of as capturing the shortest possible wagon road (or footpath) between locations. While it has the virtue of simplicity, it ignores available information about extant transport networks

[5]Cosine similarity is often used to measure similarity of textual data; see e.g. Hoberg and Phillips (2016) for an economic application. When applied to vectors of unit length, $2 - 2cos\theta$ is equal to the Euclidean distance between vectors.

[6]As an alternative to the functional form in equation (1), we also estimate our main specifications and conduct counterfactuals using the Jensen-Shannon divergence between the name share and church share vectors.

that would have made certain routes easier to travel. Thus, our preferred measure of travel costs explicitly accounts for other existing transport networks during our sample period. Specifically, we augment our footpath network with railroad and waterway transportation networks created by Donaldson and Hornbeck (2016) to produce estimates of the cost-minimizing travel path between each supercounty pair. The waterway network is time-invariant and consists of inland rivers, sea and lake routes, canals, and harbor points. The railroad network is time-varying and was expanding during our sample period. In our baseline specifications, we calculate travel costs based on the 1850 transportation network for each time period. As robustness, we also construct travel costs using the 1860 tranpsortation network and passenger costs derived from Jaremski (2012).

We assume that migrants enter the rail and water networks via the wagon road that minimizes the distance from the location centroid to the transportation network. Thus, every location is assigned an entry point into each transportation network. We assume that migrants are able to switch between transportation networks when it is cost-efficient, so we also create transshipment paths between transportation networks. Whenever two nodes on separate networks passed sufficiently closely to one another, we allow transshipment to link the networks between those nodes.

Following Fogel (1962) and Donaldson and Hornbeck (2016), we assign a per-mile cost to each mode of transportation. Each wagon road is assigned a cost of \$0.231, railways are assigned a cost of \$0.0062, and waterways are assigned a cost of \$0.0048.[7] We also include a fixed \$0.5 transshipment cost associated with each change in transport mode. For every pair of locations, we then calculate the minimum cost path between their centroids using the Dijkstra algorithm (Dijkstra, 1959). The cost of traversing this path serves as our preferred measure of travel costs, travel_{ij}. See Appendix B.2 for complete details on the construction of this measure.

2.5 An Example

For the case of Philadelphia, the locations receiving the largest migration flows are concentrated in and around Pennsylvania. There is substantial overlap with across the

[7]Using physical distance for footpaths arguably underestimates the costs of on-foot travel because it ignores topography. Following Donaldson and Hornbeck (2016), our wagon road costs include an upward adjustment.

different components of migration costs and migration flows with some notable exceptions. For Philadelphia, the cultural dissimilarity measures based on first names and the denomination of churches are lowest in a belt moving east to west and more dissimilarity is present moving farther south or to the far northeast (in the case of churches). Travel costs are lowest along the coasts and highest moving farther west, which reflects the lower relative cost of water transport and the less extensive railroad network.

Figure 1 shows migration outflows, name dissimilarity, church dissimilarity and travel costs from Philadelphia County (in Pennsylvania) to all other locations. According to Fischer (1989), Philadelphia was the principal city of the Quakers – an important regional cultural group – and was also a major economic hub and politically important. Panel A shows migration flows from Philadelphia to all other supercounties between 1850 and 1870, with darker colors indicating larger flows. Panels B and C show the component of cultural distance (i.e., name and church dissimilarity) in 1850, with darker colors indicating lower dissimilarity. Panel D shows travel costs from Philadelphia using the 1850 transport network, with darker colors indicating less costly travel.

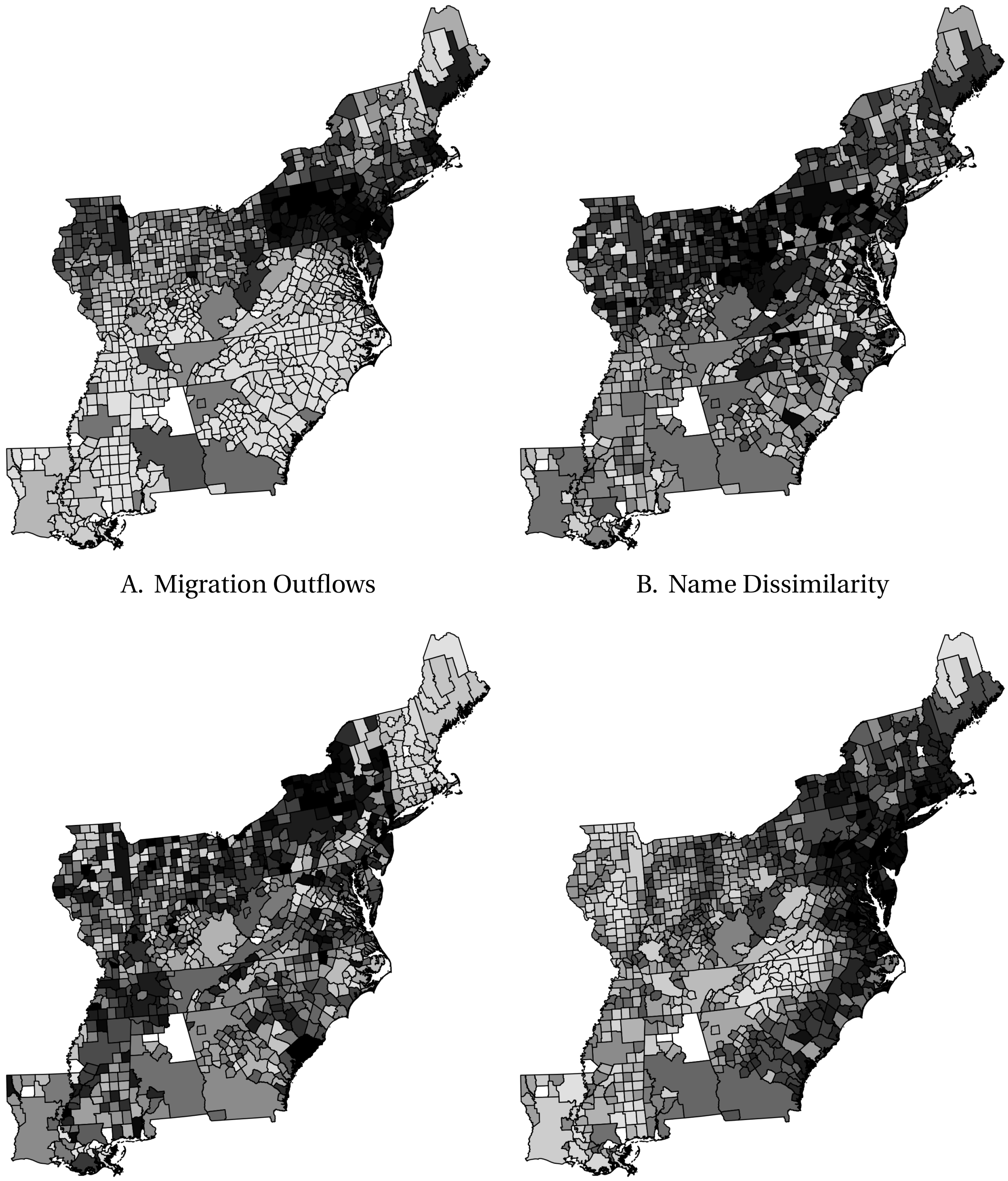
Figure 2 provides intuition for our empirical analysis. We show the binscatter (Cattaneo, Crump, Farrell and Feng, 2023) relationship between migration out-flows and components of cultural distance for Philadelphia. Panel A shows the relationship between out-flows and name dissimilarity (controlling for church dissimilarity and travel costs) and Panel B shows the relationship between out-flows and church dissimilarity (controlling for name dissimilarity and travel costs). For the case of Philadelphia, the negative relationship in each panel highlights that migrants tend to move to culturally similar locations, consistent with the view that cultural distance is an important friction in the migration decision.

The goal of our empirical analysis is to generalize this example by first estimating the strength of the relationship between migration flows and migration costs across all location pairs, which we will then use to quantify the welfare implications of the aggregate and individual components of migration costs.

