NBER WORKING PAPER SERIES

IMMIGRANT DISTRIBUTION IN THE UNITED STATES DURING THE AGE OF MASS MIGRATION

Ariell Zimran

Working Paper 28812
http://www.nber.org/papers/w28812

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2021

For helpful comments, I thank Jeremy Atack, Brian Beach, Elizabeth Cascio, William Collins, Chloe East, James Feigenbaum, Sarah Quincy; participants in the Junior Economic History Workshop; and seminar participants at Vanderbilt University. I am grateful to the Minnesota Population Center for providing data; Matt Nelson in particular provided valuable assistance. Work on this paper was completed while I was a W. Glenn Campbell and Rita Ricardo-Campbell National Fellow and the William C. Bark National Fellow at the Hoover Institution, Stanford University and while I was a Visiting Scholar at the Opportunity and Inclusive Growth Institute at the Federal Reserve Bank of Minneapolis; funding from both of these institutions is gratefully acknowledged. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. Computations were performed on the Advanced Computing Cluster for Research and Education at Vanderbilt University. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Immigrant Distribution in the United States during the Age of Mass Migration
Ariell Zimran
NBER Working Paper No. 28812
May 2021
JEL No. F22,J11,J15,J61,N31,N32,R23

ABSTRACT

Immigrant distribution--the geographic dispersion of immigrants in the destination country--was a major issue in the United States in the late Age of Mass Migration. Policy debates were influenced by the widely held view that the new immigrants were generally less geographically mobile within the United States and specifically less likely to leave urban areas than were natives and earlier immigrants. I build new linked census datasets to investigate these claims by studying the rates of, selection into, and sorting of internal migration by US immigrants. I find that contemporary claims regarding immigrant distribution were either false, oversimplified, or the product of broader national trends that applied also to natives. Nonetheless, geographic assimilation--convergence in immigrants' and natives' county-of-residence distributions over time in the United States--was almost nonexistent.

Ariell Zimran
Department of Economics
Vanderbilt University
2301 Vanderbilt Place
Nashville, TN 37235
and NBER
ariell.zimran@vanderbilt.edu

A data appendix is available at http://www.nber.org/data-appendix/w28812

1 Introduction

The year 1907 was one of the most important in US immigration history. Its nearly 1.3 million immigrant arrivals marked the highest annual total that the United States had ever seen.[1] It was also the first year in which over 80 percent of the immigrant inflow from Europe originated in its southern and eastern periphery, capping the transition from an inflow dominated by the *old immigrants* of northwestern Europe to one dominated by these *new immigrants* (Barde, Carter, and Sutch 2006a, b; Haines and Sutch 2006; US Congress 1911a).[2] The changing character of US immigration resulting from this change in its geographic origins exacerbated concerns about the effects of immigration on the US economy, and led to the formation, in 1907, of the *Dillingham Commission* to "make a full inquiry, examination, and investigation ... into the subject of immigration" (US Congress 1911a, p. II).

The Commission's 41-volume report, published in 1911, was a landmark analysis of US immigration. It also included a number of policy recommendations, many of which, including the imposition of a literacy test and national origin quotas, eventually became law. Among these recommendations was that efforts be made to achieve *immigrant distribution*—that is, that immigrants be encouraged to leave coastal cities and move inland. Some influential members of the Commission and its staff, as well as many other stakeholders in the immigration debate, viewed distribution as a way to blunt many of the perceived ills of immigration, such as the insalubrious nature of cities that were swelled by immigrants and the effects of immigration on wages in immigrant-dense areas.[3] As these were among the reasons cited as justification for restriction of immigration (e.g., Benton-Cohen 2018; Commissioner-General of Immigration 1903; Goldin 1994; Hall 1904; US Congress 1911a; Willcox 1906), successful distribution was seen by many as a way to keep the Golden Door open to immigration in the face of rising calls to close it (Benton-Cohen 2018).

The Commission's call for distribution policy, as well as those of other proponents of such efforts, were based on beliefs that the new immigrants were generally immobile, and specifically reluctant to leave urban areas, exhibiting a greater attachment than natives or earlier waves of immigrants. This tendency was seen as particularly strong among illiterate immigrants. Although this view was questioned by some (Willcox

[1] This total was not equaled until the 1990s and may have even been the highest total ever, since totals in the 1990s may be overstatements due to the Immigration Reform and Control Act of 1986 and by undocumented immigration. The per capita immigrant inflow in 1907 was 1.48 percent, which is the fifth-highest annual rate in US history. The highest per capita annual rate of immigration in US history was 1.54 percent in 1854.

[2] The distinction between new and old immigrants was popularized by the Dillingham Commission (Benton-Cohen 2018, p. 6). The old immigrants were primarily from Germany, Britain, and Ireland, and included Scandinavians. The new immigrants were primarily from Italy, the Russian Empire and the Austro-Hungarian Empire. See Gould (1980), Hatton and Williamson (1998), and Spitzer and Zimran (2021) for detailed discussions of this change in source countries.

[3] I do not evaluate whether it is true that distribution would have solved these problems. Whether this was true or not, there were many influential people who believed that it was, and this influenced their views on immigration policy.

1906), it was held across the spectrum of positions on immigration policy: in addition to those who supported distribution efforts, those who opposed immigration regardless of its distribution cited this perceived tendency as evidence of the need for greater restriction. Regardless of the prescription derived from these perceptions of immigrant immobility, there is no doubt that they played a prominent role in the early twentieth-century immigration policy debate. But they have never before been rigorously tested. Limited information on immigrants' internal migration in the United States has constrained prior scholars' ability to interrogate them. In this paper, I build new linked census datasets to investigate claims of immigrant immobility in greater detail than has previously been possible.

Issues that arose in immigration policy debates during the Age of Mass Migration often resurface in modern policy debates and in the modern economics of immigration.[4] Immigrant distribution is no exception. It is a major issue in current refugee policy debates in the United States and in the European Union (Abbott 2020; Brenke 2015; Kaufmann 2020). Moreover, the geographic distribution of immigrants in the receiving county—including the within-city segregation of immigrants and natives as well as immigrants' destination choices in the receiving country—is a central issue in the modern economics of immigration.[5] Immigrants' internal migration in the destination country is also known to be important to understanding the effects of immigration on the receiving economy (Borjas 2001; Cadena and Kovak 2016). Nonetheless, the literature on modern immigrants' internal mobility is very limited (c.f., Bartel 1989; Bartel and Koch 1991; Maré, Morten, and Stillman 2007).[6] That is, whereas the location of immigrants in the destination at any given time is well studied, changes in these locations with time in the destination, the focus of this paper, are not well understood despite their recognized importance.

My analysis focuses specifically on two closely related phenomena that shed light on immigrant distribution in the Age of Mass Migration. The first, immigrants' *geographic assimilation*, describes the degree to which immigrants' distribution of places of residence converged to that of natives over time in the United States. Because immigrants generally arrived in coastal cities, this assimilation is inherently linked to their *internal migration* patterns—the second focus of this paper. Little is known about both of these phenomena in this context, in part because they both can be studied only with data linked between censuses, which, until

[4]For instance, the modern immigration policy debate includes arguments over selection and assimilation that echo those from the Age of Mass Migration, and in many cases make explicit references to popular perceptions of historic migration to justify modern policy proposals (e.g., Collins and Zimran 2019, 2020; Spitzer and Zimran 2018).

[5]See, for instance, Borjas (1999, 2001, 2014), Cadena (2013, 2014), Cadena and Kovak (2016), Card and Lewis (2007), Cascio and Lewis (2012), Cutler, Glaeser, and Vigdor (2008a,b), Damm (2009), Edin, Frederiksson, and Åslund (2003), and Jaeger (2007).

[6]See also Beenstock (1997, 1999), Gurak and Kritz (2000), Hall (2009), Henning (1997), Kritz and Gurak (2001), Kritz, Gurak, and Lee (2011), Newbold (1999), Rashid (2009), Schündeln (2014), and Weine et al. (2011). Cadena and Kovak (2016, p. 260) point out that immigrants' internal migration in the destination is primarily studied in literatures outside of economics.

recently, have been difficult to construct. As with all facets of immigrant assimilation in the Age of Mass Migration, studying geographic assimilation requires linked data to address return migration and differences between arrival cohorts. Linked data are also required to study historic US immigrants' internal migration because no other source provides systematic information on their place of residence in the United States at two points in time.[7] To this end, I construct 10 new linked census datasets using complete-count US censuses for 1850–1930 with links spanning all possible 10- and 20-year spans. These linked data cover European- and native-born white men aged 18–40 in the initial year. They enable me to document, for the first time for this period, immigrants' rates of, selection into, and sorting of inter-county migration, to compare these patterns to those of natives to better contextualize them, and to measure immigrants' geographic assimilation.

I first test claims that the immigrants of the later Age of Mass Migration were less mobile within the United States than previous cohorts of immigrants. I do this by using the linked datasets to estimate the inter-county migration rates of immigrants. I find evidence of a large decline in immigrants' inter-county migration rates by about 15 percentage points from the beginning of my study period to the end, consistent with these claims. But the decline occurred almost entirely at the very beginning of my study period (i.e., from linkage spans beginning 1850 to those beginning 1860). From this point to the end of the Age of Mass Migration, immigrants' inter-county migration rates were largely constant or even slightly rising. I therefore conclude that the claim of reduced immigrant mobility, though technically true, was oversimplified. Immigrants in the twentieth century were no less mobile than any previous cohort of immigrants other than the very first, who arrived before the surge of mass immigration in the 1850s. I also contextualize the inter-county migration rates of immigrants by comparing them to those of natives, casting further doubt on more general claims of immigrant immobility: there is no evidence that immigrants were ever meaningfully less mobile than natives; in fact, most estimates indicate the opposite.

Contemporaries also claimed that foreign-born internal migrants were positively selected from among the foreign born, in general and specifically on the basis of literacy; that is, poorer and illiterate immigrants were more likely to remain in arrival cities. Foreign-born residents of urban areas were also seen as particularly immobile (e.g., Willcox 1906). All of this selection was argued to have become stronger over time. I evaluate these claims by measuring foreign-born individuals' selection into inter-county migration on the basis of literacy, urban status, and occupational status, measured by an individual's occupational rank. Again, I compare these patterns to those of natives in order to contextualize them. I find no evidence of positive

[7] Censuses prior to 1940 did not inquire as to an individual's previous place of residence (aside from birthplace). Absent information on prior place of residence, prior studies of internal migration over the long run in the United States (Ferrie 2006a; Hall and Ruggles 2004; Rosenbloom and Sundstrom 2004) have used birthplace information to infer internal migration, but this method cannot be used to study immigrants' mobility within the United States.

selection into inter-county migration on the basis of occupational rank among the foreign born. There is also no evidence of positive selection of immigrants into inter-county migration on the basis of literacy. I do, however, find that the foreign born were more attached to urban areas than were natives, as indicated by the relationship between initial urban status and the probability of inter-county migration. But this finding is complicated by the fact that it was true of immigrants throughout the Age of Mass Migration, rather than unique to the new immigrants as contemporaries claimed. Also consistent with contemporary claims, I find that later immigrants were more attached to urban areas than were earlier ones; but this increasing attachment was shared by natives, suggesting that it was the product of broader trends in the economy rather than of the changing character of immigration.

Immigrant distribution is determined by the rates and composition of internal migration on the one hand, and on the other by the destination choices of inter-county migrants. I investigate the latter by comparing the destination choices of foreign-born and native-born inter-county migrants. I find that foreign-born inter-county migrants were more likely than observationally similar native-born inter-county migrants from the same county of origin to choose urban destinations. This supports the claim that immigrants were drawn to urban areas to a greater extent than were natives. But this simple result again obscures the fact that it was true throughout the Age of Mass Migration rather than being unique to the new immigrants. Moreover, I find that this difference was at its *lowest* in the early twentieth century. Indeed, the destination choices of observationally similar native- and foreign-born inter-county migrants became increasingly similar over time and were at their *most similar* in the early twentieth century. This result further reinforces the notion that immigrants' attraction to urban areas in the latter portion of the Age of Mass Migration was the product of broader economy-wide trends.

This corpus of evidence leads me to conclude that the prevailing views of immigrant distribution during the Age of Mass Migration, which founded calls for restriction of immigration or for the implementation of distribution policy, were generally either false, oversimplified, or the product of broader national trends. Moreover, the attraction of immigrants to urban areas belies my findings that foreign-born inter-county migrants experienced a smaller unconditional average increase in population density than native-born inter-county migrants, and that they urbanized more slowly than immigrants and the population as a whole, whereas the reverse was true for natives. The internal migration of the foreign born was thus a force acting to moderate urbanization, even as natives' internal migration did the reverse. Nonetheless, I find that there was no meaningful geographic assimilation of immigrants relative to natives at any point during the Age of Mass Migration. That is, at no time did a cohort of immigrants experience any meaningful convergence of

their county-of-residence distribution to that of natives over time in the United States.

Although it would not be possible to study immigrants' internal migration and to reach these conclusions without linked data, the potential pitfalls of such data, especially the danger of errors in linking (Abramitzky et al. 2019b; Bailey et al. 2020), pose the main challenge to my results. I use two approaches to ensure that my results are not driven by issues particular to linked data. First, I test the robustness of my results to linkage methods of various strictness, drawing only conclusions that are robust to the choice of linkage method. In studying trends in migration rates over time and differences in migration rates by nativity, I also take another approach. In particular, I use an alternative method of measuring internal migration that does not require linkage (at the cost of being implementable for only a subset of the population of interest and only for inter-state migration) to estimate the frequency of false matches and to correct my estimates. I emphasize that, notwithstanding any of their limitations, the linked datasets that I construct provide a view of immigrants' internal migration and geographic assimilation during the Age of Mass Migration that simply cannot be duplicated in any other way—there is no other data source that enables the observation of immigrants' internal migration in this period. Thus, even if my conclusions are limited, they make a significant advance over the extremely limited prior knowledge on these important topics.

This paper makes four main contributions. The first is to the literature on internal migration in the United States, and especially to papers focused on the economic history of US internal migration. There have been several efforts to document broad patterns in internal migration through US history (Ferrie 2006a; Hall and Ruggles 2004; Molloy, Smith, and Wozniak 2011; Rosenbloom and Sundstrom 2004), but these studies generally exclude immigrants. In studies focusing on the pre-1940 period, this limitation is necessitated by data limitations that prevent the observation of immigrants' internal migration from unlinked census records. This paper thus complements these studies, building on them by providing information for immigrants during the Age of Mass Migration and comparing their patterns to those of natives. It also complements Zimran's (2021) linked data-based analysis of natives' internal migration in the period 1850–1940, as well as the studies of the rates, selection, and sorting of specific episodes of US internal migration (e.g., Collins and Wanamaker 2015; Hornbeck 2020; Sichko 2020).

Second, this paper contributes to studies on the geographic assimilation and segregation of immigrants in the destination country, both in modern settings, as discussed above, and in historical settings, where prior studies have focused on the settlement patterns of particular groups (Eriksson 2019; Spitzer 2019) and on within-city segregation (Cutler, Glaeser, and Vigdor 2008b; Eriksson and Ward 2019).[8] The settlement

[8]See also Abramitzky and Boustan (2017) and Eriksson and Ward (2020).

patterns as a whole of US immigrants during the Age of Mass Migration, how these evolved over time, and crucially how these evolved as immigrants moved within the United States, remain largely unstudied and investigating them is this paper's second contribution.

Third and relatedly, by studying how immigrants' places of residence evolved with time in the United States—their geographic assimilation—this paper adds to the broader literature on immigrant assimilation in both modern settings and in the Age of Mass Migration. The focus of both of these facets of the literature tends to be on assimilation in economic (e.g., Abramitzky, Boustan, and Eriksson 2014; Card 2005; Collins and Zimran 2020; Lubotsky 2007) and cultural (e.g., Abramitzky, Boustan, and Eriksson 2020; Bleakley and Chin 2004, 2010; Ward 2019) terms. Difficulties in tracking immigrants' place of residence over time, especially in the Age of Mass Migration, have left geographic assimilation less clearly understood than assimilation along other dimensions. This paper's evidence in this regard thus expands on the literature's understanding of immigrant assimilation.

Finally, although there have been recent studies of specific immigrant distribution movements during the Age of Mass Migration (e.g., Aaronson, Davis, and Schulze 2020; Abramitzky, Boustan, and Connor 2020), the focus of these studies on organized efforts to encourage distribution necessarily limits the scope of analysis. A complete picture, such as this paper provides, of immigrant distribution over the entirety of the Age of Mass Migration, is lacking. In light of the importance of immigrant distribution in the historical debate and in modern immigration policy debates and the economics of migration, this paper gives an important and entirely new perspective on the debate that led to one of the most important policy changes in US history—the closing of the Golden Door—and ultimately contributes to painting a clearer picture of the process by which the immigrants of the Age of Mass Migration became Americans.