3 Theory

In this section we present a standard theory of economic geography with costly migration and trade drawing closely on Eaton and Kortum (2002). The motivation for the

Figure 1: Geography of Migration and Cost Components for Philadelphia

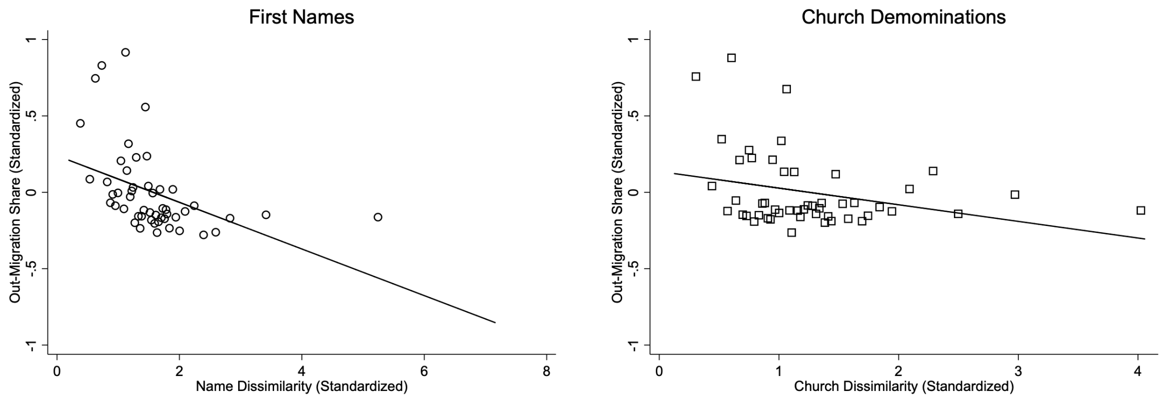


A. Migration Outflows — B. Name Dissimilarity

C. Church Dissimilarity — D. Travel Cost

Notes: This figure shows migration outflows (Panel A), name dissimilarity (Panel B), and church dissimilarity (Panel C), and travel cost via the transportation network (Panel D) between Philadelphia County and all other sample counties. In Panel A, darker coloring indicates higher out-migration from Philadelphia to other locations. In the remaining panels, darker coloring indicates lower dissimilarity (panels B and C) or lower travel costs (Panel D).

Figure 2: Relationship between Migration and Cultural Components for Philadelphia



A. First Names B. Church Denominations

Notes: This figure shows the relationship between migration outflows and name dissimilarity (Panel A) or church dissimilarity (Panel B). Each variable is standardized by subtracting the mean and dividing by the standard deviation.

theoretical framework is twofold. First, the theory provides a guide for the empirical analysis, in particular, the estimation of the parameters of the migration cost function (i.e., the elasticities of migration flows with respect to bilateral travel costs and cultural distance) and the elasticity of migration flow with respect to overall migration costs. Second, we calibrate the model and consider counterfactual scenarios with alternative migration costs to understand the role of cultural distance versus travel costs in shaping the spatial distribution of population and welfare while taking into account general equilibrium interactions.

3.1 Preferences

Consumers, indexed by ϕ, start in location i and move to location j. Each consumer derives utility from a composite tradeable consumption good ($C_j(\phi)$) and a nontradeable housing good ($H_j(\phi)$), and they are subject to bilateral moving costs (μ_{ij}). Consumers also have idiosyncratic tastes ($b_j(\phi)$) that are location-specific. Consumers maximize utility that takes the following Cobb-Douglas form:

$$U_{ij}(\phi) = \frac{b_j(\phi)}{\mu_{ij}} \left(\frac{C_j(\phi)}{\alpha}\right)^{\alpha} \left(\frac{H_j(\phi)}{1-\alpha}\right)^{1-\alpha}, \alpha \in (0,1) \tag{2}$$

subject to a budget constraint, $W_j \geq P_j C_j(\phi) + r_j H_j(\phi)$, where W_j are wages, P_j is the ideal price, and r_j is the price of land.

The index for the composite consumption good is a CES aggregate:

$$C_j = \left[\int_0^1 c_j(\ell)^{\frac{\sigma-1}{\sigma}} d\ell \right]^{\frac{\sigma}{\sigma-1}}, \sigma > 1$$

where $c_j(\ell)$ is consumption of variety ℓ in location j and σ is the elasticity of substitution between goods. The corresponding ideal price index for the consumption good is given by: $P_j = \left[\int_0^1 p_j(\ell)^{1-\sigma-1} d\ell \right]^{\frac{1}{1-\sigma}}$, where $p_j(\ell)$ is price of variety ℓ in location j.

Finally, each individual draws an idiosyncratic taste shock, $b_j(\phi)$, from a Fréchet distribution, $G_j(b) = e^{-B_j b^{-\epsilon}}$, where $B_j > 0$ is a scale parameter such that higher values indicate more valuable amenities, and $\epsilon > 1$ is a shape parameter such that higher values correspond to less dispersion.

3.2 Location Choices

Solving for the optimal choices of consumption and land, and substituting into equation (2) gives the following indirect utility function:

$$U_{ij}(\phi) = \frac{b_j(\phi)}{\mu_{ij}} \frac{W_j}{P_j^{\alpha} r_j^{1-\alpha}}$$

Then, exploiting the properties of the Fréchet distribution, the probability that an individual moves from i to j is:

$$\pi_{ij} = \frac{B_j(\mu_{ij} P_j^{\alpha} r_j^{1-\alpha})^{-\epsilon} (W_j)^{\epsilon}}{\sum_{k=1}^{J} B_k(\mu_{ik} P_k^{\alpha} r_k^{1-\alpha})^{-\epsilon} (W_k)^{\epsilon}} \tag{3}$$

In the above expression, the numerator gives the utility of moving to location j adjusted for moving costs, and the denominator is origin i's labor market access. From an initial allocation of labor (L_{i0}) at location i, we can write migration flows between i and j as $L_{ij} = \pi_{ij} L_{i0}$. Combining this expression with equation (3) gives the gravity equation for migration flows between i and j:

$$L_{ij} = (LMA_i)^{-1} L_{i0} B_j \left(\frac{W_j}{P_j^{\alpha} r_j^{1-\alpha}} \right)^{\epsilon} (\mu_{ij})^{-\epsilon} \tag{4}$$

where $LMA_i = \sum_{k=1}^{J} (\mu_{ik})^{-\epsilon} B_k (W_k / P_k^{\alpha} r_k^{1-\alpha})^{\epsilon}$ is labor market access at the origin, as in Morten and Oliveira (2018). This equation captures the idea that migration flows between i and j are higher if people in origin i have better labor market access, origin i has more people, utility is higher in destination j, or migration costs are lower between i and j. This expression is the starting point for our empirical analysis in Section 4.

The total population in location j is equal to the sum of migration flows across all origin locations:

$$L_j = \sum_{i=1} L_{ij} = \sum_{i=1} \lambda_{ij} L_{i0} \tag{5}$$

Finally, before the realization of $b_j(\phi)$ an individual starting in location i and moving to location j has expected utility,

$$\bar{U}_i = \mathrm{E}[U_{ij} | \text{move to } j] = \gamma \left[\sum_{k=1} B_k (\kappa_{ik} P_k^{\alpha} r_k^{1-\alpha})^{-\epsilon} (W_k)^{\epsilon} \right]^{1/\epsilon} \tag{6}$$

with $\gamma = \Gamma(\frac{\epsilon-1}{\epsilon})$, Γ is the gamma function, and ϵ must be a finite value greater than one.

3.3 Production and Trade

On the production side, variety ℓ is produced in location i using the following technology:

$$q_i(\ell) = z_i(\ell) n_i(\ell)$$

where $n_i(\ell)$ is the number of people employed in location i to produce variety ℓ.[8] The optimal employment level can be derived from a representative firm's cost minimization

[8] This production technology abstracts away from capital and the role of capital markets.

problem:

$$\min_{n_i(\ell)} W_i n_i(\ell) \text{ subject to } z_i(\ell) n_i(\ell) \geq q_i(\ell)$$

The optimal cost for a unit of variety ℓ produced in location i is equal to:

$$c_i(\ell) = \frac{W_i}{z_i(\ell)}$$

Following Eaton and Kortum (2002), productivity shocks ($z_i(\ell)$) are independently and identically drawn from a Fréchet distribution with scale parameter (A_i) and shape parameter (θ) given by: $F_i(z) = e^{-A_i z^{-\theta}}$. The A_i term reflects the average productivity in location i and θ governs the dispersion of productivity shocks.