2 Background

In this section, I argue that immigrant distribution was a fundamental issue in the early twentieth-century US immigration policy debate. In particular, I argue that assumptions about immigrants' geographic mobility played a key role in determining views on the effects of and appropriate policy responses to immigration and that these assumptions have not yet been rigorously evaluated. In making this argument, I rely on modern historians, especially Benton-Cohen's (2018) history of the Dillingham Commission, and on a number of historical sources, which provide illustrative quotations. These are provided in detail in Appendix B, and

are indicated by roman numeral endnotes.[9]

The changing quantity and character of immigration in the late nineteenth century arising from the change in the chief source countries of US immigrants sparked a renewed debate regarding European immigration in the United States (Benton-Cohen 2018; Daniels 2004; 1991 [2002]; Goldin 1994; Hatton and Williamson 1998).[10] Immigrant distribution was a central issue in this debate, both in the positive sense—perceptions of where immigrants lived and where they did or did not move affected beliefs regarding the effects of and ideal policy responses to immigration—and in the normative sense—the broader set of views on immigration included whether immigrants should have moved within the United States, whether they should be encouraged to move, and what impacts their moving would have. From a positive perspective, it was widely held that the new immigrants were immobile, especially in comparison with the old immigrants. In particular, the old immigrants were seen as having assimilated geographically into the country, going to the agricultural areas of the midwest, whereas the new immigrants were seen as loathe to leave the major coastal cities where they had arrived.[i] Immigrants' attachment to urban areas was also held to be greater than was that of natives.[ii] A number of causes were cited for this perceived immobility, including a lack of information regarding employment opportunities outside of the major cities,[iii] network effects,[iv] or even perceived physical and mental inferiority to prior cohorts of immigrants.[v]

A number of policy prescriptions grew out of these positive views of immigrant distribution, combined with normative views on the matter. The strongest advocates of immigration restriction in government, including President Theodore Roosevelt, and in the public saw immigrant immobility as yet more evidence of the ills of immigration and of immigrants' resistance to assimilation,[vi] leading them to double down on restriction.[vii] They supported short-term distribution efforts to alleviate the immediate (perceived) negative effects of immigration,[viii] but in the longer run, they were concerned that distribution would spread the supposed ills of immigration to other parts of the country.[ix] Ultimately, though they faulted immigrants for failing to distribute, proponents of restriction argued that distribution was no substitute for restriction of immigration to ensure that only the "right" immigrants arrived.[x]

But the doubts of advocates of restriction belie the optimism that many observers and officials, including some members of the Dillingham Commission, had for distribution (Benton-Cohen 2018).[xi] They saw in the positive facts on immigrant distribution a normative prescription to encourage distribution.[xii] Many even

[9] Willcox (1906, p. 523–524) also compiles a list of contemporary quotations documenting views on immigrant distribution.

[10] This was, of course, not the first immigration policy debate in the United States. In the 1850s, the mass inflow of Irish immigrants led to the growth of the nativist Know-Nothing party (Alsan, Eriksson, and Niemesh 2020). But European immigration was replaced by slavery, Reconstruction, and Asian immigration, among other issues, at the center of public discourse after the mid 1850s, and resurfaced in earnest only in the 1890s.

saw distribution as a way to mitigate, even in the longer run, perceived negative effects of immigration,[xiii] and thus as a way to avoid outright restriction (Benton-Cohen 2018, ch. 3). Many, in fact, saw the problems of immigration as being solvable *entirely* by distribution,[xiv] in part because it provided a way to ensure that immigration "did not undermine the American standard of living" (Benton-Cohen 2018, p. 230).[11] Distribution was also favored by regions that did not attract many immigrants but sought to do so.[xv] In the South, immigrant distribution was seen as "a way to populate [an] undeveloped region, solve the 'race problem,' and stave off new expansions of federal power" (Benton-Cohen 2018, p. 200).[12] Business interests in the West also viewed distribution as a potential source of labor (Benton-Cohen 2018, p. 83).

Across the political spectrum, concerns about immigrant distribution were often linked to urban reform movements, and many of the progressives who were concerned about immigration were also noted anti-urbanists (Benton-Cohen 2018). Indeed, many of the arguments that were raised in opposition to immigration—that it would bring disease and crime and swell urban slums—were inherently urban problems. Moving immigrants out of cities would therefore blunt these ill effects.[13]

On the whole, "Federal distribution efforts were popular in Congress because they appealed to urban immigrant constituencies [because they might dampen calls for restriction] and to western and southern business interests" (Benton-Cohen 2018, p. 83). This enthusiasm led to the creation, by the Immigration Act of 1907, of the Division of Information to provide information on opportunities in the US interior (Benton-Cohen 2018, pp. 82–83),[14] and to the Dillingham Commission's support for the strengthening of distribution efforts.[xvi] A number of extra-governmental programs encouraging distribution were also created, and were embraced with the greatest enthusiasm by leaders of the American Jewish community (Benton-Cohen 2018, pp. 73, 84).[15] These programs included the Industrial Removal Office (Abramitzky, Boustan, and Connor 2020) and the Galveston Movement (Aaronson, Davis, and Schulze 2020) among Jews and the Sunnyside Plantation among Italians (Benton-Cohen 2018, ch. 7), but all of these failed, in some cases leading some of

[11] Maintaining the "American standard of living" was seen as an important goal by the Dillingham Commission, and by President Roosevelt, who argued that immigrants' "standard of living should be such that [they] will not, by pressure of competition, lower the standard of living of our own wage-workers; for it must ever be a prime object of our legislation to keep high their standard of living" (Roosevelt 1905).

[12] Whether immigrants were seen as making a positive contribution to the racial and cultural environment of the United States was far from simple. The eugenic view of immigrants, though present for some time already by the turn of the century, did not come to dominate immigration debates until later (Benton-Cohen 2018, p. 238). Southerners also ranged from viewing immigrants as a potential force to displace blacks (a positive aspect, in these eyes), to seeing them as problematic because they were not perceived as white (Benton-Cohen 2018, pp. 209–210). An article in the *Farmer's Union News* in Union City, Georgia in February 1909 was concerned about "this present influx of brownish alien races" (in US Congress 1911c, p. 363).

[13] Distribution movements also had a goal of sparking a "rural revival" (Benton-Cohen 2018, pp. 83–85).

[14] The Division of Information replaced public immigration bureaus in some southern states that sought to recruit laborers and farmers, but which were rendered illegal by the Immigration Act of 1907 (Benton-Cohen 2018, p. 83).

[15] Benton-Cohen (2018, p. 83) explains that "Distribution [was] largely a defensive measure against prejudice [against Jews] fueled by disgust with urban overcrowding" but that "some Jewish leaders wondered if distribution schemes, by exposing more Americans to improverished, religious Jews, might actually exacerbate discrimination rather than relieve it" (p. 84)

the Dillingham Commission's members to give up on distribution and favor restriction (Benton-Cohen 2018, p. 233). In summary, for those not dogmatically opposed to immigration, "Distribution offered a palatable solution to the inseparable concerns about immigration and rapid urbanization, while also promising to solve labor shortages in the South and West" (Benton-Cohen 2018, p. 83).

Willcox (1906), however, realized that the calls for distribution and complaints about immigrant mobility were based on a fundamental misunderstanding and miscontextualization of immigrants' internal migration patterns in the United States.[16] He summarized the argument in favor of distribution in five points, which I paraphrase here: (1) immigrants were more likely than natives to live in cities and less likely to leave cities; (2) this tendency was greater among the new immigrants than the old; (3) this tendency was greater among illiterate than literate immigrants; (4) this tendency had negative effects on the immigrants and on the US society and economy; and (5) governmental or private action was necessary to reverse this tendency. Using a variety of data from the published volumes of the 1890 and 1900 censuses, Willcox (1906) challenged these claims. Although he conceded that immigrants were more urban than natives, he argued that they were not less likely to leave urban areas, and that claims to the contrary failed to consider that immigrants were initially located in major cities on arrival in the United States and that the country had urbanized dramatically over the course of the Age of Mass Migration, which led natives to move to cities.[17]

Working contemporaneously, Willcox (1906) was limited in the nature and quantity of data that he could bring to bear on this question. By focusing on the data in the published volumes of the census, as the constraints of the time required, Willcox (1906) was not actually able to directly observe migration,[18] but instead was forced to deduce it from changes in population counts by place of birth. That is, he did not observe gross migration flows—simply each county or city's net migration by nativity, which cannot separate internal migration, return migration of immigrants to Europe, and the arrival of new immigrants.[19] Thus, while Willcox's (1906) analysis casts strong doubt on contemporary views of immigrant distribution, it is not dispositive. Indeed, I will show that some of his claims, such as that there was "no evidence on [immigrants']

[16]This is not to say that, had the proponents of distribution accepted these arguments, they would not have insisted that immigration distribution was necessary. Indeed, the anti-urbanists would likely have cited the contribution of immigrants to city growth, which I show below to simply be the product of the urbanization of the economy, as reason enough to restrict urbanization. Willcox's (1906) analysis, like the present paper, simply serves to show that the assumptions on which the policy prescriptions were made were not accurate.

[17]He did not directly engage with the fourth and fifth points, arguing that by disproving the first three he had *a fortiori* disproved the latter two. As Willcox (1906, pp. 531–532) put it, "the swarms of immigrants found at any time in New York are no more conclusive evidence of a tendency to remain there than the clouds hanging around a mountain are proof that there is no wind at the summit to blow it away." Willcox (1906, p. 537) also pointed out that the view of immobility arose from the fact that "nine-tenths of [immigrants] arrive in cities and that it takes them a long time to disperse from these centres."

[18]See Zimran (2021) for more discussion of this issue. A Department of Commerce Report in 1933 pointed out that "While the census figures for State of birth form practically the only source of extensive information with regard to interstate movements of population, extreme care should be exercised in the use of these figures as representing or measuring migration . . . It is therefore with some hesitancy that the term 'migration' is used at all . . ." (Department of Commerce 1933, p. 135).

[19]The latter concern can be addressed by using the census's information on immigrants' year of arrival.

part of a tendency to cling to or stagnate in the cities of the country" (Willcox 1906, p. 532), that "recent immigrants [did not] show any characteristic tendency towards or fondness for city life" (Willcox 1906, p. 544), and that there was no "evidence that the tendency of the foreign born towards urban life [was] any stronger or weaker than the tendency of natives" (Willcox 1906, p. 537), may have swung too far in the opposite direction of the contemporary conventional wisdom.

Modern studies have used census microdata to delve more deeply into historical immigrants' internal migration, though with the exception of Aaronson, Davis, and Schulze's (2020) and Abramitzky, Boustan, and Connor's (2020) studies of specific Jewish distribution efforts (the Galveston Movement and the Industrial Removal Office, respectively), there has not been any application of these data to understanding immigrant distribution in the Age of Mass Migration.[20] One common way in which such data are used to study internal migration in the United States is to use the census's information on place of birth, which can be used to infer migration over the lifespan (e.g., Ferrie 1997, 2006a, b; Hall and Ruggles 2004). However, this method cannot be used to study immigrants' internal migration because their birthplace information is not informative of prior place of residence in the United States.[21] An alternative method, used by Rosenbloom and Sundstrom (2004) to measure internal migration, can be applied to immigrants, and I use a variant of this method below to generate alternate estimates of internal migration rates. This method uses census information on children's ages and birthplaces to infer parents' migration. Unlike the birth state-residence state comparison method, this approach does enable the observation of immigrants' internal migration. Using this method, Rosenbloom and Sundstrom (2004) find a negligible difference in the migration rates of the native and foreign born. But a number of limitations of this method constrain these conclusions. First, it applies only to families with children, whereas most immigrants were unaccompanied men (Barde, Carter, and Sutch 2006c). Second, because the US census reports only state of birth (for those born in the United States) and not birthplace at a more disaggregated level, this approach enables observation only of inter-state migration. Intra-state inter-county migration, which is particularly important, for instance, in the case of rural-to-urban migration (Ferrie 2005), cannot be observed. Finally, because the information on migration is retrospective, no pre-migration characteristics are observed, limiting the ability to study selection into internal migration.

To most effectively measure and study immigrants' internal migration, it must be directly observed by determining an individual's place of residence and characteristics at two points in time. This is possible

[20] Eriksson (2019) and Spitzer (2019) have used census microdata to study the settlement patterns of Norwegian and Russian Jewish immigrants, respectively, but do not study internal migration.

[21] Of course, immigrants not living in ports of entry were very likely to be internal migrants, but even those living in ports of entry may have been internal migrants, either by moving from some other port or by moving inland and then returning to the ports. Only linked data can capture such movements.

only with linked census data. Such data can be constructed for individuals regardless of marital status or nativity,[22] and the combination of data from the two linked censuses enables the observation of pre-migration characteristics and of intra-state moves. The panel perspective that the linked census data provide are also important in studying all types of immigrant assimilation (Abramitzky, Boustan, and Eriksson 2014; Lubotsky 2007). In studying geographic assimilation, it is necessary to distinguish between immigrants living in cities with large immigrant populations because they had recently arrived there, and those who had lived there for some time, as well as to ensure that changes in the share of immigrants living in cities over a cohort's time in the United States is not affected by return migration.

Ferrie (1995, 1999) pioneered the use of such data to study immigrants' internal mobility, giving evidence on the settlement and internal migration patterns of immigrants and natives in the United States prior to 1860. More recently, Biavaschi and Facchini (2020) use the linked census data of Abramitzky, Boustan, and Eriksson (2014) to compare the rates of inter-state migration of the native and foreign born in the early twentieth century.[23] Linked data have also been crucial to the studies of specific distribution movements by Aaronson, Davis, and Schulze (2020) and Abramitzky, Boustan, and Connor (2020).

Because of the limitations of unlinked data and because it has only recently become feasible to make links across all US censuses, the claims of immigrant immobility have not been scrutinized by the lens of modern economic history. To my knowledge, no study exists providing a comprehensive description of immigrants' internal migration during the Age of Mass Migration, nor is there an existing study comparing the internal migration of immigrants and natives or a systematic study of the distribution question in this period.[24] This important aspect of immigrant assimilation in the historic United States is thus poorly understood.

3 Data

My analysis relies on complete-count identified census data for the US censuses of 1850–1930, compiled by Ruggles et al. (2019). Although none of these censuses inquired directly about internal migration, they all provide information on place of residence, making it possible to determine whether an individual migrated between two census years by linking him between the censuses of those years and comparing his places of residence in the two. The individual's characteristics in the first census also provide pre-(internal) migration

[22] But they cannot be constructed for women, who can be studied with the unlinked approaches. And of course, not everyone can be linked, though I will argue below that the linked sample is largely representative of the population of interest.

[23] Biavaschi and Facchini (2020) do not address the possibility that their results are driven by linkage errors, which I show below to be an important issue.

[24] Immigrants are included in a number of studies of historic US internal migration with linked data (e.g., Stewart 2006, 2009), but are treated as simply another set of observations. No comparisons are made to natives.

characteristics.[25] Aside from specific linkage issues, described below, the chief limitation of these data is that, by design, they exclude individuals who lived in the United States in one census, but returned to their home countries before the second. This is a necessary limitation in any study of assimilation, as Abramitzky, Boustan, and Eriksson (2014) show, and is required in this setting because, as discussed above, immigrants' internal migration cannot be observed without linkage. But it has implications for the proper contextualization of my results insofar as the concerns of advocates of immigration restriction were based on immigrants who arrived in cities and remained there before returning to their home countries.[xvii] These data would not capture this form of immobility. Nonetheless, the data do enable the description of the distribution, mobility, and geographic assimilation of permanent immigrants.

3.1 Linkage

I constructed my linked data sets as follows. Beginning with all native- and European-born males in each census 1850–1910 (other than 1890), I attempted to make a link to the records of the census taken 10 and 20 years later, where that census is available (i.e., not 1890). I make links over both 10- and 20-year spans so that I can comparably cover the period 1880–1900, which is complicated by the lack of microdata from the 1890 census. In total, this approach resulted in 10 distinct census-to-census linked datasets. The beginning of my study period is determined by the availability of census data—the 1850 census was the first to include information, including birthplace, on all US residents, and so it is not possible to begin the analysis earlier. I end the study with linked samples beginning in 1910 (and ending in 1920 or 1930) because this is the last census taken before World War I and the imposition of the literacy test in 1917 fundamentally altered US immigration. Although there were large immigrant inflows in 1919, 1920, and 1921, there is evidence of fundamental changes in the character of immigration (Spitzer and Zimran 2018), its gender composition (Barde, Carter, and Sutch 2006c), and immigrant behavior (Ward 2017) relative to the pre-War period. The 1850–1910 coverage of initial censuses concentrates on the period of essentially unrestricted migration.[26]

I use three primary methods to create the linked data sets. The first is a method developed and implemented by Collins and Zimran (2019, 2020) and laid out in detail in Appendix A. Following the example of Ferrie's (1996) pioneering matching approach, and based on best practices outlined by Abramitzky et al. (2019b) and Bailey et al. (2020), this method makes links on the basis of fixed and pre-determined rules, uses string distance measures to make comparisons of names (to address potential differences in spelling across

[25] Characteristics in the second census also provide post-migration characteristics, but these are not a focus of my analysis.

[26] This does lead to missing the very last set of immigrants arriving 1911–1914, but it would not be possible to have even a first observation of these individuals until 1920 and later, by which time the Age of Mass Migration had effectively ended.

censuses), and removes cases where individual records are not unique. I refer to this linkage method as the *CZ* method. As has become typical in papers using linked data, I also use records linked by other methods. In particular, I use two approaches described by Abramitzky et al. (2019b) and Abramitzky, Boustan, and Rashid (2020)—ABE-NYSIIS with Robustness and ABE-Exact with Robustness, which I refer to as the *ABEN* and *ABEE* methods, respectively. Like the CZ method, these approaches make links on the basis of fixed and pre-determined rules, but do not use string distance measures.[27]

Recent research has come to recognize two primary difficulties in the use of linked data. The first, highlighted prominently by Bailey et al. (2020), is that of *false positives*—that is, that the linked data may not actually identify the same individual at each point in time. This is a particularly severe problem in studying internal migration because a false match will almost always appear as an instance of inter-county migration. I use two strategies to address this difficulty. First, I draw only conclusions that are qualitatively robust to the choice of linkage method, as well as to the use of two stricter linkage methods. The first such stricter method makes a link only in cases where all three of the methods described above agree. I call this the *intersection-of-matches method*, and refer to it in brief as the *Int* method. This method is founded on Abramitzky et al.'s (2019b) finding that using the intersection of different sets of matches reduces the probability of a false match, at the expense of a lower match rate and potentially less representative sample. I also use an even stricter set of matches than this, which, in addition to requiring an intersection of matches (which in practice requires an exact match of names between records because this is a linkage criterion of the ABEE method),[28] also requires an absolute difference in age-implied birth years between records of no more than one year.[29] I refer to this as the *Exact* linkage method.