Under the assumption that output markets are perfectly competitive and bilateral trade costs between location i and j take the iceberg form ($\tau_{ij} \geq 1$), the price of variety ℓ produced in location i and consumed in location j is:

$$p_{ij}(\ell) = \tau_{ij} \frac{W_i}{z_i(\ell)}$$

Combining the fact that consumers source each variety from the supplier with the lowest cost with the properties of the Fréchet distribution governing productivity shocks, we can write expenditure shares as:

$$\lambda_{ij} = \frac{A_i(\tau_{ij} W_i)^{-\theta}}{\sum_{k=1}^{J} A_k(\tau_{ik} W_k)^{-\theta}} \tag{7}$$

and we can then write the gravity equation for bilateral trade flows as:

$$X_{ij} = A_j W_j^{-\theta} (\tau_{ij})^{-\theta} (CMA_j)^{-1} X_i$$

where $CMA_j = \sum_{k=1}^{J} \tau_{ik}^{-\theta} A_k W_k^{-\theta}$ is consumer market access (Donaldson and Hornbeck, 2016) and X_i is total expenditures on consumption goods. Intuitively, this expression captures the idea that trade flows between i and j are higher if productivity is higher and wages are lower in destination j, if trade costs are lower between i and j, if people in the origin i have better market access, and if origin i has higher total expenditures.

Finally, the price index in any location j is given by:

$$P_j^{-\theta} = \gamma^{-\theta}\left[\sum_{k=1}^{J} A_k(\tau_{ik}W_k)^{-\theta}\right] \tag{8}$$

where $\gamma = [\Gamma(\frac{\theta-(\sigma-1)}{\theta})]^{\frac{1}{1-\sigma}}$ and $\Gamma(\cdot)$ is the Gamma function.

3.4 Labor Income

In the spirit of Monte, Redding and Rossi-Hansberg (2018), we model land as owned by local immobile landowners. These landowners spend their entire income on consumption goods in the location where they live. This implies that total expenditure on goods in location k is given by: $X_k = \alpha W_k + (1-\alpha)W_k$. The first and second terms give expenditures by workers and landowners, respectively.

Labor income in each location is equal to the total expenditures on goods produced in that location. Therefore, the expenditures of people in region k on goods produced in i is equal to k's expenditure share on goods from i multiplied by k's total expenditures: $X_{ki} = \pi_{ki}X_k = \pi_{ki}W_k$. Then total income can be written as:

$$W_iL_i = \sum_{k=1}^{J} X_{ki} = \sum_{k=1}^{J} \pi_{ki}W_k \tag{9}$$

which says that total expenditures are equal to the total wage bill in each location.

3.5 Land Prices

Income derived from land and total expenditures are equal in equilibrium. Since the utility function is Cobb-Douglas, the fraction of income spent on land is equal to $(1-\alpha)$. Therefore, exploiting the condition for land market clearing, the price of land can be written as:

$$r_j = (1-\alpha)\frac{W_jL_j}{H_j} \tag{10}$$

4 Results

In this section we first present results from estimating specifications that correspond to the theory and, therefore, yield estimates of model parameters. In particular, we estimate the migration elasticity and migration costs parameters associated with cultural distance and travel costs. We discuss the direct interpretation of these parameters. We then use the parameters to construct migration costs and carry out counterfactuals to quantify the implications of removing all migration costs or the specific components due to cultural distance and travel costs.

4.1 Determinants of Migration Flows

We conduct our analysis for 927 locations over three periods: between 1850 and 1860, between 1860 and 1870, and between 1850 and 1870. The starting point for our empirical analysis is equation (3) in which bilateral migration flows are a function of the origin i labor market access, destination j amenities and real wages, and bilateral migration costs between origin i and destination j.

$$\pi_{ij} = \underbrace{(LMA_i)^{-1}}_{\Xi_i=\text{origin } i \text{ FE}} \left(\underbrace{B_j \frac{W_j}{P_j^{\alpha} r_j^{1-\alpha}}}_{\Xi_j=\text{destination } j \text{ FE}} \right)^{\epsilon} \left(\underbrace{\mu_{ij}}_{\text{migration costs between } i \text{ and } j} \right)^{-\epsilon} \tag{11}$$

which we estimate using Poisson pseudo-maximum-likelihood and report robust standard errors clustered on origin and destination. The dependent variable, π_{ij}, is the fraction of the population in i moving to j. Origin fixed effects Ξ_i capture labor market access at origin i and destination fixed effects Ξ_j capture real wages at destination j. In addition, all specifications include a stayer dummy that is equal to one if i and j are the same location. The migration cost term is the focus of this paper. We model these costs in the following way: $\mu_{ij} = \exp(\beta_{\text{names}}\text{names}_{ij} + \beta_{\text{churches}}\text{churches}_{ij} + \beta_{\text{travel}}\text{travel}_{ij})$. The variables names_{ij} and churches_{ij} capture the bilateral distance between first name or church denomination distributions, respectively, as described in Section 2.3; travel_{ij} is the monetary cost of using the transportation network between i and j, as described in Section 2.4. Each variable is standardized so that the β parameters reflect the effect of a one standard deviation change on migration costs. The coefficients on names_{ij}, β_{churches},

Table 1: Results for Composite Migration Cost Parameters

	1850-60		1860-70		1850-70	
	(1)	(2)	(3)	(4)	(5)	(6)
Name Dissimilarity	-0.243	-0.197	-0.232	-0.185	-0.228	-0.181
	(0.036)	(0.040)	(0.035)	(0.041)	(0.034)	(0.041)
Church Dissimilarity	-0.845	-0.650	-0.858	-0.650	-0.837	-0.660
	(0.054)	(0.045)	(0.053)	(0.044)	(0.050)	(0.041)
Travel Cost		-1.871		-1.929		-1.698
		(0.042)		(0.043)		(0.038)

Notes: The table shows results for estimating equation (11). Columns 1 and 2 use migration flows between 1850 and 1860, columns 3 and 4 use migration flows between 1860 and 1870, and columns 5 and 6 use migration flows between 1850 and 1870. The dependent variable is the share of migrants between i and j. The names and church dissimilarity variables, and the travel cost variables are calculated as described in sections 2.3 and 2.4. All specifications include a stayer dummy, and origin and destination fixed effects. The method of estimation in all columns is Poisson pseudo-maximum-likelihood. Robust standard errors in parentheses are clustered on origin and destination. The number of locations (supercounties) is 927.

and travel_{ij} estimate composite parameters $\epsilon\beta_{\text{names}}$, $\epsilon\beta_{\text{churches}}$, and $\epsilon\beta_{\text{travel}}$, respectively; we can recover β parameters by assuming a value for ϵ from the literature. We also show how to estimate this parameter within our framework. The results are shown in Table 1.

Table 1 shows results for estimating equation (11) using migration flows between 1850 and 1860 (columns 1 and 2), 1860 and 1870 (columns 3 and 4), and 1850 and 1870 (columns 5 and 6). For each time period, the first column provides the estimated coefficients associated with cultural component of migration costs and the second column adds the travel cost component of migration costs. For each time period, the estimated coefficients on name dissimilarity range between -0.22 and -0.25 and on church dissimilarity range between -0.83 and -0.85 when the travel cost component is excluded (in columns 1, 3, and 5). Each set of coefficients decreases by about 20 percent when travel costs are added (in columns 2, 4, and 6), however, all coefficients are statistically significant. Since all variables are standardized the relative magnitude of the coefficients indicate the relative importance of each variable in shaping migration costs and, ultimately, migration flows. For example, based on the coefficients in column 6 of Table 1, a one standard deviation increase in (i) name dissimilarity decreases migration flows by 0.18 percent, (ii) church dissimilarity decreases migration flows by 0.67 percent, and (iii) travel costs decreases migration flows by 1.70 percent.