In studying trends in internal migration rates over time and in making comparisons across nativities, I use an additional method to protect agains spurious conclusions. In particular, I use an alternative non-linkage-based method similar to that of Rosenbloom and Sundstrom (2004) to compute an alternative estimate of inter-state migration based on the birthplaces and ages of children in the household. By comparing inter-state migration rates from the linked samples to those arising from this method, I can estimate a rate of

[27] The data provided by Abramitzky, Boustan, and Rashid (2020) include links only by the ABEE and ABEN methods. The CZ-method links that I use thus complement these links. I do not use the links provided by Abramitzky, Boustan, and Rashid (2020) directly; I implement these methods using code provided by the authors.

[28] Although the Int method requires an exact match on name, the difference from the ABEE method is that the record must be unique not only on the basis of exact name, but also on the basis of the name's NYSIIS code and on the basis of potential string distance matches to the name. This is thus a much more conservative match criterion.

[29] The CZ method permits links with up to a four-year difference in age-implied birthyear to accommodate age heaping. The ABEE and ABEN method allows links with up to a two-year difference, which implies that the Int method also allows links with up to a two-year difference. In general, widening the permitted age difference is a double-edged sword, in that it creates more permissible matches, but also raises the threshold for uniqueness, which is required by all 5 linkage methods. The Int and Exact methods are conservative in that they permit matches only in cases of greater similarity, but also require less similarity between records to exclude them on the basis of non-uniqueness.

false matching in linkage and use this estimate to adjust my estimated rates of internal migration. I develop this method in further detail below and in Appendix A (see also Zimran 2021).

The second difficulty in using linked data is that linked samples may not be representative of the population of interest. Table 1 shows the rates of successful linkage for each linkage method for immigrants in my samples.[30] In this table and throughout the paper, I make the following sample limitations. First, the sample is limited to white men aged 18–40 in the initial year. The gender restriction is imposed because the changing of women's names at marriage poses a significant obstacle to linking their records, and so I do not attempt to do so. The age restriction is imposed in order to ensure that men are initially observed as adults—so that they are less likely to move as part of their parents' internal migration and so that they are more likely to be in the labor market, which provides occupational information needed for the study of selection into migration—and yet are not too old after linkage for mortality to be a major concern. Second, I limit immigrants to those born in Europe, as this is the main group of interest in the literature on immigration during the period in question (Abramitzky and Boustan 2017; Hatton and Ward 2019).[31] The sample thus includes individuals born in the United States or Europe. Finally, I limit the sample to whites in order to maintain comparability of natives to immigrants.[32] The first row of each grouping in Table 1 shows the number of individuals who were at risk for linkage. As I explain below, it is desirable, to the extent possible, to make the linked sample representative of the population in the second period of each span, and in the case of immigrants to limit this to those that had arrived by the initial year. This can only be done for samples ending 1900–1930 because these are the only available censuses that report immigrants' arrival year. Thus, the number of individuals at risk for linkage for spans ending 1900–1930 is based on the final year and for other spans this figure is based on the initial year. This is the reason for the difference in the number of individuals at risk for linkage across methods in the same span.[33] The second row of each grouping shows the number of successful links, and the third row shows the implied fraction of successful links.

The linkage rates for each of the three main methods range from about 6 percent in the nineteenth century to about 9 percent by the end of the study period for the foreign born (Table 1), and from about 10 to 19 percent for the native born (Online Appendix Table C.1). As these figures suggest, linkage rates

[30] Online Appendix Table C.1 presents analogous figures for the native born. I do not include these in the main text because they largely mirror Zimran's (2021) results, where they are discussed in more detail.

[31] The Dillingham Commission and the debate over immigration policy around the turn of the twentieth century also focused exclusively on European immigration (Benton-Cohen 2018, p. 7). The continued exclusion of Asian immigrants was also an issue. There was no focus on immigration from Mexico or Canada

[32] I limit the sample to those reporting a white race in the initial year. Because I do not link on the basis of race, there will be some links to individuals with a non-white race in the second year; these are retained.

[33] Part of the difference between the CZ and the other methods is that the former allows a match with up to 4 years' difference in age-implied birth years, whereas the other methods permit only a two-year gap, and the Exact method permits only a one-year age difference, thus changing the sample at risk for linkage to an 18–40 year old in the initial census.

increase over time and are greater for the native born than for the foreign born. As expected, the successful linkage rates are lower for the two stricter methods, ranging from only about 1 to 3 percent for the foreign born, and from about 5 to 9 percent for the native born, again increasing over time and greater for natives than immigrants. Although these linkage rates are typical (as evidenced in part by their similarity across the three methods), the fact that less than one-fifth, and in some cases less than even 5 percent of the population of interest can be linked, raises clear concerns regarding the representativeness of the resulting samples.

Figure 1 compares the linked sample of foreign-born men to the samples at risk for linkage (as defined in the next paragraph) on the basis of all of the observable characteristics available in each census year. These figures present the ratios of the mean of each observable characteristic in the linked sample to that in the sample at risk for linkage; a value of one indicates perfect representativeness on the basis of a particular characteristic. As is typical in studies using linked data, the linked samples over-represent individuals of higher socioeconomic status and under-represent urbanites, consistent with linkage favoring individuals who would be better able to consistently report accurate information. To address concerns regarding representativeness on the basis of observables, I reweight each sample to match (as closely as possible) the sample at risk for linkage on the basis of the available observables. Specifically, I estimate probit regressions for the probability of a successful link as a function of name frequency and length (as in Feigenbaum and Tan 2019) and all observables available in the initial census of the linkage span, which are presented in Table 2, and use the estimated conditional linkage probabilities to construct inverse probability weights.[34] The weighting method cannot address selection into linkage on the basis of unobservables (e.g., Antonie et al. 2018)—a limitation that must be kept in mind in all studies using linked data.

In reweighting, it is desirable to the extent possible to match the observables of the linked sample to those of the population in the second census of the span.[35] This is because the population of immigrants in a particular initial census year includes individuals who would return to their home countries before the final census. This was a relatively rare phenomenon earlier in the Age of Mass Migration, but became increasingly common among the later waves of immigrants (Bandiera, Rasul, and Viarengo 2013). Returners are likely to have been different from the persisters, who are the only ones at risk for linkage but who cannot be identified ex ante. To avoid including returners in the set of individuals at risk for linkage, I compare the observables

[34] That is, the weights are $\frac{1}{\hat{p}}$, where $\hat{p}$ is the estimated conditional linkage probability for each successfully linked observation. I include continuous variables in these regressions as higher ordinary polynomials for flexibility. In each sample, I truncate the linkage probability from below at the 0.5 percentile to avoid extreme weights dominating the analysis. A consequence of this weighting is that individuals for whom any variable used to construct weights is not available are excluded. Most consequentially, this implies that anyone without a reported occupation (i.e., anyone with occ1950 code above 970) will not be included in the analysis. In 1900–1930 there are some individuals with an occupation listed as 979, "Not Yet Classified," whom I omit from the main results. I discuss robustness to this sample limitation below and in Online Appendix E.2.

[35] For instance, in linking 1900–1920, I reweight the sample to match the distribution of observables in 1920 rather than 1900.

of linked individuals to those of the population at risk for linkage (based on age) in the latter census. In the case of immigrants, I include in the sample at risk for linkage only individuals reporting an arrival year of the first census or earlier (to avoid including newer immigrants in the sample at risk for linkage);[36] the samples of individuals at risk for linkage in Table 1, Online Appendix Table C.1, Figure 1 and Online Appendix Figure C.1 are determined in this way. Because arrival year is available only in the censuses of 1900–1930, this approach can be applied only for spans ending in these years; for other spans, the population at risk for linkage is based on the initial census of the span. The lower rates of return migration earlier in the Age of Mass Migration reduce substantially the danger of failing to do this in the earlier samples.

3.2 Defining Migration

The finest unit of place of residence that can be observed in all censuses is the county, which motivates my focus on inter-county migration. With an individual's county of residence reported in the census, it is, in principle, straightforward to determine if an individual made an inter-county move between two censuses. A difficulty arises, however, because of changes in county boundaries between two censuses, which are documented by Manson et al. (2019). Consider, for instance, the example presented in Figure 2. This figure depicts several counties in northeastern Illinois in 1850 and in 1870 that experienced boundary changes in this span. In 1850, points A and B on the one hand and C and D on the other are each in the same county and cannot be distinguished from one another (because only the county of residence is known). Similarly, in 1870, points B and C are in the same county and cannot be distinguished from one another. A move from point A to C, which crosses county boundaries as defined in either year, thus cannot be considered an example of internal migration because it cannot be differentiated from a move from point A to B, from B to C, or from continuous residence at point B. This issue arises whenever there is any overlap between an individual's residence county in two census years. Thus, only moves between counties with no geographic overlap—in the example of Figure 2, moves from A or B to D or from C or D to A—are considered inter-county migration.[37]

3.3 Measures of Migrant Selection

To evaluate claims of positive selection into internal migration, a measure of socioeconomic status is required. Selection is typically measured on the dimensions of wages or years of education (e.g., Chiquiar and Hanson

[36] In cases where a link is made to a person whose reported year of arrival is after the initial census, I retain the link.

[37] Due to slight variations in the drawing of county boundaries by Manson et al. (2019) in different years, a more nuanced definition is required to avoid seeing an overlap in two non-overlapping counties simply because of a drawing imprecision. I consider an overlap to occur only if the overlapping area is at least 0.5 percent of at least one county's area.

2005; McKenzie and Rapoport 2010). Both of these measures became available in the US census beginning in 1940. Instead, I use two measures that are consistently available throughout my study period. As a measure of education, I use the indicator of literacy available in each census. The only indication of economic status that is available in every census that I use is occupation. Following the literature on immigrant assimilation in the United States (Abramitzky, Boustan, and Eriksson 2014; Collins and Zimran 2019, 2020), I use the 3-digit occupational codes created by Ruggles et al. (2019) to create occupational scores.[38] My preferred measure is constructed as follows. First, I implement three occupational scoring systems. One is based on the average wealth holdings in the 1870 census by individuals with a particular occupation, following the method of Collins and Zimran (2019, 2020). The second is based on Preston and Haines's (1991) estimated average income by occupation in 1900 with accommodations for farmers (Collins and Wanamaker 2020; Collins and Zimran 2020). The last is the 1950-based occscore variable provided by Ruggles et al. (2019). For each of these scoring methods, I determine the rank of each occupation in the population of employed white men aged 18–74 in each census. My preferred occupational status measure is the simple average of these three ranks. As with all such occupational status measures, this measure is limited in that it captures differences across but not within occupations (Inwood, Minns, and Summerfield 2019; Saavedra and Twinam 2020). Nonetheless, the value of such measures as the best available indicator of an individual's economic status in historical contexts is broadly recognized. I also study selection into migration on the basis of urban residence to better investigate claims that immigrants were attached to urban areas. My benchmark results use the standard census definition of an urban place as one with at least 2,500 inhabitants. In Online Appendix E, I explore the consequences for my results of using alternative definitions of urban residence. I discuss the details of this and other robustness checks below.

4 Inter-County Migration Rates

4.1 Trends Over Time

Figure 3 presents estimated rates of inter-county migration—simply the estimated probability that a foreign-born individual is observed in two different counties in the initial and final census of each linked sample—for the foreign born for 10- and 20-year spans and for each linkage method. For the most part, the estimates for each of the three simple methods are similar to one another (though the ABEE method shows somewhat

[38] I also create indicators for fitting into one of five occupational categories—farmer, white collar, craft, operative, and unskilled. I do not analyze these as primary variables of interest, but they are included in creating linkage weights and in making conditional comparisons between immigrants and natives.

lower migration in the early twentieth century), as are those for each of the two stricter methods. But, consistent with false matches inducing spurious observations of internal migration (and also consistent with the stricter methods over-representing less mobile types), the estimated rates of internal migration according to the stricter methods are substantially lower than those according to the more permissive methods.

Despite this difference, the trends in internal migration rates over time as implied by each method are broadly similar, showing a steady decline in the nineteenth century before stabilizing at a lower level in the twentieth, consistent with the claims of reduced mobility over time. According to the more permissive linkage methods, the 10-year internal migration rates declined from about 70 percent in the 1850s to about 50–55 percent in the twentieth century. According to the stricter linkage methods, the decline is from about 50 percent to about 35 percent. Predictably, the estimated migration rates for the 20-year spans are greater than for 10-year spans, but show similar declines over time. These declines are crucial to comparing the mobility of the new and old immigrants. But just as the difference between the migration rates implied by the different linkage methods is suggestive of false matches generating spurious migration, it is possible that declining rates of false linkage over time, perhaps due to improvements in the quality of the census, are responsible for spurious declines in the estimated internal migration rates.

To determine the extent to which these estimates are biased by false matches, I use methods described by Zimran (2021) and in Appendix A to estimate the rate of false linkage for each linkage method and span and to use these estimates to correct the estimated inter-county migration rates. In brief, for each 10-year linkage span and method, I identify a subsample of the linked data set for which it is possible to use the birthplaces and ages of children to determine the individual's inter-state migration status over the 10-year span (a process similar to that used by Rosenbloom and Sundstrom 2004). For this (unrepresentative) subset, comparison of the inter-state migration rates by the linkage method and by the household-composition method enables the estimation of a rate of false matching by Bayes Theorem. Estimated false match rates can be computed by this strategy only for 10-year spans because the household composition method cannot be used for 20-year spans due to the small likelihood that children would reside with their parents for 20 years or more.

Figure 4 plots the estimated false match rate for each linkage method and 10-year span. As expected, the estimated false match rates are considerably greater for the more permissive linkage methods than for the stricter methods, and decline over time. For the more permissive methods, the estimated false match rates decline from about 40 percent in the nineteenth century to about 20 percent in the twentieth. For the stricter methods, the decline is from about 15 percent to about 5 percent.

Figure 5 shows the estimated rates of inter-county migration after correcting the estimates of Figure 3 for

Figure 4's estimated rate of false matches according to another procedure described by Zimran (2021) and in Appendix A. In essence, this method uses Bayes's Theorem, the estimated rate of false linkage, and the probability that a false match creates spurious inter-county migration (essentially one) to correct estimated inter-county migration rates. Panel (a) of Figure 5 covers 10-year spans and panel (b) covers 20-year spans. Because I cannot use the method above to estimate false match rates for 20-year spans, I use the false match rate from the 10-year span beginning in the same year to correct these estimates. The 20-year spans do not include estimates for the 1880–1900 span because there is no estimate of the rate of false matching for periods beginning 1880. In Figure 5, there are still differences in the estimates coming from each linkage method, but they are more similar than before correction, and, as shown in Figure 6, which standardizes the migration rate for the last span beginning in the nineteenth century (i.e., 1870–1880 for 10-year spans and 1860–1880 for 20-year spans), show substantially similar trends over time.

The estimates of Figures 5 and 6 yield my first main result, regarding the change in internal migration rates of the foreign born over time. According to all linkage methods, there was a large decline in immigrants' internal mobility from spans beginning in 1850 to spans beginning in 1860. There was then a further, but smaller, decline in the 10-year rates from the span beginning in 1860 to that beginning in 1870. From the last spans of the nineteenth century to the first of the twentieth, the change in internal migration rates varies by linkage method; but, crucially, there is no indication of a decline. All methods then show a slight *increase* in internal migration rates through the twentieth century.

In the strictest sense, the decline in the internal migration rates of the foreign born from the beginning of the study period to the latter part of the Age of Mass Migration confirms the claims that immigrants became increasingly immobile over the Age of Mass Migration. But this claim is complicated by the fact that the decline occurred almost entirely from spans beginning 1850 to spans beginning 1860, and entirely in the nineteenth century. Indeed, once the linkage spans beginning in 1850 are removed from consideration, there is essentially no evidence of a decline. This result provides important context for the decline in immigrants' internal mobility. Whereas a substantial number of immigrants did arrive in the United States prior to 1850, the greatest surge in immigration, even of the old immigrants, occurred after 1850 (Barde, Carter, and Sutch 2006b; Haines and Sutch 2006). This implies that the highly mobile group of immigrants first observed in 1850 did not include the mass waves of old immigrants from Britain, Ireland, and Germany in the 1850s. Thus, the view of declining immigrant mobility, while technically true, is severely oversimplified. It was not that the new immigrants were less mobile than the old, but that all immigrants, including the old immigrants arriving in the 1850s, were less mobile than the very first cohort that arrived before 1850.