Table 2: Results for Composite Migration Costs Parameters with Controls

	(1)	(2)	(3)	(4)
Name Dissimilarity	-0.181	-0.208	-0.233	
	(0.041)	(0.051)	(0.046)	
Church Dissimilarity	-0.660	-0.698		-0.672
	(0.041)	(0.053)		(0.041)
Travel Cost	-1.698	-1.698	-1.792	-1.703
	(0.038)	(0.038)	(0.042)	(0.038)
Name × Church Dissimilarity		0.030		
		(0.029)		

Notes: The table shows results for estimating equation (11) using migration flows between 1850 and 1870. The dependent variable is the share of migrants between i and j. The dissimilarity (using first names and churches) and travel cost variables are calculated as described in sections 2.3 and 2.4, respectively. Column 1 includes travel costs and both first name and church dissimilarity, column 2 adds an interaction between name and church dissimilarity. Column 3 includes only name dissimilarity and travel costs. Column 4 includes only church dissimilarity and travel costs. All specifications include a stayer dummy, and origin and destination fixed effects. The method of estimation in all columns is PPML. Robust standard errors in parentheses are clustered on origin and destination. All specifications include a stayer dummy, and origin and destination fixed effects. The method of estimation in all columns is PPML. Robust standard errors in parentheses are clustered on origin and destination. The number of locations (supercounties) is 927.

The estimates presented in column 1 of Table 2 reproduce estimates in column 6 of Table 1 using migration flows between 1850 and 1870. The remaining columns also use migration flows from 1850 to 1870 and consider alternative ways of capturing cultural distance. Column 2 adds an interaction between first name and church dissimilarity. The coefficients on first name and church dissimilarity are unchanged, and the coefficient on the interaction is not statistically significant. This suggests we do not need to include the interaction between first name and church dissimilarity when constructing migration costs for the counterfactual results in Section 4.3. Columns 3 and 4 of Table 2 allow each component of cultural distance to enter the specification separately. The estimated coefficients are slightly larger than in columns 1 and 2; we will consider configurations of migration costs in Section 4.3 in which cultural distance is a function of only one or the other of these variables to assess how these specification changes influence our estimates of the welfare effects of counterfactually removing various cost components.

The model in Section 3 does not account for heterogeneity in labor types, an as-

Table 3: Results for Culture Distance by State-of-Birth and -Residence

		Born in Place-of-Residence:		Difference
	All	Yes	No	(2) − (3)
	(1)	(2)	(3)	(4)
Names Dissimilarity	-0.1811	-0.1982	-0.1088	-0.0894
	(0.0338)	(0.0363)	(0.0307)	(0.0244)
Church Dissimilarity	-0.6602	-0.6842	-0.5003	-0.1839
	(0.0191)	(0.0213)	(0.0188)	(0.0186)
Travel Cost	-1.6981	-1.9136	-1.3840	-0.5296
	(0.0172)	(0.0188)	(0.0190)	(0.0147)

Notes: The table shows results for estimating equation (11) using migration flows between 1850 and 1870. The dependent variable is the share of migrants between i and j born in the same place in which they reside in 1850 (column 1) and not born in the same place in which they reside in 1850 (column 2). Column 3 presents the difference between columns 1 and 2. All specifications include travel costs, a stayer dummy, and origin and destination fixed effects. The method of estimation in columns 1 and 2 is PPML. Bootstrapped standard errors in parentheses (from 500 replications) are clustered on origin and destination. The number of locations (supercounties) is 927.

sumption that warrants further exploration to understand its implications. To address this, Table 3 provides estimates of the coefficients on first name and church dissimilarity across three groups. The dependent variable in Column 1 includes all migrants between 1850 and 1870, offering an aggregate view of migration sensitivity to cultural distance. Columns 2 and 3 focus on more specific samples: Column 2 includes only those migrants still living in their place of birth in 1850, capturing a group for whom cultural familiarity may play a more significant role, while Column 3 includes individuals not living in their place of birth by 1850, covering both US-born migrants from other states and foreign-born migrants. Column 4 then reports the differences between each pair of coefficients in Columns 2 and 3, illustrating how cultural distance may impact these distinct groups differently. These results indicate that cultural distance influences migration decisions differently for those who had previously chosen to stay versus those who had already migrated by 1850.

There are two observations regarding the interpretation of the coefficients and differences. First, the differences between each pair of coefficients on first name and church dissimilarity are statistically significant, which suggests that the two groups of migrants are differentially sensitive to the components of cultural distance. These results are consistent with a greater role for cultural dissimilarity in shaping migration decisions of people who have never migrated or only migrated within the state in which they were born versus the group of people who have already chosen to move before 1850. Second, although the estimates are different, all coefficients are statistically significant, which reflects at least some role for cultural distance in shaping the migration decisions of people in each group. It useful to note that quantitatively the group living in their state-of-birth in 1850 is larger in size and therefore plays the most important role in shaping the aggregate responsiveness to the components of cultural distance (i.e., the coefficients in columns 1 and 2 are similar).

During our sample period, there were sharp differences and ultimately conflict between the North and South. These differences had economic, political, and socio-cultural dimensions, all of which could have lead to differences in migration patterns across regions. The key motivation of this paper is to better understand and quantify the cultural aspects of migration. The historical literature emphasizes and our main hypothesis allows that some (but not all) of what we term culture is due to Differences between the North and South.

Empirically, therefore, we are interested in whether our two measures of cultural distance only reflect broad interregional or more granular intraregional differences. In this spirit, the panels of figures C2 and C3 report estimated coefficients on first name and church dissimilarity each interacted with a set of directional (i.e., North-North, North-South, South-North, and South-South) migration flow indicators. We do this separately across each 10-year period of our sample (i.e., 1850 to 1860, 1860 to 1870) to examine whether our estimated coefficients are stable across these two decades (including the Civil War).

The results for the two variables indicate some interesting regional differences. The estimated coefficient on first name dissimilarity is statistically significant for the within-North (Figure C2A) and within-South (Figure C2B) migration flows in the 1850s or 1860s. In contrast, for migration flows across the North and South (in figures C2B and C2D), neither set of coefficients is statistically significant in either decade. The standard errors are large enough to warrant caution in making detailed comparisons. However, the pattern of the coefficients suggests that the component of cultural distance due to first name dissimilarity played more of a role in inhibiting migration within regions but less of a role across regions.

For church dissimilarity, the estimated coefficients in each panel (and for each time period) are statistically significant and have a similar magnitude. That is, church dissimilarity seems to inhibit migration both within- and between-regions in both time periods we study.

Taken together, Figures C2 and C3 suggest that the Civil War did not immediately alter the influence of cultural distance on migration flows. The abolition of slavery and subsequent Reconstruction policie, aimed to reshape Southern society, especially in a region long divided by racial and economic differences. It might be expected that these changes would lead to greater acceptance of cultural differences and increase the fluidity of migration across regional boundaries. However, our data indicates that by 1870–just a few years after the Civil War's end–these shifts had not yet significantly impacted migration patterns. This suggests that deeply ingrained cultural factors and long-standing regional identities continued to play a substantial role in migration decisions, likely due to the persistence of social and economic divides that Reconstruction alone could not quickly bridge.

Finally, we consider two sets of robustness checks for the estimates of the elasticities of migration flows with respect to different components of migrations costs. The first focuses on variations in the underlying samples of linked census data. Table 4 shows the estimated coefficients on first name and church dissimilarity as well as travel costs. Column 1 reproduces the baseline results in column 6 of Table 1 but instead linking individuals across censuses according to the procedure from the Census Tree (Price, Buckles, Van Leeuwen and Riley, 2021; Buckles, Haws, Price and Wilbert, 2023) covering both men and women between 1850 and 1870. Columns 2 and 3 show that restricting the sample to males and household heads does not substantially alter the results.