These results suggest that the very first wave of immigrants in the Age of Mass Migration held a substantial, and perhaps undue, influence in shaping views of the immigrant experience in the United States. In this sense, my results are consistent with those of Collins and Zimran (2020) in that they establish that certain "myths" of US immigration—high economic and geographic mobility—were true for the first wave of immigrants in the Age of Mass Migration and not thereafter, and yet remain persistent.[39]

4.2 Comparison to Natives

Was the nineteenth-century decline in internal mobility unique to immigrants? Zimran (2021) shows that natives' internal migration rates were largely constant over time with evidence of a slight increase from the nineteenth to the twentieth centuries, implying that the decline in internal migration rates in the nineteenth century was unique to immigrants.[40]

Were immigrants relatively mobile or relatively immobile in the broader context of the US economy? Figure 7 compares estimated inter-county migration rates of immigrants and natives. Panels (a) and (b) show the differences in uncorrected migration rates (based on the results presented in Figure 3 and Online Appendix Figure C.2). Panels (c) and (d) present differences in these rates after correcting, as above, for errors in linkage (based on estimates presented in Figure 5 and Online Appendix Figure C.4). After correction, a clear premium in internal migration rates of about 15–20 percentage points for the foreign born is evident in linkage spans beginning 1850 before declining considerably, contemporaneously with the decline in immigrants' internal migration rates documented above. From the nineteenth to the twentieth century, the premium is roughly constant before rising somewhat for spans beginning 1910. Although the range of estimates of the difference in internal migration rates in each span differs by linkage method by about 5–10 percentage points, even after correction, it is notable that there is no evidence that immigrants were meaningfully less mobile than natives during the Age of Mass Migration. The stricter linkage methods do indicate slightly lower migration rates for immigrants than natives for spans beginning 1900, but even there the difference is never larger than 3 percentage points.

To delve further into these comparisons, I compare the internal migration rates of immigrants and natives after controlling for all covariates available in each census. That is, I estimate an equation of the form

$$y_{it} = \beta_t f_{it} + \mathbf{x}'_{it}\delta_t + \varepsilon_{it},$$

[39] Similarly, Long and Ferrie (2013) show that related myths of US intergenerational mobility were also true in the early years of the nation, but ceased to be (c.f., Ward 2020).

[40] See also Online Appendix C, which repeats the analysis above for natives. I do not include these results in the main text because they are very similar to those of Zimran (2021), who discusses them in detail.

where y_{it} is an indicator equal to one if individual i made an inter-county move in span t and zero otherwise, f_{it} is an indicator for foreign birth, and $\mathbf{x}_{it}$ is a vector of controls, including fixed effects for initial county of residence.[41] Panels (e) and (f) of Figure 7 report the coefficients β_t for each linkage method and span.[42] Because the correction method used above cannot be applied here, as it is not possible to determine which individuals are falsely matched, these estimates are not corrected for false matches. However, insofar as the probability of making a false match varies according to the observables included in $\mathbf{x}_{it}$, the danger is reduced relative to panels (a) and (b). The decline in the immigrant-native difference in internal migration rates in the nineteenth century is moderated relative to the unconditional results, and there remains no evidence of lower mobility by immigrants than natives.

5 Migrant Selection

Beyond the general, and evidently misplaced, concerns that immigrants were immobile, immigrants living in urban areas were seen as being especially unlikely to move. Moreover, as pointed out by Willcox (1906), illiterate immigrants were thought to be particularly immobile; that is, it was claimed that foreign-born internal migrants were positively selected from among the foreign-born population as a whole on this dimension. The availability of pre-migration characteristics in the linked data enables me to assess these claims directly and to expand the analysis to measure selection on the basis of occupational status.

To measure migrant selection, I estimate, separately for each linked sample, an equation of the form

$$y_{it} = \alpha_t u_{it} + \beta_t \ell_{it} + \gamma_t r_{it} + \mathbf{x}'_{it}\theta_t + \varepsilon_{it},$$

where y_{it} is an indicator equal to one if individual i made an inter-county move in span t and zero otherwise, u_{it} is an indicator equal to one if individual i lived in an urban area in the initial year of span t, ℓ_{it} is an indicator equal to one if individual i was literate in the initial year of span t, r_{it} is individual i's occupational rank in the initial year of span t, and $\mathbf{x}_{it}$ is a vector of initial-year controls. The control vector includes initial-state or initial-county fixed effects, age-category indicators, birthplace indicators, marital status, household size, and an indicator for being the head of a household in the initial year, which are the variables available

[41] In addition to these fixed effects, $\mathbf{x}_{it}$ includes all the observables listed in Table 2 that are observed for both immigrants and natives. I include continuous covariates as higher order polynomials for flexibility, and include all measures of occupational status separately (i.e., I include the occupational category indicators and each of the three components of the occupational rank separately) for the greatest flexibility.

[42] Note that this is not a panel regression. Each linked individual constitutes a single observation. The time index simply denotes different linkage spans.

in all 6 initial censuses, and which are unlikely to confound the interpretation of results.[43] The coefficients of interest are α_t, β_t, and γ_t, which capture the relationship between an individual's migration probability and the dimensions of migrant selection, conditional on all other observables. Figure 8 presents the results of this estimation, with each panel focusing on a different metric of selection for either a 10- or 20-year span. I include estimates for both immigrants and natives to contextualize the results.

There is, of course, a danger arising from errors in linkage. Because it does not identify which particular observations are linked in error, the approach above of correcting for the probability of a false match cannot be applied. I address concerns over false linkage by repeating my analysis with all five of the linkage methods presented above, and drawing only conclusions that are robust to this variation of the strictness of the linkage method. In this section, I present results using the CZ linkage method; results with alternative linkage methods are presented in Online Appendix D.

Panels (a) and (b) of Figure 8 focus on the relationship between initial urban status and migration probability. The vector $\mathbf{x}_{it}$ includes initial-state fixed effects.[44] These estimates show that foreign-born urbanites were slightly more likely to move than foreign-born ruralists in the nineteenth century, and about 5 percentage points less likely in the twentieth. While the greater migration probability for urbanites in the nineteenth century is not robust to varying the strictness of the matching method, the decline in the magnitude of the coefficient—and thus the increasing attachment of the foreign born to urban areas over time—is robust to the choice of linkage method.[45] Also robust is the fact that the coefficients for natives are consistently greater than those for immigrants, indicating a larger attachment to urban areas than among immigrants, and that these coefficients, like those of immigrants, also declined over time.[46]

The negative coefficient on urban status in the twentieth century supports claims that immigrants exhibited an attachment to urban areas in general and to a greater extent than natives—claims that Willcox (1906) declared to be untrue. These patterns also support claims of an increasing tendency over time among immigrants to remain in cities. But as with changing migration rates over time, the technical truth of these claims belies a greater complexity. The greater attachment of immigrants to urban areas than natives is

[43] Keeping the control vector unchanged ensures that changes in selection over time are not driven by changes in the control variables included in the estimation. I do not include an indicator for farm residence or occupational category indicators because these would confound the interpretation of the urban residence indicator and occupational rank measure, respectively.

[44] I do not include county fixed effects because the identification of the urban coefficient would only come from within-county comparisons. This would imply that a number of large and important cities would not drive any of the results.

[45] Lower migration rates for urbanites support the validity of the results in the face of concerns over false matches. In general, urban status is associated with lower linkage rates, and it is reasonable to expect that this implies that urbanites would be more likely to be linked in error, which would tend to show higher migration rates among urbanites. The absence of this pattern suggests validity of the links. The loss of the urban migration premium in moving to stricter matching methods supports this.

[46] If false matches were more common among urbanites and declined in frequency over time, the true changes over time might have been different. A more negative coefficient on the urban indicator in the nineteenth century would moderate evidence of increasing attachment by immigrants to urban areas over time.

complicated by the fact that it was true throughout the Age of Mass Migration, not just in its later portion. Insofar as the distribution issues raised regarding the new immigrants were based on perceived differences from the old, these results show that a greater attachment by immigrants than natives to urban areas was not new. Similarly, the greater attachment by immigrants to urban areas over time is complicated by the fact that it was paralleled by natives. Thus, rather than being the product of some feature unique to immigrants, this pattern suggests that it was likely the product of the structural transformation and urbanization of the US economy over this period. The idea of the structural transformation of the economy generating changes that affected the perception of immigrants' role in the US economy mirrors the findings of Collins and Zimran (2020), who argue that falling rates of immigrant occupational catch up relative to natives over the Age of Mass Migration were also the product of this change.

Panels (c)–(f) of Figure 8 investigate whether foreign-born internal migrants were positively or negatively selected by focusing on the coefficients on occupational rank and on the literacy indicator. The vector $\mathbf{x}_{it}$ in all of these regressions includes initial-county fixed effects. Panels (c) and (d) focus on the occupational rank measure. For the foreign born as a whole, these estimates show clear evidence of negative selection into internal migration, though the magnitude of the estimate is reduced considerably and rendered economically insignificant by the use of stricter matching methods. There is also clear and robust evidence of more negative selection into internal migration by the foreign born than by the native born. These results contradict claims that the foreign born were characterized by positive selection into internal migration—that is, that it was the particularly poor immigrants who remained in cities and required encouragement to move inland.[47] Indeed, there is no evidence that this was true. It should be emphasized that the estimated magnitude of selection even in the less restrictive methods (where it is larger) is small. The occupational rank measure ranges from 0 to 1. Thus, a coefficient of -0.1 implies that a one-standard deviation increase in rank (about 0.25) is associated with a 2.5-percentage point decline in migration probability, which is not negligible but small.

Panels (e) and (f) focus on selection into internal migration on the basis of literacy. Although the estimates presented here show slightly negative selection into migration on the basis of literacy among immigrants in the nineteenth century followed by more negative selection in the twentieth, this result is not robust to the strictness of the linkage method. What is robust is that there is no evidence to support the contemporary claims of *positive* selection into internal migration on the basis of literacy.[48]

[47] A decline in false matches over time, together with a greater likelihood of a false match for individuals of lower occupational status, would tend to drive the coefficient in the nineteenth century to be more negative than it actually was. Thus, the presence of a greater rate of false matches in the nineteenth than the twentieth century would actually indicate that, if anything, the selection into internal migration became more negative over time, contrary to contemporary claims.

[48] As above, if false matches were more common for the illiterate, and since they are more common in the nineteenth century, there was, if anything, more negative selection into internal migration on the basis of literacy over time.

6 Destination Choice

Although immigrants were evidently mobile, the analysis above does not dispute per se the claims that immigrants were particularly attached to urban areas. That is, immigrants were moving, but they may have simply moved from one urban area to another. In this section, I study the destination choices of foreign-born internal migrants. I restrict attention to movers. As with the analysis of migrant selection, the approach that I take to ensure that my conclusions are not driven by errors in linkage is to draw only conclusions that are robust to the use of different linkage methods. In this section, I present results using CZ links; results with alternative linkage methods are presented in Online Appendix D. Rather than focusing solely on the characteristics of the chosen destinations, such that changes over time would largely reflect changes in the economy,[49] I compare the destination choices of foreign- and native-born inter-county migrants.

The simplest way to compare the destination choices of native- and foreign-born internal migrants is to compute an index of dissimilarity between them (Cutler, Glaeser, and Vigdor 2008b; Duncan and Duncan 1955). Figure 9 plots this index for each linkage span. The dissimilarity index between immigrants' and natives' destination choices was never below 0.4, which can be interpreted as indicating that at least 40 percent of foreign-born internal migrants in any particular span would have had to change their destinations for the destination-county distributions of the two groups to match. This difference could, in part, be the product of differences in the initial places of residence and characteristics of immigrants and natives. For this reason, Figure 9 also presents dissimilarity indices that are computed conditional on initial characteristics, including fixed effects for county of initial residence.[50] When comparing immigrants to observationally similar natives from the same initial county, the dissimilarity in destination choices is substantially reduced. Moreover, I find that the conditional dissimilarity in the destination choices of observationally similar native- and foreign-born internal migrants declined over time. Indeed, at no time were the internal migration patterns of immigrants and natives as similar as the early twentieth century, when the calls for distribution efforts and general concerns over the internal migration patterns of immigrants were at their peak.

Next, I focus on differences in the characteristics of destinations chosen by native- and foreign-born

[49] For instance, an increase in the probability that internal migrants chose urban destinations over time would be consistent with greater attachment to urban areas, but would also be consistent with rising urbanization of the economy generally.

[50] I compute this by estimating, for each county, an equation of the form

$$y_{ijt} = \beta_{jt} f_{it} + \mathbf{x}'_{it}\delta_{jt} + \varepsilon_{ijt}, \tag{1}$$

where y_{ijt} is an indicator equal to one if internal migrant i in linkage span t chose destination j and zero otherwise, f_{it} is an indicator equal to one if individual i in linkage span t was foreign born, and $\mathbf{x}_{it}$ is a vector of controls, including initial county-of-residence fixed effects. I then compute the dissimilarity index as $\hat{D}_t = \frac{1}{2}\sum_j \left|\hat{\beta}_{jt}\right|$, where $\hat{\beta}_{jt}$ are the estimates from equation (1). Note that when $\mathbf{x}_{it}$ is omitted from equation (1), this collapses to the standard Duncan and Duncan (1955) index.

movers. In particular, I estimate a series of regressions of the form

$$y_{it} = \beta_t f_{it} + \mathbf{x}'_{it}\delta_t + \varepsilon_{it},$$

where y_{it} is some characteristic of the destination chosen by internal migrant i in span t, f_{it} is an indicator equal to one if individual i in span t was foreign born and zero otherwise, and $\mathbf{x}_{it}$ is a vector of initial-year characteristics of individual i, including initial county-of-residence fixed effects.[51] The coefficients β_t capture the difference in the means of each destination characteristic between immigrants and natives. Figure 10 presents the results of the estimation both with and without initial-year controls. These figures also include the mean of the characteristic among the foreign born.

Panels (a) and (b) of Figure 10 focus on the probability of choosing an urban destination. The results parallel those for selection into migration on the basis of urban residence. I find that foreign-born internal migrants were indeed increasingly attached to urban areas over time, with about half choosing urban destinations in the nineteenth century and nearly 80 percent in the twentieth century. But this change is unsurprising given increasing urbanization over this time. Comparing the probability of choosing an urban destination across nativities shows that foreign-born internal migrants were more likely to choose urban destinations, even in comparison to observationally similar natives from the same initial county of residence. But this difference lessened over time. Conditional differences in the probability of choosing an urban destination declined from about 20 percentage points at the beginning of the study period to about 10 percentage points at the beginning of the twentieth century. It then rose slightly, but never exceeded 15 percentage points. Thus, when immigrants' attachment to urban areas was seen to be at its strongest, the probability of moving to urban areas was actually at the most similar across nativities that it would ever be. The gap is even smaller in context given the rise in the base probability of choosing an urban destination.

As with the results on selection, these results show that Willcox's (1906) claims that immigrants were not attached to urban areas in general and relative to natives were not true. At the same time, the results again show that the claims of greater attachment over time and across nativities are vastly oversimplified. Not only was the greater rate of urban destination choice over time shared with natives, indicating that it was an economy-wide phenomenon, but my results indicate that the old immigrants, whose internal migration patterns were discussed favorably by immigration critics in the early twentieth century, were, in fact, the *most* likely to choose urban areas relative to natives; the new immigrants, who were unfavorably compared

[51] The controls include all variables in Table 2, including all distinct measures of occupational status (including category indicators), with continuous variables included as higher order polynomials for flexibility.

to them, were most similar to natives in this regard. These results suggest that the growing attraction of immigrants to urban areas, rather than being the product of the change in immigrant origins, was simply the product of changing economic incentives in the county, which also attracted natives to urban areas.

Panels (c) and (d) of Figure 10 test the responsiveness of destination choices to labor demand growth. I measure this as $b_{jt} = \sum_k s_{jkt}\Delta_{kt}$, where s_{jkt} is the share of prime-aged men with reported occupations in the initial year of span t in county j in industry k,[52] where industries are defined according to Ruggles et al.'s (2019) ind1950 code, and Δ_{kt} is the growth rate in employment in industry k in the country as a whole over span t. This measure captures labor demand growth in county i without the endogeneity that would come from using the actual growth in employment. In the nineteenth century, the destinations chosen by foreign- and native-born internal migrants were qualitatively indistinguishable from one another in this regard. In the twentieth century, the gap grew somewhat with foreign-born internal migrants choosing destinations with slightly greater growth, even when making conditional comparisons. Importantly, there is no evidence that immigrants were meaningfully less responsive to labor demand growth than natives. This result supports the idea that immigrants were attracted to the locations where there were employment opportunities. In a sense, this is evidence that any distribution scheme was destined to fail. Indeed, the Dillingham Commission admitted as much in its report—immigrants would go where conditions were good.[xviii]

This is perhaps why the hopes of advocates of distribution policy that immigrants might be moved to the South and West never came to fruition, as borne out in panels (e) and (f) of Figure 10. These panels focus on the probability of choosing a destination in the South or West regions. The difference between native- and foreign-born internal migrants in the probability of choosing such a destination was large and increasing over time, approaching 30 percentage points. But this gap is reduced substantially when making conditional comparisons. In the nineteenth century, the difference is less than 5 percentage points, and although this difference grew slightly over time, it never exceeded 10 percentage points (relative to a base probability among natives of nearly 50 percent). This difference is non-negligible, but also small relative to a perceived unwillingness of immigrants to move to the south and west despite claims of a labor shortage there.

7 Summary of Robustness Checks

In addition to verifying the robustness of my conclusions to the choice of linkage methods in Online Appendix D, I verify in Online Appendix E the robustness of my conclusions to a variety of other potential perturbations

[52] To address changing boundaries, I apply initial-year employment shares to final-year boundaries following Hornbeck (2010).

of the data and analysis. All of the results in Online Appendix E are based on CZ-method links.

First, I use a different definition of migration. In addition to requiring that an individual reside in non-overlapping counties in the first and last year of the span to be considered a migrant, I classify such individuals as having migrated only if the distance of the move, defined as the distance between the centroids of the two counties, was at least 150 miles. The effect of this re-definition is to limit attention to longer distance moves, which may have been fundamentally different in their frequency, selection, and destination than moves generally. By definition, this rate of migration must be less than that presented in Figure 5. Of greater interest is that the main conclusion—that the internal migration rates of the foreign born fell initially and then remained largely stable, or even increased slightly, and were never meaningfully less than those of the native born during the Age of Mass Migration—remains unaffected. I continue to find that immigrants remained more attached to urban areas than did natives, but the general magnitude of the coefficient is such that the long-distance migration flow over-represented urbanites relative to the migration flow generally. There remains evidence of increasing relative attachment to urban areas over time. Selection patterns according to occupational rank and literacy are essentially unchanged. Patterns of destination choice for farther moves also became more similar between immigrants and natives over time—albeit by somewhat less than in the main results—in general and in terms of the probability of choosing urban destinations.

Next, I repeat the main analysis, but incorporate individuals in the censuses of 1900–1930 with "Not Yet Classified" occupations, according to a procedure described in the Online Appendix. There are no appreciable differences relative to the main results.

I also omit from spans beginning 1900–1930 immigrants who reported having been in the United States for more than 10 years. This cannot be done for other spans due to a lack of data on year of arrival. This exercise removes immigrants who had been in the United States for some time, whose internal migration patterns might have differed from those of more recent immigrants. This omission increases the internal migration rates of the foreign born in the twentieth century so as to magnify the otherwise slight rise in immigration rates from the nineteenth to the twentieth century, casting further doubt on claims of falling mobility. Patterns of selection are not appreciably affected. Destination choice results are also very similar, though the magnitude of the decline in destination dissimilarity over time is smaller than in the main results.

I next exclude natives with foreign-born fathers, a limitation possible only from 1880 onwards. This refinement prevents the comparison of first-generation immigrants to second-generation ones, who may have differed meaningfully from other natives. This adjustment slightly increases the mobility of natives in the twentieth century so that the more restrictive linkage methods estimate a slightly larger migration premium

for natives than in the main results, but this premium is still small. There is also a slightly smaller decline in the dissimilarity of destination choices over time between immigrants and natives.

Next, I remove any individual living in the South in the initial year of the span. With immigrants particularly unlikely to settle in the South, this approach compares them to the natives among whom they resided. This is a limitation often made in other studies of immigrant assimilation (Abramitzky, Boustan, and Eriksson 2014; Collins and Zimran 2020). When imposing this limitation, the similarity in destination choices between immigrants and natives is increased relative to the baseline results. It also reduces the difference between native- and foreign-born internal migrants in the probability of choosing a southern or western destination. Indeed, immigrants in the nineteenth century were actually more likely than observationally similar natives to choose a southern or western destination. But even when the direction of this difference reverses in the twentieth century, the difference remains unremarkable at about 5 percentage points.

Finally, I explore robustness of my results on selection into migration on the basis of urban status and for the probability that movers chose an urban destination to different definitions of urban residence.[53] I focus on three alternatives to the benchmark urban indicator based on a population of at least 2,500: an indicator based on a population of at least 25,000; an indicator for being in one of the 20 largest cities of the country in each census year;[54] or an indicator for being in a main immigrant arrival port (Boston, New York, Philadelphia, Baltimore, or New Orleans). The first two alternative definitions address concerns that the debate in the early 20th century was focused on immigrants' presence in larger cities only. The last focuses in particular on the perceived tendency of immigrants to remain in arrival cities. The main results are not qualitatively affected by using the first two definitions, and using the third yields substantially similar results, though there is less evidence of changes over time in selection or destination choice.

8 Geographic Assimilation and Urbanization

To what extent did immigrants' internal migration contribute to their geographic assimilation in the US economy? Figure 11 presents dissimilarity indices between the county-of-residence distributions of immigrants and natives in my linked samples (including both movers and stayers) for the initial and final years of each span. In each case, there is a decline in this index from the initial to the final year, showing that geographic assimilation did take place. But these changes were universally small, indicating that immigrants' geographic assimilation was severely limited. This lack of geographic assimilation stands in contrast to the

[53] I do not replace the urban definition in the creation of linkage weights, simply in the analysis.
[54] These first two alternate definitions are based on the citypop variable provided by Ruggles et al. (2019).

rapid economic assimilation of immigrants in the nineteenth century (Collins and Zimran 2020), the relative similarity of immigrants' and natives' occupational status in the twentieth century (Abramitzky, Boustan, and Eriksson 2014; Collins and Zimran 2020), and the rapid cultural assimilation of immigrants in the Age of Mass Migration (Abramitzky, Boustan, and Eriksson 2020). Instead, it is more in line with Eriksson and Ward's (2019) findings of persistence in immigrants' segregation from natives within cities.

To what extent did immigrants' internal mobility contribute to US urbanization? Figure 12 plots the urbanization rates of individuals in the linked samples for the initial and final years of each sample (the solid lines) and the urbanization rates of the native- and foreign-born populations as a whole (the dotted lines). Panels (a) and (b) include all individuals in the linked samples whereas panels (c) and (d) include only those who moved between the initial and final year of the span. In panels (a) and (b), the urbanization rates of the native born in the linked samples track fairly closely with the aggregate rate of urbanization. Immigrants, on the other hand, largely follow national trends in the nineteenth century, but show shallower increases in the linked samples than in the aggregate data in the twentieth century. While urbanization still increased in these samples, it did so more slowly than for natives and for all immigrants, indicating that some of the persistent immigrant-native premium in urbanization was due to new arrivals. Panels (c) and (d), which focus only on movers, strengthen this view. Native movers in the twentieth century urbanized faster than natives as a whole, whereas immigrant movers did so at rates slower than immigrants as a whole. Far from failing to leave urban areas, immigrants' internal migration worked against the growing urbanization of the economy, while that of natives did contribute to growing urbanization. In Online Appendix E, I repeat this analysis with the alternative definitions of urban, finding similar results, and in some cases evidence of a *decline* in immigrants' urbanization with time in the United States.

This notion receives further support from Figure 13, which presents similar destination choice results to Figure 10 but with the change in log population density of the residence place as the outcome.[55] This figure shows that, throughout the study period, immigrants moved to destinations with considerably lower population density. Moreover, although immigrant movers experienced a smaller decrease in population density than did observationally similar native movers from the same initial county of residence, the unconditional difference was consistently negative, indicating that immigrant mobility on the whole tended to reduce density to a greater extent than did that of natives.

[55]For instance, the outcome for an individual living in county j in year t and in county k in year $t+10$ is $\log(d_{kt}) - \log(d_{jt})$, where d denotes population density. Note that the population density of the destination county is from the initial year of the span rather than the final. Adjustments for changing county borders over time are according to Hornbeck's (2010) method.

9 Conclusion

Immigrant distribution was a major issue in the early twentieth-century US immigration policy debate and remains important in modern immigration policy and in the economics of immigration (though perhaps under a different name). During the Age of Mass Migration, the encouragement of immigrants to leave coastal cities and move inland was seen as away to dampen opposition to immigration and to avoid restriction. Indeed, some even claimed that all of the problems arising from immigration could be solved by distribution policy. The calls for distribution were based on the view—which was also shared by advocates of outright immigration restriction, who saw it as further evidence that immigration should be restricted—that immigrants were not mobile while in the United States, and were particularly unlikely to leave urban areas. I construct new linked census datasets of immigrants and natives in the United States spanning the Age of Mass Migration to test these claims, giving a clearer view of the internal migration patterns of immigrants over time than has ever before been possible, and ultimately developing a clearer understanding of immigrants' assimilation into the US economy. The hazards of using linked data to study internal migration limit the conclusions that I can draw, but so little was previously known about those patterns that any insights from linked data are new.

I find that the claims underlying the push for immigrant distribution are either unsubstantiated by the data, oversimplified, or reflected broader changes in the economy and in the landscape of internal migration, rather than changes specific to immigrants' distribution patterns. Nonetheless, it seems unlikely that many of those concerned about immigrant distribution would have been moved in their views, even with knowledge of these results. On the one hand, my results apply only to persistent immigrants, and some of the concern regarding immigrant distribution was directed at immigrants who returned to their home countries without having left their arrival city. But it is also unlikely that contemporaries would have taken a more accepting view even of persistent immigrants. Although I have shown that it was in no way unusual, immigrants did, in fact, go to cities. Presumably this was because this is where there was economic opportunity, as evidenced by the contemporaneous growth of natives' attachment to cities and by immigrants' and natives' similar responsiveness to labor demand growth. But the continued presence of immigrants put them in competition with natives (Abramitzky et al. 2019a; Price, vom Lehn, and Wilson 2020). As long as immigrants did not, as distribution efforts intended for them to do, defy these incentives in choosing a place of residence, it seems inevitable that their presence would have engendered opposition.

References

Aaronson, Daniel, Jonathan Davis, and Karl Schulze (2020). "Internal Immigrant Mobility in the Early 20th Century: Evidence from Galveston, Texas." *Explorations in Economic History* 76, 101317.

Abbott, Greg (2020). "Governor Greg Abbott Secretary Michael Pompeo." Letter dated January 10.

Abramitzky, Ran, Philipp Ager, Leah Platt Boustan, Elior Cohen, and Casper Worm Hansen (2019a). "The Effects of Immigration on the Economy: Lessons from the 1920s Border Closure." NBER Working Paper 26536.

Abramitzky, Ran and Leah Platt Boustan (2017). "Immigration in American Economic History." *Journal of Economic Literature* 55:4, pp. 1311–1345.

Abramitzky, Ran, Leah Platt Boustan, and Dylan Connor (2020). "Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office." NBER Working Paper 27372.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson (2014). "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122:3, pp. 467–506.

——— (2020). "Do Immigrants Assimilate More Slowly Today than in the Past?" *American Economic Review: Insights* 2:1, pp. 125–141.

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J. Feigenbaum, and Santiago Pérez (2019b). "Automated Linking of Historical Data." *Journal of Economic Literature* Forthcoming.

Abramitzky, Ran, Leah Platt Boustan, and Myera Rashid (2020). *Census Linking Project: Version 1.0 [dataset].* `https://censuslinkingproject.org`.

Alsan, Marcella, Katherine Eriksson, and Gregory Niemesh (2020). "Understanding the Success of the Know-Nothing Party." NBER Working Paper 28078.

Antonie, Luiza, Kris Inwood, Chris Minns, and Fraser Summerfield (2018). "Bias, Accuracy, and Sample Size in the Systematic Linking of Historical Records." *International Journal of Population Data Science* 3:4.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey (2020). "How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth." *Journal of Economic Literature* 58:4, pp. 997–1044.

Bandiera, Oriana, Imran Rasul, and Martina Viarengo (2013). "The Making of Modern America: Migratory Flows in the Age of Mass Migration." *Journal of Development Economics* 102, pp. 23–47.

Barde, Robert, Susan B. Carter, and Richard Sutch (2006a). "Table Ad1–2: US Immigrants and Emigrants: 1820–1998." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.541–1.543.

——— (2006b). "Table Ad106–120: Immigrants, by country of last residence—Europe, 1820–1997." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.560–1.563.

——— (2006c). "Table Ad222–225: Immigrants, by sex, 1820–1997." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.586–1.587.

Bartel, Ann P. (1989). "Where Do the New US Immigrants Live?" *Journal of Labor Economics* 7:4, pp. 371–391.

Bartel, Ann P. and Marianne J. Koch (1991). "Internal Migration of US Immigrants." In *Immigration, Trade and the Labor Market*. John M. Abowd and Richard B. Freeman (ed.). Chicago: University of Chicago Press. Chap. 4, pp. 121–134.

Beenstock, Michael (1997). "The Internal Migration of Immigrants: Israel 1969–1972." *Quarterly Review of Economics and Finance* 37, pp. 263–284.

——— (1999). "Internal Migration by Immigrants in the Short-Run: Israel 1992–1994." *International Migration Review* 33:4, pp. 1098–1106.

Benton-Cohen, Katherine (2018). *Inventing the Immigration Problem: The Dillingham Commission and its Legacy*. Cambridge: Harvard University Press.

Biavaschi, Costanza and Giovanni Facchini (2020). "Immigrant Franchise and Immigration Policy: Evidence from the Progressive Era." CEPR Discussion Paper 14684.

Bleakley, Hoyt and Aimee Chin (2004). "Language Skills and Earnings: Evidence from Childhood Immigrants." *Review of Economics and Statistics* 86:2, pp. 481–496.

——— (2010). "Age at Arrival, English Proficiency, and Social Assimilation Among US Immigrants." *American Economic Journal: Applied Economics* 2:1, pp. 165–192.

Borjas, George J. (1999). *Heaven's Door: Immigration Policy and the American Economy*. Princeton: Princeton University Press.

Borjas, George J. (2001). "Does Immigration Grease the Wheels of the Labor Market?" *Brookings Papers on Economic Activity* 1, pp. 69–133.
——— (2014). *Immigration Economics*. Cambridge: Harvard University Press.
Brenke, Karl (2015). "Distribution of Refugees Very Uneven Among EU Member States—Even When Accounting for Economic Strength and Total Population." *DIW Economic Bulletin* 39, pp. 511–523.
Cadena, Brian C. (2013). "Native Competition and Low-Skilled Immigrant Inflows." *Journal of Human Resources* 48:4, pp. 910–944.
——— (2014). "Recent Immigrants as Labor Market Arbitrageurs: Evidence from the Minimum Wage." *Journal of Urban Economics* 80, pp. 1–12.
Cadena, Brian C. and Brian K. Kovak (2016). "Immigrants Equilibriate Local Labor Markets: Evidence from the Great Recession." *American Economic Journal: Applied Economics* 8:1, pp. 257–290.
Card, David (2005). "Is the New Immigration Really So Bad?" *Economic Journal* 115, F300–F323.
Card, David and Ethan G. Lewis (2007). "The Diffusion of Mexican Immigrants during the 1990s: Explanations and Impacts." In *Mexican Immigration to the United States*. George J. Borjas (ed.). Chicago: University of Chicago Press. Chap. 6, pp. 193–227.
Cascio, Elizabeth U. and Ethan G. Lewis (2012). "Cracks in the Melting Pot: Immigration, School Choice, and Segregation." *American Economic Journal: Economic Policy* 4:3, pp. 91–117.
Chiquiar, Daniel and Gordon H. Hanson (2005). "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." *Journal of Political Economy* 113:2, pp. 239–281.
Collins, William J. and Marianne H. Wanamaker (2015). "The Great Migration in Black and White: New Evidence on the Selection and Sorting of Southern Migrants." *Journal of Economic History* 75:4, pp. 947–992.
——— (2020). "Up from Slavery? African American Intergenerational Mobility since 1880." NBER Working Paper 23395.
Collins, William J. and Ariell Zimran (2019). "The Economic Assimilation of Irish Famine Migrants to the United States." *Explorations in Economic History* 74, 101302.
——— (2020). "Working Their Way Up? US Immigrants' Changing Labor Market Assimilation in the Age of Mass Migration." NBER Working Paper 26414.
Commissioner-General of Immigration (1903). *Annual Report of the Commissioner-General of Immigration for the Fiscal Year Ended June 30, 1903*. Washington: Government Printing Office.
Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor (2008a). "When are Ghettos Bad? Lessions from Immigrant Segregation in the United States." *Journal of Urban Economics* 63, pp. 759–774.
——— (2008b). "Is the Melting Pot Still Hot? Explaining the Resurgence of Immigrant Segregation." *Review of Economics and Statistics* 90:3, pp. 478–497.
Damm, Anna Piil (2009). "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence." *Journal of Labor Economics* 27:2, pp. 281–314.
Daniels, Roger (2004). *Guarding the Golden Door: American Immigration Policy and Immigrants since 1882*. New York: Hill and Wang.
——— (1991 [2002]). *Coming to America: A History of Immigration and Ethnicity in American Life*. New York: Harper Perennial.
Department of Commerce (1933). *Fifteenth Census of the United States: 1930, Population*. Vol. II. Washington: Government Printing Office.
Department of the Interior (1895). *Report on Population of the United States at the Eleventh Census: 1890, Part I*. Washington: Government Printing Office.
Duncan, Otis Dudley and Beverly Duncan (1955). "A Methodological Analysis of Segregation Indexes." *American Sociological Review* 20:2, pp. 210–217.
Edin, Per-Anders, Peter Frederiksson, and Olof Åslund (2003). "Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment." *Quarterly Journal of Economics* 118:1, pp. 329–357.
Eriksson, Katherine (2019). "Ethnic Enclaves and Immigrant Outcomes: Norwegian Immigrants during the Age of Mass Migration." *European Review of Economic History* 24, pp. 427–446.
Eriksson, Katherine and Zachary Ward (2019). "The Residential Segregation of Immigrants in the United States from 1850 to 1940." *Journal of Economic History* 79:4, pp. 989–1026.
——— (2020). "Immigrants and Cities during the Age of Mass Migration." *Regional Science and Urban Economics* Forthcoming.
Feigenbaum, James J. and Hui Ren Tan (2019). "The Return to Education in the Mid-20th Century: Evidence from Twins." NBER Working Paper 26407.