The remaining columns of Table 4 show results using the Multigenerational Longitudinal Panel (Ruggles, Fitch, Goeken, Hacker, Helgertz, Roberts, Sobek, Thompson, Warren and Wellington, 2020) and the Census Linking Project (Abramitzky, Boustan, Eriksson, Pérez and Rashid, 2020b) to construct links across census waves for men and household heads. There are minor differences in the coefficients for the cultural distance components of migration costs. For example, the magnitude of the coefficients on first name dissimilarity are larger using the Multigenerational Longitudinal Panel

Table 4: Results with Alternative Linked Samples

	Census Tree			MLP		CLP	
	All	Male	HHH	Male	HHH	Male	HHH
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Name Dissimilarity	-0.181	-0.179	-0.172	-0.204	-0.210	-0.171	-0.164
	(0.041)	(0.040)	(0.040)	(0.045)	(0.045)	(0.040)	(0.041)
Church Dissimilarity	-0.660	-0.649	-0.649	-0.709	-0.708	-0.675	-0.671
	(0.041)	(0.039)	(0.040)	(0.053)	(0.053)	(0.040)	(0.040)
Travel Cost	-1.698	-1.596	-1.575	-2.445	-2.476	-1.571	-1.539
	(0.038)	(0.038)	(0.037)	(0.049)	(0.049)	(0.038)	(0.037)

Notes: The table shows results for estimating equation (11) using migration flows between 1850 and 1870. The dependent variable is the share of migrants between i and j. The names dissimilarity and travel cost variables are calculated as described in sections 2.3 and 2.4, respectively. Columns 1 through 3 calculate migration flows by linking people across censuses via the Census Tree (Price, Buckles, Van Leeuwen and Riley, 2021; Buckles, Haws, Price and Wilbert, 2023) using all observations, only men, or only household heads. Columns 4 and 5 calculate migration flows using links from the Multigenerational Longitudinal Panel (Ruggles, Fitch, Goeken, Hacker, Helgertz, Roberts, Sobek, Thompson, Warren and Wellington, 2020) using men or only household heads. Columns 6 and 7 calculate migration flows using links from the Census Linking Project (Abramitzky, Boustan, Eriksson, Pérez and Rashid, 2020b) using men or only household heads. All specifications include a stayer dummy, and origin and destination fixed effects. The method of estimation in all columns is PPML. Robust standard errors in parentheses are clustered on origin and destination. The number of locations (supercounties) is 927.

and smaller using the Census Linking Project. The differences for the travel cost component of migration costs are larger, i.e., contrast coefficients between -1.5 and -1.6 using the Census Tree and Census Linking Project with coefficients around -2.5 using the Multigenerational Longitudinal Panel. To understand the implications of this difference, when we conduct counterfactuals, we consider a variant of our baseline scenario that uses migration flows constructed from these alternative procedures to ensure that our overall interpretation is not sensitive to particular method of linking and to provide a range of estimates of the welfare effects of cultural distance.

The second set of robustness checks considers alternative functional forms used to construct cultural distance based on first names and church denominations. Our baseline results, reproduced in column 1 of Table 5, use the functional form in equation (1) to construct dissimilarity for first names and church denominations.[9] In the baseline,

[9]The construction of the travel cost variable is unchanged. That said, in the counterfactuals, we consider versions of the travel costs based the 1860 transportation network and alternative parameterizations of

Table 5: Results with Alternative Culture Distance Functional Forms

	Dissimilarity:			Jensen-Shannon Divergence:		
	All	5K	1K	All	5K	1K
	(1)	(2)	(3)	(4)	(5)	(6)
Name Distance	-0.181	-0.183	-0.187	-0.222	-0.229	-0.237
	(0.041)	(0.040)	(0.040)	(0.038)	(0.039)	(0.040)
Church Distance	-0.660	-0.660	-0.660	-0.756	-0.754	-0.752
	(0.041)	(0.041)	(0.041)	(0.048)	(0.048)	(0.048)
Travel Cost	-1.698	-1.698	-1.698	-1.649	-1.648	-1.649
	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)

Notes: The table shows results for estimating equation (11). using migration flows between 1850 and 1870. The dependent variable is the share of migrants between i and j. Travel costs are calculated as described in section 2.4. The first three columns calculate first names distance using equation 1 as in section 2.3, but varying the set of names: all names (column 1) and the 5,000 (column 2) or 1,000 (column 3) most common first names in the national distribution. The second three columns calculate first names distance using Jensen-Shannon Divergence with all names (column 4) and the the 5,000 (column 5) or 1,000 (column 6) most common first names in the national distribution. All specifications include a stayer dummy, and origin and destination fixed effects. The method of estimation in all columns is PPML. Robust standard errors in parentheses are clustered on origin and destination. The number of locations (supercounties) is 927.

the set of first names is all 50,000+ names that we were able to identify in the national distribution in 1850. For a particular location, this can mean a distribution with many entries equal to zero. To ensure that our results are not driven by an abundance of zeros in the local first name distributions, columns 2 and 3 restrict the set of first names to the 5,000 or 1,000 most common names in the national distribution; the results are similar. The remaining columns of Table 5 use an alternative measure of differences based on the Jensen-Shannon Divergence to construct both first name and church distances. As above, the first name distance is based on all first names (column 4), the most popular 5,000 first names (column 5), or the most popular 1,000 first names (column 6) in the national distribution. The results are slightly larger in magnitude using this alternative distance metric. For our counterfactuals, we thus also consider robustness to this alternative functional form for the distance metric.

the network. We discuss the relevant details in Section 4.3.

Table 6: Results for Migration Elasticity and Migration Cost Function

Parameter	(1)	(2)	(3)	(4)
A. Elasticity of migration flows w.r.t migration costs				
ϵ	-15.8980			
	(0.1473)			
B. Elasticity of migration costs w.r.t cost components				
β_{names}		0.0114	0.0146	
		(0.0019)	(0.0021)	
β_{church}		0.0415		0.0422
		(0.0012)		(0.0012)
β_{travel}		0.1068	0.1127	0.1071
		(0.0003)	(0.0001)	(0.0003)

Notes: The table shows the estimated values of the migration elasticity and migration cost parameters based on migration flows between 1850 and 1870. Column 1 is the migration elasticity. The remaining columns provide results for different configurations of the migration cost function using name dissimilarity, travel costs, and church dissimilarity (column 2), name dissimilarity and travel costs (column 3), and church dissimilarity and travel costs (column 4). The method of estimation is PPML controlling for a stayer dummy and origin and destination fixed effects. Bootstrapped standard errors in parentheses (from 500 replications) are clustered on origin and destination. The number of locations (supercounties) is 927.

4.2 Migration Costs and Other Model Parameters

The empirical results so far recover composite parameters that combine the migration elasticity (ϵ) and the elasticities of migration costs with respect to first name and church dissimilarity (β_{names} and β_{church}) and travel costs (β_{travel}). To reconstruct migration costs and carry out the counterfactual we require estimates of the migration elasticity and the parameters of the migration cost function. First, we can estimate the migration elasticity by transforming the travel costs portion of migration costs into an iceberg form by using $\tilde{\mu}_{ij} = 1 + (\text{travel}_{ij}/\text{average income in US}_{1850})$. We use $\tilde{\mu}_{ij}$ in place of μ_{ij} to estimate equation (11). Second, we use estimates of ϵ to recover the implied estimates of β_{names}, β_{church}, and (β_{travel}. We estimate all parameters jointly and provide bootstrapped standard errors (based on 500 replications) clustered on origin and destination.

The results are reported in Table 6. Panel A (column 1) reports an estimate of the migration elasticity, 15.9, that is statistically significant. This is larger than the value of 3.3 in Monte, Redding and Rossi-Hansberg (2018) from more recent times. For the counterfactuals, we use a migration elasticity of 16.0 for our baseline simulations; we also consider alternative values of 12 and 20. Panel B provides estimates of the migration cost function parameters under alternative configurations; all estimates are statistically significant. In column 2, we allow migration costs to be a function of first name dissimilarity, church dissimilarity, and travel costs; in column 3, migration costs are a function of first name dissimilarity and travel costs; and, in column 4, migration costs are a function of church dissimilarity and travel costs. The estimates are not too sensitive to the alternative configurations. We use the estimates in column 2 as our baseline migration cost parameters for counterfactuals; we also report counterfactual results that use the configurations and corresponding estimates of migration costs in columns 3 and 4.

Finally, to carry out counterfactuals we also require values of the share of the tradeable good in consumption (α), the trade elasticity (θ), and the wage bill 1870. We let $\alpha = 0.3$ in all counterfactuals (Yang, 2024). For the trade elasticity we use $\theta = 8$, which we take from Donaldson and Hornbeck (2016), in our baseline simulations. We also consider alternative trade elasticities of 6 and 10. For the wage bill, we use information from agricutural wages with board and manufacturing wages at the county level from Haines (2010).