Ferrie, Joseph P. (1995). "Up and Out or Down and Out? Immigrant Mobility in the Antebellum United States." *Journal of Interdisciplinary History* 26:1, pp. 33–55.
——— (1996). "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census Manuscript Schedules." *Historical Methods* 29:4, pp. 141–156.
——— (1997). "Long-Term Perspectives on Geographic and Economic Mobility, 1800–1900." Mimeo., Northwestern University.
——— (1999). *Yankeys Now: Immigrants in the Antebellum United States, 1840–1860*. New York: Oxford University Press.
——— (2005). "How Ya Gonna Keep 'Em Down on the Farm [When They've Seen Schenectady]?: Rural-to-Urban Migration in 19th Century America, 1850–70." Mimeo., Northwestern University.
——— (2006a). "Internal Migration." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.489–1.494.
——— (2006b). "Table Ac1–42: Interstate migration—native-born population, by race and residence within or outside the state of birth: 1850–1990." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.495–1.497.
Gershteyn, Yefim (2000). "Use of SPEDIS Function in Finding Specific Values." Proceedings of the Twenty-fifth Annual SAS Users Group International Conference, Indianapolis.
Goldin, Claudia (1994). "The Political Economy of Immigration Restriction in the United States, 1890 to 1921." In *The Regulated Economy: A Historical Approach to Political Economy*. Claudia Goldin and Gary D. Libecap (ed.). Chicago: University of Chicago Press, pp. 223–258.
Gould, John D. (1980). "European Inter-Continental Emigration: The Role of 'Diffusion' and 'Feedback'." *Journal of European Economic History* 9:2, pp. 267–315.
Gurak, Douglas T. and Mary M. Kritz (2000). "The Interstate Migration of US Immigrants: Individual and Contextual Determinants." *Social Forces* 78:3, pp. 1017–1039.
Haines, Michael R. and Richard Sutch (2006). "Table Aa6–8: Population: 1790–2000 [Annual estimates]." In *Historical Statistics of the United States*. Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (ed.). Cambridge: Cambridge University Press, pp. 1.28–1.30.
Hall, Matthew (2009). "Interstate Migration, Spatial Assimilation, and the Incorporation of US Immigrants." *Population, Space and Place* 15, pp. 57–77.
Hall, Patricia Kelly and Steven Ruggles (2004). "'Restless in the Midst of Their Prosperity': New Evidence on the Internal Migration of Americans, 1850–2000." *Journal of American History* 91:3, pp. 829–846.
Hall, Prescott F. (1904). "Selection of Immigration." *Annals of the American Academy of Political and Social Science* 24:1, pp. 169–184.
——— (1906). *Immigration and its Effects upon the United States*. New York: Henry Holt and Company.
Hatton, Timothy J. and Zachary Ward (2019). "International Migration in the Atlantic Economy 1850–1940." In *Handbook of Cliometrics*. Claude Diebolt and Michael Haupert (ed.). Springer.
Hatton, Timothy J. and Jeffrey G. Williamson (1998). *The Age of Mass Migration: Causes and Economic Impact*. New York: Oxford University Press.
Henning, Sabine (1997). "The Internal Migration of the Foreign-Born Population in the United States: 1985–1990." PhD thesis. University of Colorado at Boulder.
Hornbeck, Richard (2010). "Barbed Wire: Property Rights and Agricultural Development." *Quarterly Journal of Economics* 125, pp. 767–810.
——— (2020). "Dust Bowl Migrants: Identifying an Archetype." NBER Working Paper 27656.
Inwood, Kris, Chris Minns, and Fraser Summerfield (2019). "Occupational Income Scores and Immigrant Assimilation. Evidence from the Canadian Census." *Explorations in Economic History* 72, pp. 114–122.
Jaeger, David A. (2007). "Green Cards and the Location Choices of Immigrants in the United States, 1971–2000." *Research in Labor Economics* 27, pp. 131–183.
Kaufmann, David (2020). "Debating Responsibility-Sharing: An Analysis of the European Parliament's Debates on the Common European Asylum System." *European Policy Analysis*, pp. 1–19.
Kritz, Mary M. and Douglas T. Gurak (2001). "The Impact of Immigrationon the Internal Migration of Natives and Immigrants." *Demography* 38:1, pp. 133–145.
Kritz, Mary M., Douglas T. Gurak, and Min-Ah Lee (2011). "Will They Stay? Foreign-Born Out-Migration from New US Destinations." *Population Research and Policy Review* 30, pp. 537–567.

Long, Jason and Joseph P. Ferrie (2013). "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." *American Economic Review* 103:4, pp. 1109–1137.
Lubotsky, Darren (2007). "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings." *Journal of Political Economy* 115:5, pp. 820–867.
Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles (2019). *IPUMS National Historical Geographic Information System: Version 14.0* [database]. Minneapolis: IPUMS.
Maré, David C., Melanie Morten, and Steven Stillman (2007). "Settlement Patterns and the Geographic Mobility of Recent Migrants to New Zealand." *New Zealand Economic Papers* 41:2, pp. 163–196.
McKenzie, David and Hillel Rapoport (2010). "Self-Selection Patterns in Mexico-US Migration: The Role of Migration Networks." *Review of Economics and Statistics* 92:4, pp. 811–821.
Molloy, Raven, Christopher L. Smith, and Abigail Wozniak (2011). "Internal Migration in the United States." *Journal of Economic Perspectives* 25:3, pp. 173–196.
Newbold, K. Bruce (1999). "Internal Migration of the Foreign-Born: Population Concentration or Dispersion?" *Population and Environment* 20, pp. 259–276.
Norton, Eliot (1904). "The Diffusion of Immigration." *Annals of the American Academy of Political and Social Science* 24, pp. 161–165.
Preston, Samuel H. and Michael R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America*. Princeton: Princeton University Press.
Price, Joseph, Christian vom Lehn, and Riley Wilson (2020). "The Winners and Losers of Immigration: Evidence from Linked Historical Data." NBER Working Paper 27156.
Rashid, Saman (2009). "Internal Migration and Income of Immigrant Families." *Journal of Immigrant & Refugee Studies* 7:2, pp. 180–200.
Roosevelt, Theodore (1905). *State of the Union Address, December 5, 1905*. URL: https://www.gutenberg.org/files/5032/5032-h/5032-h.htm.
Rosenbloom, Joshua L. and William A. Sundstrom (2004). "The Decline and Rise of Interstate Migration in the United States: Evidence from the IPUMS, 1850–1990." *Research in Economic History* 22, pp. 289–325.
Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek (2019). *Integrated Public Use Microdata Series: Version 9.0* [dataset]. Minneapolis: University of Minnesota.
Saavedra, Martin and Tate Twinam (2020). "A Machine Learning Approach to Improving Occupational Income Scores." *Explorations in Economic History* 75, 101304.
Schündeln, Matthias (2014). "Are Immigrants More Mobile than Natives? Evidence from Germany." *Journal of Regional Science* 54:1, pp. 70–95.
Sichko, Christopher (2020). "Migrant Selection and Sorting in the Great American Drought." Mimeo., Vanderbilt University.
Speranza, Gino Carlo (1904). "Solving the Immigration Problem." *Outlook* 76, pp. 928–933.
Spitzer, Yannay (2019). "Pale in Comparison: Jews as a Rural Service Minority." CEPR Discussion Paper 14262.
Spitzer, Yannay and Ariell Zimran (2018). "Migrant Self-Selection: Anthropometric Evidence from the Mass Migration of Italians to the United States, 1907–1925." *Journal of Development Economics* 134, pp. 226–247.
——— (2021). "Like an Ink Blot on Paper: Testing the Diffusion Hypothesis of Mass Migration, Italy 1875–1920." Mimeo., Vanderbilt University.
Stewart, James I. (2006). "Migration to the Agricultural Frontier and Wealth Accumulation, 1860–1870." *Explorations in Economic History* 43, pp. 547–577.
——— (2009). "Economic Opportunity or Hardship? The Causes of Geographic Mobility on the Agricultural Frontier, 1860–1880." *Journal of Economic History* 69:1, pp. 238–268.
US Congress (1911a). *Reports of the Immigration Commission: Abstracts of Reports of the Immigration Commission*. Vol. 1. Washington: Government Printing Office.
——— (1911b). *Reports of the Immigration Commission: Immigrants in Cities*. Vol. 26. Washington: Government Printing Office.
——— (1911c). *Reports of the Immigration Commission: Statements and Recommendations Submitted By Societies and Organizations Interested in the Subject of Immigration*. Vol. 41. Washington: Government Printing Office.
Ward, Zachary (2017). "Birds of Passage: Return Migration, Self-Selection and Immigration Quotas." *Explorations in Economic History* 64, pp. 37–52.
——— (2019). "The Low Return to English Fluency during the Age of Mass Migration." *European Review of Economic History* 24, pp. 219–242.
——— (2020). "Intergenerational Mobility in American History: Accounting for Race and Measurement Error." Mimeo., Baylor University.

Weine, Stevan Merrill, Yael Hoffman, Norma Ware, Toni Tugenberg, Leonce Hakizimana, Gonwo Dahnweigh, Madeleine Currie, and Maureen Wagner (2011). "Secondary Migration and Relocatin Among African Refugee Families in the United States." *Family Process* 50, pp. 27–46.

Whelpley, James Davenport (1905). "Emigration: An International Affair." *Fortnightly Review* 77:458, pp. 317–326.

Willcox, Walter F. (1906). "The Distribution of Immigrants in the United States." *Quarterly Journal of Economics* 20:4, pp. 523–546.

Zimran, Ariell (2021). "Internal Migration in the United States: Rates, Selection, and Sorting, 1850–1940." Mimeo., Vanderbilt University.

Tables

Table 1: Linkage rates for the foreign born

Initial Year	*Final Year*	(1) CZ	(2) ABEN	(3) ABEE	(4) Int	(5) Exact
1850	1860	722,500	722,500	722,500	722,500	722,500
		47,333	46,793	43,410	10,872	7,473
		(0.066)	(0.065)	(0.060)	(0.015)	(0.010)
1850	1870	722,500	722,500	722,500	722,500	722,500
		45,087	43,728	40,138	10,060	6,873
		(0.062)	(0.061)	(0.056)	(0.014)	(0.010)
1860	1870	1,260,709	1,260,709	1,260,709	1,260,709	1,260,709
		71,747	76,315	73,270	17,475	12,209
		(0.057)	(0.061)	(0.058)	(0.014)	(0.010)
1860	1880	1,260,709	1,260,709	1,260,709	1,260,709	1,260,709
		69,304	74,176	69,184	17,332	11,994
		(0.055)	(0.059)	(0.055)	(0.014)	(0.010)
1870	1880	1,444,826	1,444,826	1,444,826	1,444,826	1,444,826
		102,539	106,047	101,074	28,968	20,890
		(0.071)	(0.073)	(0.070)	(0.020)	(0.014)
1880	1900	1,408,927	1,269,809	1,266,563	1,251,351	1,180,481
		118,279	120,796	116,337	36,999	27,664
		(0.084)	(0.095)	(0.092)	(0.030)	(0.023)
1900	1910	2,610,947	2,387,360	2,376,707	2,361,864	2,230,596
		228,522	225,616	210,178	76,374	59,943
		(0.088)	(0.095)	(0.088)	(0.032)	(0.027)
1900	1920	2,180,293	1,997,002	1,985,144	1,968,741	1,868,546
		220,396	215,530	196,091	73,753	56,193
		(0.101)	(0.108)	(0.099)	(0.037)	(0.030)
1910	1920	3,529,487	3,236,767	3,225,170	3,211,236	3,038,454
		309,801	299,172	261,623	100,626	78,894
		(0.088)	(0.092)	(0.081)	(0.031)	(0.026)
1910	1930	3,102,661	2,858,216	2,849,589	2,834,830	2,699,346
		271,049	262,454	235,517	90,703	71,651
		(0.087)	(0.092)	(0.083)	(0.032)	(0.027)

Notes: For each linkage span, this table presents the number of individuals at risk to be linked (the first row of each grouping), the number successfully linked (the second row of each grouping), and the fraction successfully linked (the third row of each grouping) by each method. For links terminating 1860-1880, the table is limited to white men aged 18-40 in the initial year of the span. For links terminating 1900-1930, the table is limited to men in the final census at risk for linkage to someone aged 18–40 in the initial census and who are either linked or who report an arrival date before the initial census.

Table 2: Observables available in each initial census

Variable	1850	1860	1870	1880	1900	1910
Age	X	X	X	X	X	X
Urban	X	X	X	X	X	X
City Population[a]	X	X	X	X	X	X
Farm	X	X	X	X	X	X
Literacy	X	X	X	X	X	X
Marital status	X	X	X	X	X	X
Occupation[b]	X	X	X	X	X	X
Household size	X	X	X	X	X	X
Household head	X	X	X	X	X	X
Birthplace	X	X	X	X	X	X
Place of residence	X	X	X	X	X	X
Real property	X	X	X			
Personal property		X	X			
Father's nativity[c]				X	X	X
Years in US[d]					X	X
Speaks English[d]					X	X
Owns home (mortgaged)					X	X
Owns home (free and clear)					X	X

[a]: used only for alternate definitions of urban residence in Online Appendix F.
[b]: used to create occupational categories, the three occupational rank measures, and the average occupational rank measure, which are available for every census year
[c]: native born only
[d]: foreign born only

Figures

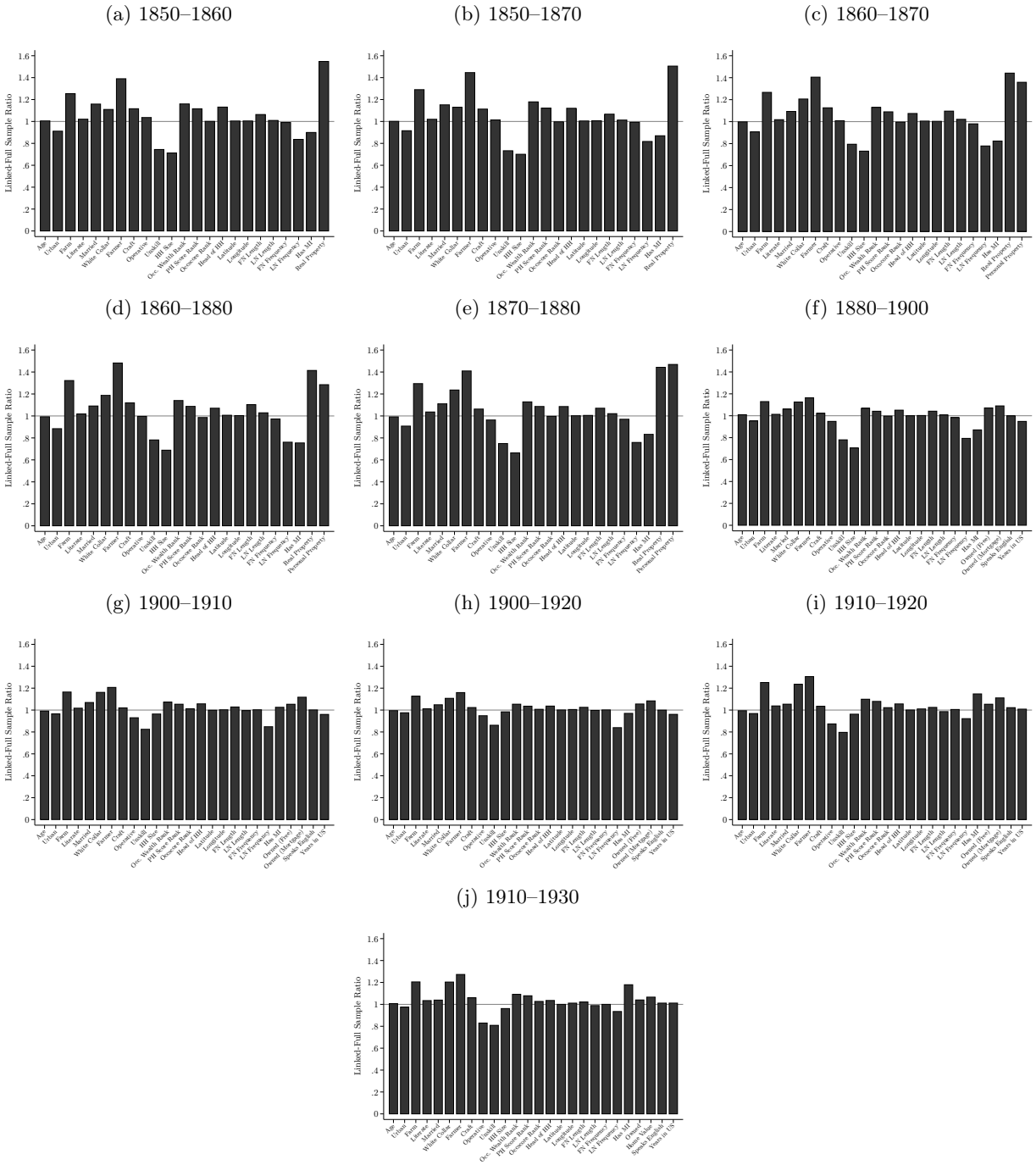


Figure 1: Representativeness of the linked sample

Note: Each bar presents the ratio of the mean of each variable in the linked sample relative to the full sample at risk for linkage. For panels (a)–(e), the sample at risk is from the initial census year. For panels (f)–(j), the sample at risk is from the final census year, limiting to individuals whose ages and years of arrival put them at risk for linkage. Because I retain links to immigrants whose arrival year is later than the initial year, these individuals are also included in the sample at risk for linkage.

Figure 2: Definition of inter-county migration

Note: These figures show a set of counties in northeastern Illinois that experienced changing boundaries between 1850 and 1870. In 1850, it is not possible to distinguish between points A and B or between points C and D. In 1870, it is not possible to distinguish between points B and C. As a result, only moves from A or B to D or from C or D to A will be considered an inter-county move.

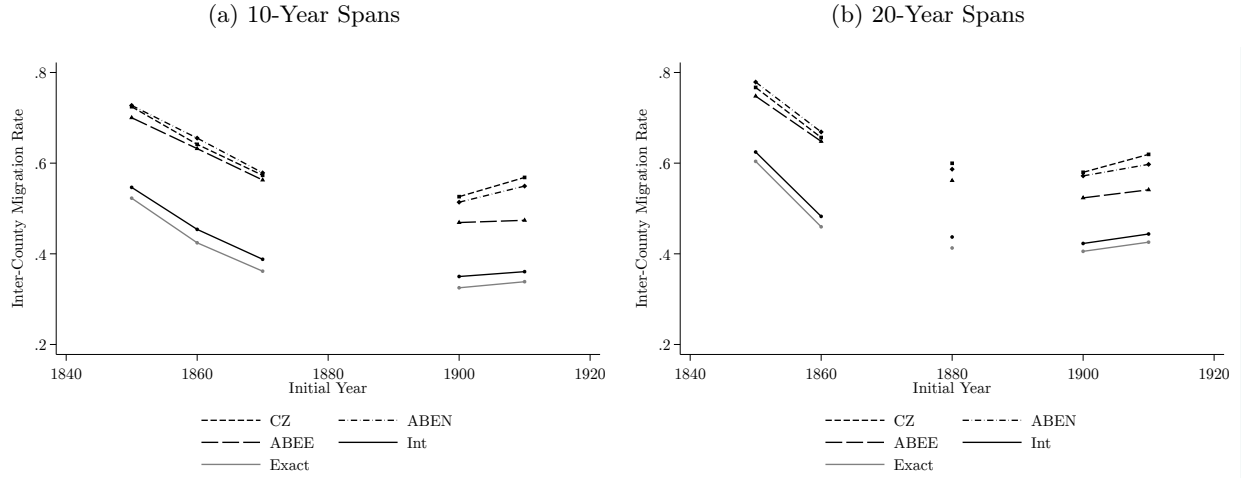


Figure 3: Inter-county migration rates for the foreign born

Note: Each figure shows the probability that a foreign-born individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, according to each linkage method. All observations are weighted by inverse linkage probability.

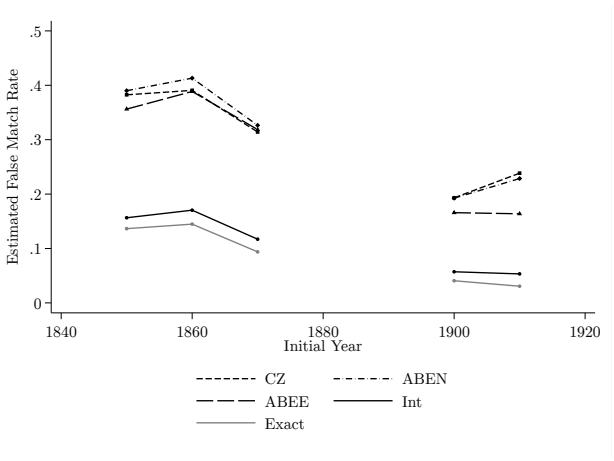


Figure 4: Estimated rates of false linkage

Note: This figure shows the estimated rate of false matching for each linkage method and 10-year span according to the comparison of inter-state migration estimates by the linkage method and by the method based on the birth places and ages of children. The rate is computed according to Bayes's Theorem, as explained in text and in Appendix A.

(a) 10-Year Inter-county Migration

(b) 20-Year Inter-county Migration

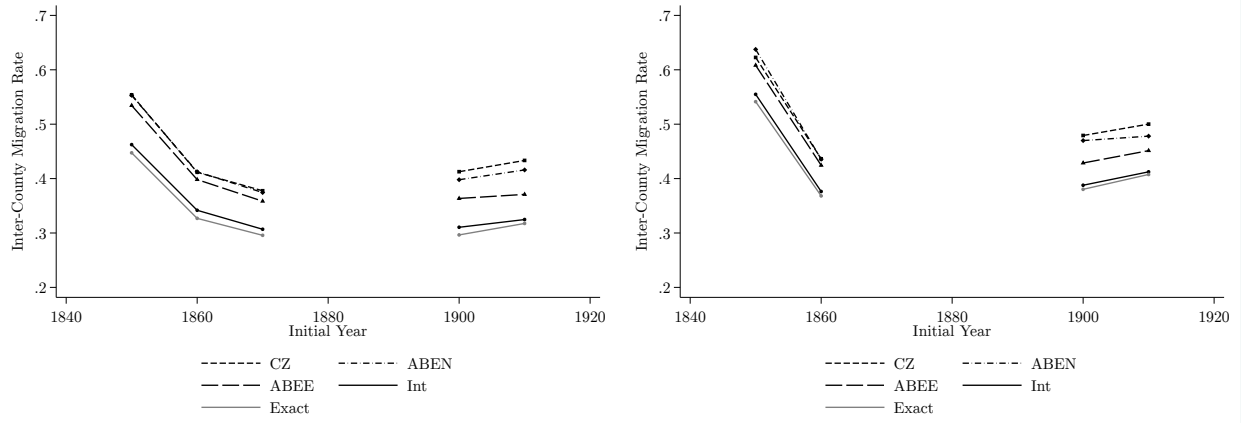


Figure 5: Inter-county migration rates by linkage method and span, corrected for false matches

Note: Each figure shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, according to each linkage method. All observations are weighted by inverse linkage probability. The estimates are then adjusted for false matches according to the method presented in text.

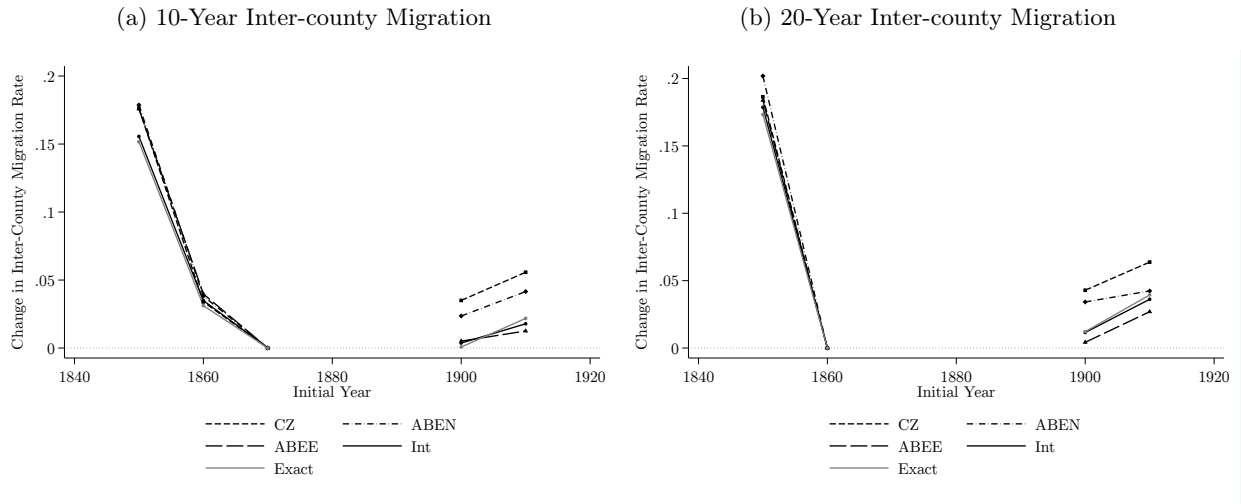


Figure 6: Changes in inter-county migration rates by linkage method and span, corrected for false matches

Note: Each figure shows the probability that an individual in the linked sample beginning in the year on the x-axis was observed living in a different non-overlapping county 10 or 20 years later, according to each linkage method. All observations are weighted by inverse linkage probability. The estimates are then adjusted for false matches according to the method presented in text. Finally, the estimates are adjusted so that the migration rate for the last span beginning in the nineteenth century is set to zero, so that the figure presents the change in migration rates over time.

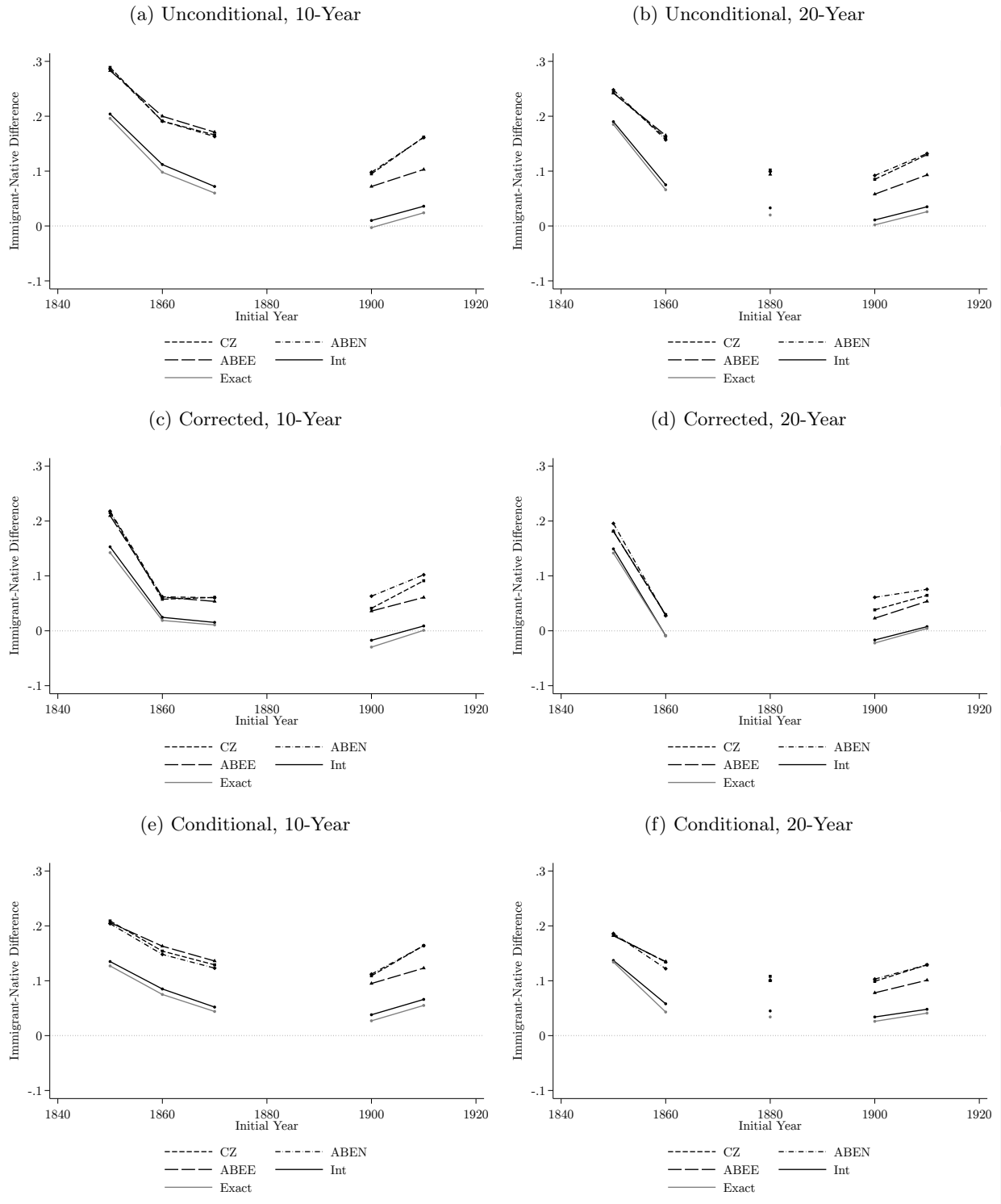


Figure 7: Differences in inter-county migration rates by nativity and span

Note: These figures present estimated differences in inter-county migration rates by nativity. Panels (a) and (b) present unconditional differences. Panels (c) and (d) are unconditional differences after correcting for false matches as described in text and in Appendix A. Panels (e) and (f) are from regressions including all controls available for both immigrants and natives in the census of the initial year of the span. All observations are weighted by inverse linkage probability.

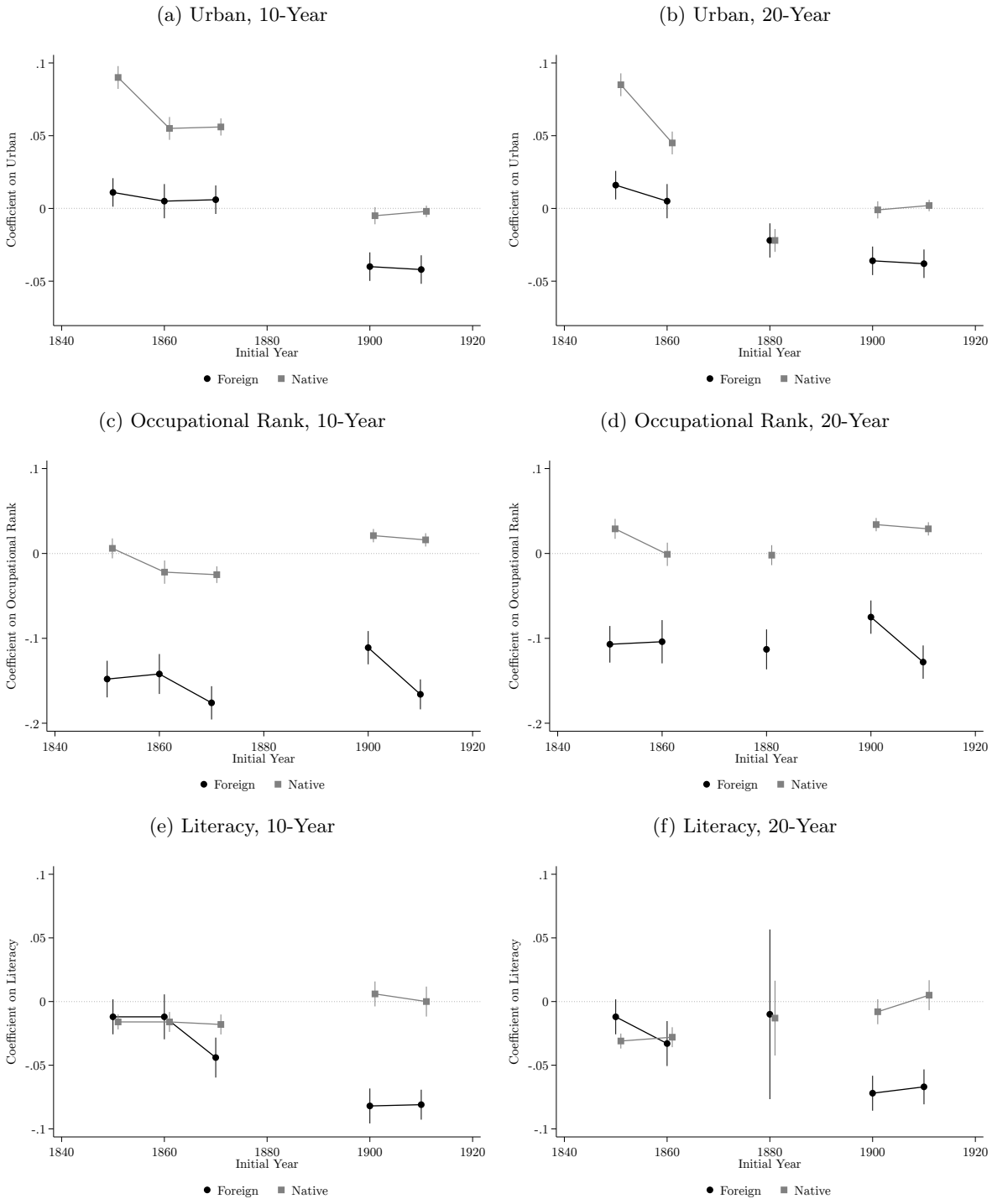


Figure 8: Selection into migration

Note: These are coefficients from regressing a migration indicator on various individual and household characteristics, the variables in this figure, and state- and birthplace-specific fixed effects, with separate regressions for each nativity-sample. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent confidence intervals.

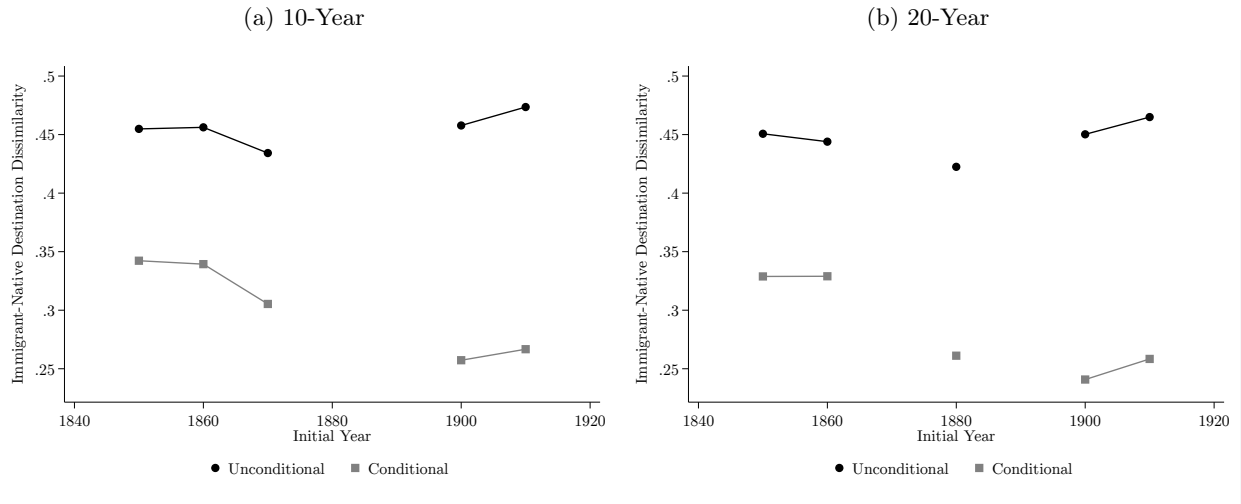


Figure 9: Dissimilarity in destination choices

Note: The unconditional results are dissimilarity indices in the destination county distribution of foreign- and native-born internal migrants for spans beginning in the year listed on the x-axis. The conditional results are computed as explained in footnote 50.