4.3 Counterfactual Migration Costs, Welfare, and the Spatial Population Distribution

In this section we quantify the implications of counterfactual migration costs. In particular, we consider changes in welfare under three types of scenarios. First, we counterfactually remove all migration frictions (i.e., both the cultural distance and travel cost components). Second, we remove only migration frictions associated with cultural distance. Finally, we remove only the travel cost component. The results are reported in Table 7 where the different rows allow for alternative ways of measuring cultural distance: the baseline results define cultural distance using both first names and church denominations, the first alternative defines cultural distance using only first names, and the second alternative uses only church denomination. This exercise provides a decomposition for the welfare effects of each migration cost component.

For the baseline case in the first row, removing all migration frictions counterfactually increases welfare by 48.6 percent (column 1) and this can be decomposed into a welfare increase of 5.5 percent due to cultural distance (column 2) versus 44.3 per-

Table 7: Results for Welfare Change from Counterfactual Migration Costs

	Cultural Distance: Names	Cultural Distance: Churches	Travel Costs in 1850	Welfare change (in %) from removing: Cultural & Travel Components (1)	Cultural Distance (2)	Travel Costs (3)
Baseline	✔	✔	✔	48.6	5.5	44.3
Alternative #1	✔		✔	48.4	1.4	47.3
Alternative #2		✔	✔	48.2	4.5	44.5

Notes: The table shows counterfactual changes in welfare for three configurations of migrations costs. In the baseline scenario (row 1), the cultural distance component of migration costs includes both first name and church dissimilarity. The first alternative scenario (row 2), the cultural distance component includes only first name dissimilarity. In the second alternative scenario (row 3), the cultural distance component includes only church dissimilarity. In all scenarios the travel cost component of migration costs is based on the 1850 transportation network. For each scenario, column 1 gives the welfare gain associated of removing both cultural distance and travel cost components of migration costs, column 2 gives the welfare gain of removing only the cultural distance component, and column 3 gives the welfare gain of removing only the travel cost component.

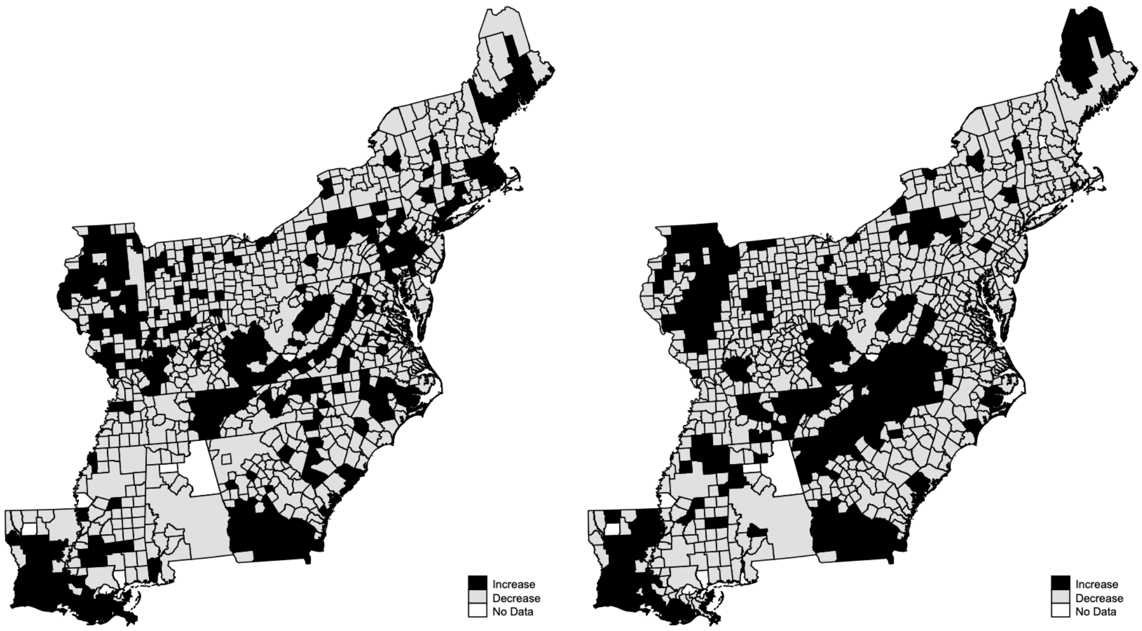
cent due to travel costs (column 3).[10] These results indicate that the welfare gains in the absence of migration costs are substantial and that travel costs are the most important component of total migration costs: cultural distance accounts for just 11 percent of the total effect. The two alternative ways of capturing cultural distance in the second and third rows indicate that differences due to church denominations account for a larger fraction of the welfare impact than differences due to first names.

Finally, to understand the implications of the different components of migration costs for the spatial distribution of population, Figure 3 shows the locations with an increase (black) or a decrease (light gray) in population as a result of removing cultural distance (Panel A) versus travel costs (Panel B). Counterfactual population increases more toward the west in both scenarios, however, the overlap is not perfect This suggests that these two components of migration costs have different qualitative implications for where people lived in the middle of the nineteenth century.

Table 8 examines the robustness of the baseline counterfactuals to alternative metrics for cultural distance and travel costs as well as inputs based on census linking and values of the migration elasticity. The baseline results from Table 7 appear in the first row. To start, we consider three variants of cultural distance: (i) uses only the 1,000

[10]The values is columns 2 and 3 do not necessarily sum to the value in column 1 due to non-linear, general equilibrium interactions in the model.

Figure 3: Results for Population Change from Removing Cost Components



A. Cultural Distance B. Travel Costs

Notes: The figure shows the locations where the counterfactual change in population is positive due to eliminating cultural distance (Panel A) or travel costs (Panel B), while holding the other component fixed.

most popular first names in the national distribution to compute distance, (ii) calculates distance as the Jensen-Shannon Divergence between county pairs, and (iii) uses Jensen-Shannon Divergence and includes only the 1,000 most popular first names in the national distribution. In each scenario, the welfare change from removing both the cultural and travel cost components (column 1) are similar. The welfare change due to the cultural components (column 2) is larger and ranges from 5.6 to 9.6 percent, while the welfare change due to the travel cost component (column 3) is smaller.

The next set of results considers alternative travel cost variables. In particular, we allow for travel costs to be determined by the 1860 transportation network and either use the parameterization of the network costs from Donaldson and Hornbeck (2016) or use the a parameterization based on estimates of passenger costs from Jaremski (2012). Making these changes leads to an aggregate impact of removing migration costs between 43.6 and 49.3 percent, while the effect of removing the cultural distance components ranges from 2.6 to 6.0 percent.

Finally, Table 8 includes robustness results using alternative samples to construc-

Table 8: Robustness for Cultural Distance, Travel Costs, Linking, and Parameters

	Welfare change (in %) from removing:		
	Cultural & Travel Components	Cultural Distance	Travel Costs
	(1)	(2)	(3)
Baseline	48.6	5.5	44.3
Cultural Distance			
Dissimilarity, Top 1,000 Names	48.6	5.6	44.3
Jensen-Shannon, All Names	49.8	9.3	42.7
Jensen-Shannon, Top 1,000 Names	49.5	8.7	42.7
Travel Costs			
Travel Costs, 1860	49.3	6.0	44.4
Passenger Costs, 1860	43.6	2.5	40.1
Census Linking			
Census Tree, Male, 20-year	44.0	5.1	39.6
Census Linking Project, Male, 20-year	45.9	5.4	42.6
Multi. Longitudinal Panel, Male, 20-year	39.6	4.1	33.3
Census Tree, All, 10-year	43.6	3.6	38.2
Migration Elasticity			
$\epsilon = 12$	38.6	4.5	33.6
$\epsilon = 20$	56.6	6.9	53.5

Notes: The table shows the robustness of counterfactual changes in welfare for alternative measures of cultural distance, travels costs, census linking, and migration elasticity. For each scenario, column 1 gives the welfare gain associated of removing both cultural distance and travel cost components of migration costs, column 2 gives the welfare gain of removing only the cultural distance component, and column 3 gives the welfare gain of removing only the travel cost component.

tion migration rates as well as alternative values of the migration elasticity. Across samples of 20-year migration horizons for men constructed using the Census Tree, Census Linking Project, and Multigenerational Longitudinal Panel, the aggregate impact of removing migration costs is between 39.6 and 45.9 percent with the cultural component between 4.1 and 5.4 percent. Using a 10-year migration horizon constructed using men and women from the Census Tree yields an estimated 3.6 percent change in welfare from removing the cultural component of migration costs. The results for alternative values of the migration elasticity are as expected: a lower (higher) value of the migration elasticity decreases (increases) the welfare impact of total migration costs and each component.

5 Conclusion

In addition to the standard "push" and "pull" forces as well as bilateral costs that shape migration flows, potential migrants may also be attentive to cultural differences between locations. In this paper, we propose new measures of cultural distance and quantify its importance for migration flows in the United States in the middle of the nineteenth century. In particular, our measures of cultural distance are based on (1) the dissimilarity of first name distributions, where we rely on the hypothesis that first names carry distinctive information about culture following Zelinsky (1970, 1992) as well as more recent work in economics and economic history, and (2) the dissimilarity of church denominations, since religious differences played an important role in early US history (Baltzell, 1979; Fischer, 1989; Phillips, 1999).

We then use a gravity framework to estimate the parameters of a migration cost function that includes both pecuniary travel costs and cultural distance. We find that the contribution of cultural distance to migration costs is quantitatively important and statistically significant. We then use a simple economic geography model to quantify the overall importance of the two components of travel costs. Our findings suggest that travel costs are substantially more important than cultural distance in terms of welfare: removing cultural distance from migration costs increases welfare by 5.5 percent versus welfare gains of 44.3 percent from removing travel costs. Overall, we view our estimates for the aggregate impact of cultural distance as a lower bound given our focus only on the bilateral component of location decisions. Building on this, future work should focus on disentangling the culturally-specific components that makes origins and destinations relatively more (or less) attractive.

Interestingly, we also show that cultural distance and travel costs have different qualitative effects on the distribution of population: removing cultural distance counterfactually pushes population both northeast and to the boundaries of the states in our sample, while removing travel costs pushes population towards Appalachia and towards more distant regions. This has important implications for pace and pattern of western settlement as well as the types of economic, political, and social institutions that are established across the United States.

While our results suggest a relatively larger role for travel costs in the nineteenth century, technological change has made travel costs much more uniform today. This, combined with the significant and sizable welfare impact of cultural distance in our data, suggests that cultural distance may remain an important determinant of internal migration today. Indeed, it seems plausible that culture matters relatively more in an era of inexpensive travel and migration. We thus suggest that cultural distance remains one plausible explanation for the observation that people today migrate less than expected given economic fundamentals. Additional work is needed to test this hypothesis.

It is also worthwhile to think about how changing patterns of migration resulting from changes in travel costs over time may lead to changes in regional cultural configurations (Fouka and Serlin, 2023). Changes in "market access" have been argued to result in more individualistic cultural patterns (Posch and Raz, 2024), and future work should more generally consider to what extent increasing ease of mobility contributed to the emergence of a national culture in the US.

References

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson, “Do Immigrants Assimilate More Slowly today than in the past?,” *American Economic Review: Insights*, 2020, *2* (1), 125–41.

_ , _ , _ , Santiago Pérez, and Myera Rashid, “Census Linking Project: Version 1.0 [dataset],” *URL https://censuslinkingproject.org*, 2020.

Assouad, Lydia, “Charismatic Leaders and Nation Building,” 2020. halshs-02873520v2.

Bahrami-Rad, Duman, Anke Becker, and Joseph Henrich, “Tabulated Nonsense? Testing the Validity of the Ethnographic Atlas,” *Economics Letters*, 2021, *204*, 109880.

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, “Social Sonnectedness: Measurement, Determinants, and Effects,” *Journal of Economic Perspectives*, 2018, *32* (3), 259–80.

Baltzell, E. Digby, *Puritan Boston and Quaker Philadelphia: Two Protestant Ethics and the Spirit of Class Authority and Leadership*, New York: The Free Press, 1979.

Barrett, Tyson, Matt Dowle, Arun Srinivasan, Jan Gorecki, Michael Chirico, and Toby Hocking, *data.table: Extension of ‘data.frame‘* 2024. R package version 1.15.4.

Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse, “Frontier Culture: The Roots and Persistence of “Rugged Individualism” in the United States,” *Econometrica*, 2020, *88* (6), 2329–2368.

Binbin Lu, Huabo Sun, Miaozhong Xu, Paul Harris, and Martin Charlton, “Shp2graph: Tools to Convert a Spatial Network into an Igraph Graph in R,” *ISPRS International Journal of Geo-Information*, 2018, *7* (8), 293.

Bleakley, Hoyt and Paul Rhode, “The Economic Effects of American Slavery, Redux: Tests at the Border,” 2022. Working Paper.

Bleidorn, Wiebke, Felix Schönbrodt, Jochen E Gebauer, Peter J Rentfrow, Jeff Potter, and Samuel D Gosling, “To live among like-minded others: Exploring the links between person-city personality fit and self-esteem,” *Psychological science*, 2016, *27* (3), 419–427.

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley E.B. Wilbert, “Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project,” Technical Report, National Bureau of Economic Research 2023.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng, *binsreg: Binscatter Estimation and Inference* 2023. R package version 1.0.

Cebula, Richard J, “Migration and the Tiebout-Tullock Hypothesis Revisited,” *American Journal of Economics and Sociology*, 2009, *68* (2), 541–551.

Clark, David E and William J Hunter, "The impact of economic opportunity, amenities and fiscal factors on age-specific migration rates," *Journal of Regional Science*, 1992, *32* (3), 349–365.

Cooke, Thomas J, "Internal migration in decline," *The Professional Geographer*, 2013, *65* (4), 664–675.

Csárdi, Gábor, Tamás Nepusz, Vincent Traag, Szabolcs Horvát, Fabio Zanini, Daniel Noom, and Kirill Müller, *igraph: Network Analysis and Visualization in R* 2024. R package version 2.0.3.

de Tocqueville, Alexis, *Democracy in America*, University of Chicago Press, Chicago, IL (2002), 1835.

Deming, David J., "The Growing Importance of Social Skills in the Labor Market," *The Quarterly Journal of Economics*, 2017, *132* (4), 1593–1640.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg, "Culture, Ethnicity, and Diversity," *American Economic Review*, September 2017, *107* (9), 2479–2513.

Dijkstra, Edsger W., "A Note on Two Problems in Connexion with Graphs," *Numerische Mathematik*, 1959, *1* (1), 269–271.

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

Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Eckert, Fabian, Andrés Gvirtz, Jack Liang, and Michael Peters, "A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790," 2020. NBER Working Paper.

Egger, Peter and Doina Maria Radulescu, "The influence of labour taxes on the migration of skilled workers," *World Economy*, 2009, *32* (9), 1365–1379.

Eli, Shari, Laura Salisbury, and Allison Shertzer, "Ideology and Migration After the American Civil War," *Journal of Economic History*, 2018, *78* (3), 822–861.

Faust, Drew Gilpin, *The creation of Confederate nationalism: Ideology and identity in the Civil War South*, LSU Press, 1989.

Fields, Gary S, "Labor force migration, unemployment and job turnover," *The Review of Economics and Statistics*, 1976, pp. 407–415.

Finke, Roger and Rodney Stark, *The Churching of America, 1776-2005: Winners and Losers in our Religious Economy*, Rutgers University Press, 2005.

Finley, John, *The Hoosier's Nest and Other Poems*, Moore, Wilstach, & Baldwin, Printers, 1866.

Fischer, David Hackett, *Albion's Seed: Four British Folkways in America*, New York: Oxford University Press, 1989.

_ and James C. Kelly, *Bound Away: Virginia and the Westward Movement*, University of Virginia Press, 2000.

Fogel, Robert William, "A quantitative approach to the study of railroads in American economic growth: a report of some preliminary findings," *The Journal of Economic History*, 1962, *22* (2), 163–197.

Fouka, Vasiliki and Theo Serlin, "Industry and Identity: The Migration Linkage Between Economic and Cultural Change in 19th Century Britain," 2023. Working Paper.

Gandrud, Christopher, *DataCombine: Tools for Easily Combining and Cleaning Data Sets* 2016. R package version 0.2.21.

Garbuszus, Jan Marvin and Sebastian Jeworutzki, *readstata13: Import 'Stata' Data Files* 2023. R package version 0.10.1.

Haines, Michael R., *Historical, Demographic, Economic, and Social Data: The United States, 1790-2000*, Vol. ICPSR02896-v3, Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010.

Hall, Patricia Kelly and Steven Ruggles, ""Restless in the Midst of Their Prosperity": New Evidence on the Internal Migration of Americans, 1850–2000," *The Journal of American History*, 2004, *91* (3), 829–846.

Harris, John R and Michael P Todaro, "Migration, unemployment and development: a two-sector analysis," *The American economic review*, 1970, *60* (1), 126–142.

Hijmans, Robert J., *geosphere: Spherical Trigonometry* 2022. R package version 1.5-18.

_ , *raster: Geographic Data Analysis and Modeling* 2023. R package version 3.6-26.

Hoberg, Gerard and Gordon Phillips, "Text-based network industries and endogenous product differentiation," *Journal of Political Economy*, 2016, *124* (5), 1423–1465.

Jaremski, Matthew, "Estimating Antebellum Passenger Costs: A Hub-and-Spoke Approach," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2012, *45* (2), 93–101.

Jaworski, Taylor, Erik Kimbrough, Pranit Nanda, and Manav Chakravarthy, "Using Machine Learning to Identify Regional Cultures in the United States," 2020. Unpublished.

Jr, Henry W Herzog and Alan M Schlottmann, "Labor force mobility in the United States: migration, unemployment, and remigration," *International Regional Science Review*, 1984, *9* (1), 43–58.

Jurajda, Štěpán and Dejan Kovač, "Names and behavior in a war," *Journal of Population Economics*, 2021, *34* (1), 1–33.

Knudsen, Anne Sofie Beck, "Those Who Stayed: Selection and Cultural Change during the Age of Mass Migration," 2019. Working Paper.

Larson, John Lauritz, *The Market Revolution in America: Liberty, Ambition, and the Eclipse of the Common Good*, Cambridge University Press, 2009.

Main, Gloria L., "Naming Children in Early New England," *The Journal of interdisciplinary history*, 1996, *27* (1), 1–27.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles, "IPUMS National Historical Geographic Information System: Version 16.0 [dataset]. Minneapolis, MN," 2021.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak, "Internal Migration in the United States," *Journal of Economic Perspectives*, 2011, *25* (3), 173–196.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, 2018, *108* (12), 3855–3890.

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *NBER Working Paper*, 2018, *22158*, 1–64.

Munshi, Kaivan, "Community Networks and the Process of Development," *Journal of Economic Perspectives*, 2014, *28* (4), 49–76.

Padgham, Mark and Michael D. Sumner, *geodist: Fast, Dependency-Free Geodesic Distance Calculations* 2024. R package version 0.1.0.

Partridge, Mark D and Dan S Rickman, "Differences in state unemployment rates: The role of labor and product market structural shifts," *Southern Economic Journal*, 1995, *62* (1), 89–106.

_ and _ , "The dispersion of US state unemployment rates: the role of market and non-market equilibrium factors," *Regional Studies*, 1997, *31* (6), 593–606.

Pebesma, Edzer, "Simple Features for R: Standardized Support for Spatial Vector Data," *The R Journal*, 2018, *10* (1), 439–446.

Pebesma, Edzer J. and Roger Bivand, "Classes and methods for spatial data in R," *R News*, November 2005, *5* (2), 9–13.

Pedersen, Thomas Lin, *tidygraph: A Tidy API for Graph Manipulation* 2024. R package version 1.3.1.

Phillips, Kevin, *The Cousins' Wars: Religion, Politics, Civil Warfare, and The Triumph of Anglo-America*, Basic Books: New York, NY, 1999.

Posch, Max and Itzchak Tzachi Raz, "Markets Make Humans WEIRDer: Evidence from 1850-1920 United States," 2024. Working Paper.

Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley, "Combining Family History and Machine Learning to Link Historical Records: The Census Tree data set," *Explorations in Economic History*, 2021, *80*, 101391.

Putnam, Robert D, *Bowling alone: America's declining social capital*, Springer, 2000.

R Core Team, *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing 2020.

Ravenstein, Ernest George, "The laws of migration," *Journal of the statistical society of London*, 1885, *48* (2), 167–235.

_ , "The laws of migration," *Journal of the royal statistical society*, 1889, *52* (2), 241–305.

Rhode, Paul W and Koleman S Strumpf, "Assessing the importance of Tiebout sorting: Local heterogeneity from 1850 to 1990," *American Economic Review*, 2003, *93* (5), 1648–1677.

Rosenbloom, Joshua L. and William A. Sundstrom, "The Decline and Rise of Interstate Migration in the United States: Evidence from the IPUMS, 1850–1990," in "Research in economic history," Emerald Group Publishing Limited, 2004.

Ruggles, Steven, Catherine Fitch, Ron Goeken, J. David Hacker, Jonas Helgertz, Evan Roberts, Matt Sobek, Kelly Thompson, John Robert Warren, and Jacob Wellington, "IPUMS Multigenerational Longitudinal Panel: Version 1.0 [dataset]. Minneapolis, MN," 2020.

_ , Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Robers, and Warren J. Robert, "IPUMS Ancestry Full Count Data: Version 4.0 [dataset]. Minneapolis, MN," 2024.

Sellers, Charles, "The Market Revolution: Jacksonian America, 1815-1846 364-69," 1991.

Smith, Daniel Scott, "Child-Naming Practices, Kinship Ties, and Change in Family Attitudes in Hingham, Massachusetts, 1641 to 1880," *Journal of Social History*, 1985, *18* (4), 541–566.

_ , "Continuity and Discontinuity in Puritan Naming: Massachusetts, 1771," *William and Mary Quarterly*, 1994, *51* (1), 67–91.

Stark, Oded and David E Bloom, "The new economics of labor migration," *The American Economic review*, 1985, *75* (2), 173–178.

Stewart, George R., *Men's Names in Plymouth and Massachusetts in the Seventeenth Century*, University of California Press, 1948.

Taylor, Alan, *American Republics: A Continental History of the United States, 1783-1850*, W.W. Norton & Company, 2021.

Tennekes, Martijn, “tmap: Thematic Maps in R,” *Journal of Statistical Software*, 2018, *84* (6), 1–39.

Teucher, Andy and Kenton Russell, *rmapshaper: Client for 'mapshaper' for 'Geospatial' Operations* 2023. R package version 0.5.0.

Thernstrom, Stephan and Peter R. Knights, “Men in Motion: Some Data and Speculations about Urban Population Mobility in Nineteenth-Century America,” *The Journal of Interdisciplinary History*, 1970, *1* (1), 7–35.

Tiebout, Charles M, “A pure theory of local expenditures,” *Journal of Political Economy*, 1956, *64* (5), 416–424.

van der Meer, Lucas, Lorena Abad, Andrea Gilardi, and Robin Lovelace, *sfnetworks: Tidy Geospatial Networks* 2024. R package version 0.6.4.

Wickham, Hadley, “Reshaping data with the reshape package,” *Journal of Statistical Software*, 2007, *21* (12).

_ , Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani, “Welcome to the tidyverse,” *Journal of Open Source Software*, 2019, *4* (43), 1686.

Wilson, Riley, “Isolated States of America: The Impact of State Borders on Mobility and Regional Labor Market Adjustments,” *Available at SSRN*, 2021.

Woodard, Colin, *American Nations: A History of the Eleven Rival Regional Cultures of North America*, New York: Penguin, 2011.

Yang, Dongkyu, “A General Equilibrium Investigation of the American Dust Bowl,” Technical Report 2024.

Zelinsky, Wilbur, “Cultural variation in personal name patterns in the Eastern United States,” *Annals of the Association of American Geographers*, 1970, *60* (4), 743–769.

_ , *The Cultural Geography of The United States: A Revised Edition*, Pearson, 1992.