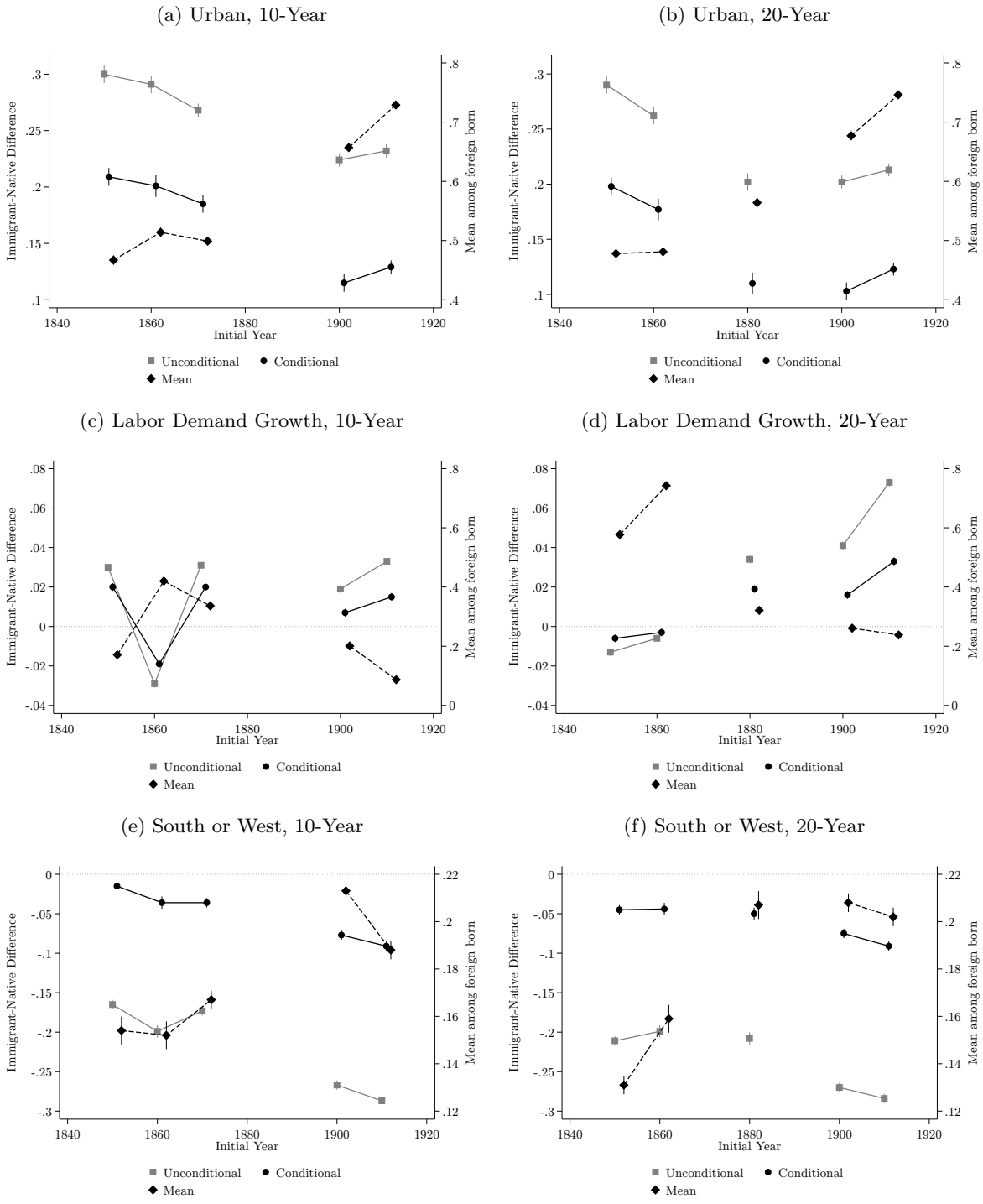


Figure 10: Immigrant-native differences in destination choice

Note: These figures plot the coefficients from a regression of the listed destination characteristic on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

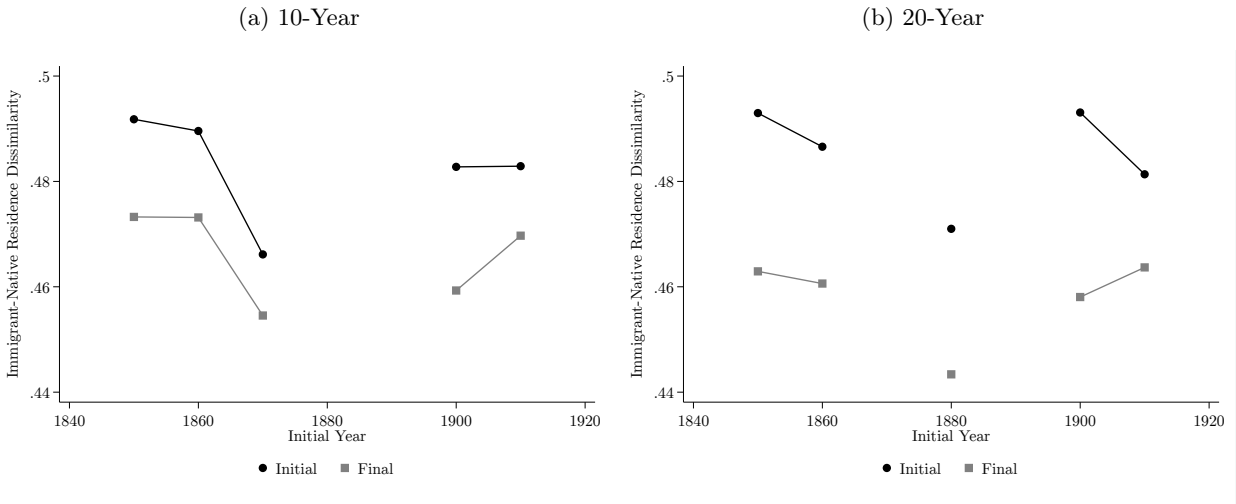


Figure 11: Dissimilarity in counties of residence

Note: These figures present the dissimilarity index between the counties of residence of the foreign and native born in my linked samples in the initial and final year of the span beginning in the year on the x-axis.

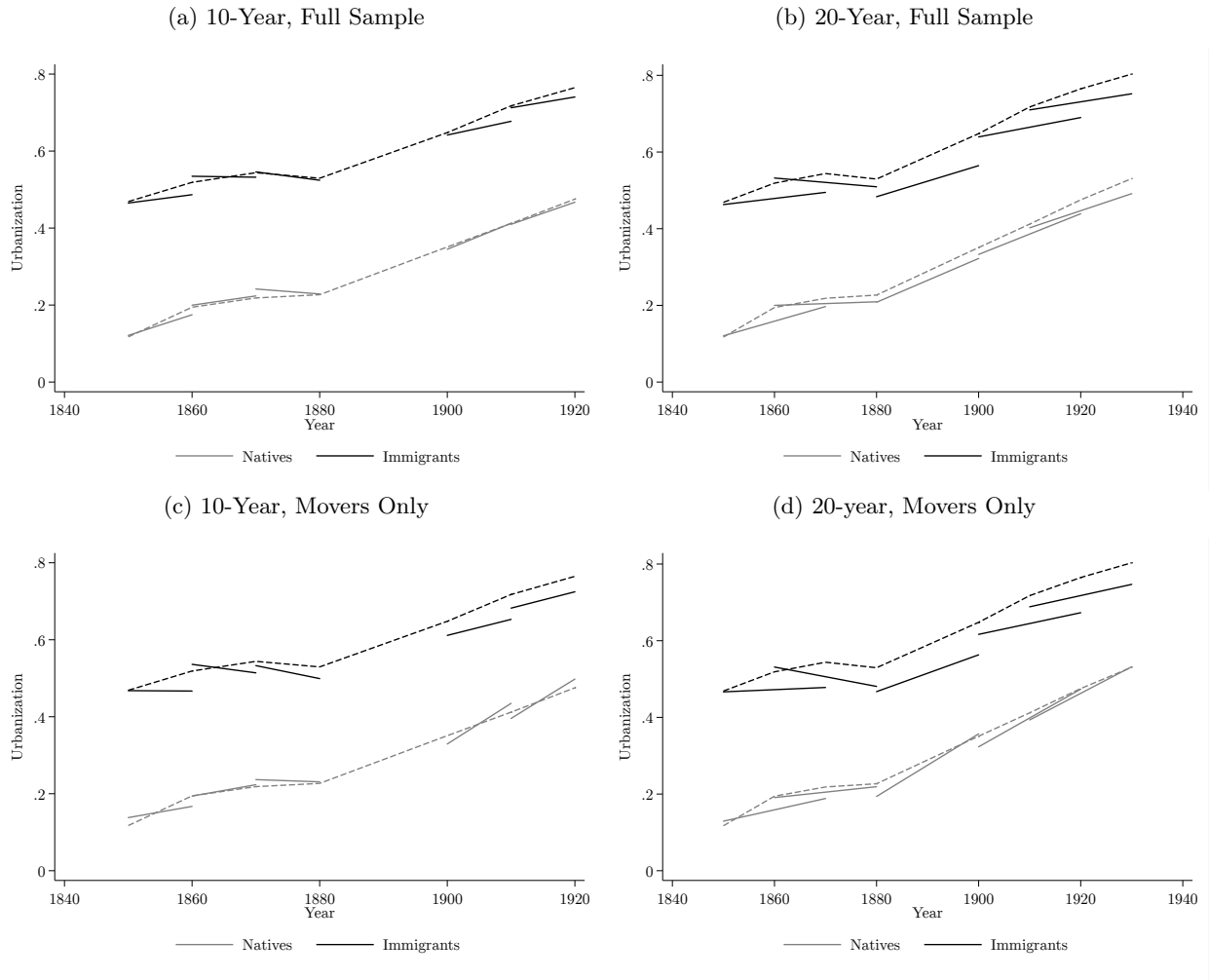
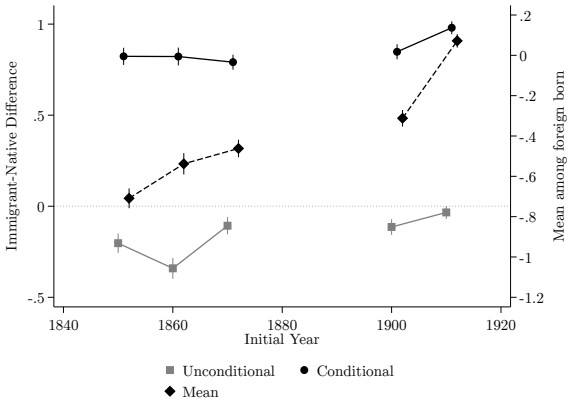


Figure 12: Urbanization of natives and immigrants

Note: The solid lines in these figures plot the urbanization rates of immigrants and natives in each of my linked samples in the initial and final year of each sample. The dashed lines present the general level of urbanization for each nativity group. Panels (a) and (b) include all individuals in the sample in the lines for the linked samples. Panels (c) and (d) limit these samples to movers only.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

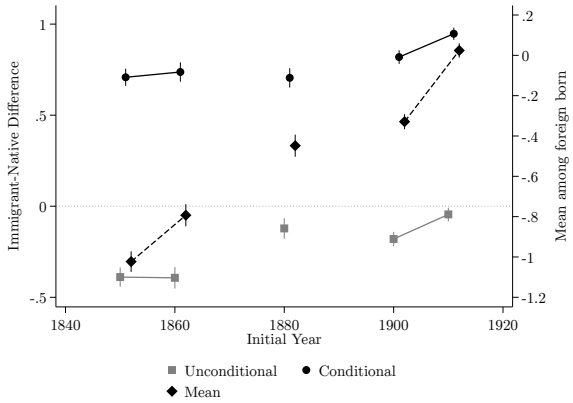


Figure 13: Immigrant-native differences in population density change

Note: These figures plot the coefficients from a regression of movers' change in population density on a foreign indicator. The conditional estimates also include the full battery of controls available for the sample. Each figure also presents the mean of the outcome for the foreign born. The year on the x-axis is the first year of the sample span. Observations weighted by inverse linkage probability. Vertical lines are 95-percent confidence intervals.

A Procedures for Linkage and Estimation of and Correction for False Match Rates

A.1 Record Linkage

The ABEE and ABEN linkage procedures are described in detail by Abramitzky et al. (2019b), and so I do not repeat them here, as my links were made using the code provided by those authors. The CZ method is as follows:[56]

1. I extracted males with non-empty first and last names from the full count censuses of 1850–1930 and removed punctuation.

2. I divided the listed first name into a given name and middle initial, when one was present.

3. I replaced standard first name abbreviations (e.g., "Wm" was replaced with "William").

4. I removed any remaining spaces from the names.

5. After steps 1–4, I linked each data set to itself according to the following criteria.

   (a) The birthplace (US state or foreign country) matches exactly.

   (b) The absolute difference in birth years is less than or equal to 4

   (c) The first three characters of the last name soundex match, and either of the following two conditions are true.

      i. The last name soundexes are identical and the SAS spelling distance (using the SAS function SPEDIS) is less than or equal to 20.[57]

      ii. The last name soundexes are non-identical, and the SAS spelling distance is less than or equal to 17.

   (d) First letter of first name matches, and spelling distance between first names is less than or equal to 20.

6. For each census, two versions were created, with names cleaned according to steps 1, 2, and 4 above (no standardization of name abbreviations was made):

[56] This description is closely based on the one provided by Collins and Zimran (2019, 2020), and is nearly identical to that of Zimran (2021), which uses the same linkage procedure but applies it only to the native born.

[57] The SAS spelling distance is not a symmetric measure. Whenever I use spelling distance, I make a match when $\min\{\text{spedis}_{a,b}, \text{spedis}_{b,a}\} \leq c$, where c is the linkage cutoff. The SPEDIS function is described in detail by Gershteyn (2000).

(a) One eliminating any individual who had a candidate match in step 5 that was someone other than himself. This is a "starting" sample. This was not created for 1940.

(b) One with all men. This is a "target" sample. This was not created for 1850.

7. I linked males from each starting sample to the target sample of each census 10 and 20 years later according to the same criteria listed in step 5, as well as the following additional criterion: where both records report a middle initial, the middle initials must match for a match to be made.

8. I dropped any case in which more than one individual in the starting sample matched to an individual in the target sample, or vice versa.

A.2 Estimation of False Match Rates

The procedure used to estimate the rates of false matching in linkage is based on the observation that, according to Bayes's Theorem, the probability that a pair of linked observations (i.e., one ostensibly matched individual) will be characterized as moving can be written as

$$P(m) = P(m|t)[1 - P(f)] + P(m|f)P(f), \tag{A.2}$$

where m denotes the binary outcome of observing a move (whether true or not), t denotes a true match and f denotes a false match. Ultimately, the object of interest in the analysis is to determine $P(m|t)$, but as a first step, it is necessary to compute $P(f)$. Solving for the probability of a false match, equation (A.2) can be written as

$$P(f) = \frac{P(m) - P(m|t)}{P(m|f) - P(m|t)}. \tag{A.3}$$

To estimate $P(f)$, I take advantage of the fact that, for a subset of the linked sample, it is possible to measure the probability of inter-state migration by linkage and by the ages and birthplaces of children, similar to Rosenbloom and Sundstrom (2004). That is, I create an alternative non-linkage based alternative measurement of inter-state migration by defining inter-state movers as those with a child aged less than 10 years born in a different state than the current state of residence and no older child born in the state of residence, and defining stayers as those with a child born in the state of residence at least 10 years old and no children younger than 10 born in a different state. This approach uses information in one census to determine migration over the previous 10 years (e.g., the 1860 census is informative of migration 1850–1860).

It is then possible to compute equation (A.3) by setting the linkage-implied inter-state migration rate

in the sample of linked individuals whose family structure also enables the application of the alternative method as $P(m)$, the birthplace-implied inter-state migration rate in this sample as $P(m|t)$, and computing $P(m|f)$ based on the distribution across states for each individual of individuals whose age and birthplace make them candidates for linkage to him. For example, for a man in the 1850 census born in New York in 1830 and living in New York in 1850, I determine $P(m|f)$ for an 1850-1860 link by determining, in 1860 the fraction of all New York-born men aged 30 (or sufficiently similar to this age based on each linkage method's linkage criteria) who lived outside New York.

This method makes the implicit assumption, by using the rate of inter-state migration as $P(m|t)$, that there is no error arising from using the birthplaces and ages of children to measure internal migration. Clearly there would be errors in this method, but as long as these errors are rare relative to those of the linkage method, it is unlikely to adversely affect results.

A.3 Correcting for False Matches

With the estimates of $P(f)$ from equation (A.3) in hand, the next step is to correct the estimated migration rates. An alternative rearrangement of equation (A.2) yields

$$P(m|t) = \frac{P(m) - P(m|f)P(f)}{1 - P(f)}, \tag{A.4}$$

which is an expression that can be used to correct the estimated rates of inter-county migration for false matches, yielding an estimate of the desired object $P(m|t)$. The components of this expression are arrived at as follows. The uncorrected probability of observing a move is $P(m)$. The estimated rate of false matching from equation (A.3) is $P(f)$. I use the $P(f)$ calculated for the 10-year span beginning in the same year as the span in question, since the procedure to calculate false match rates can be applied only in 10-year spans.[58] The probability that a false match would yield an observation of inter-county migration is $P(m|f)$; in most cases this is essentially 1, and so I simply set it to that value.[59] The exception is in Online Appendix E.1, as I discuss further there.

In practice, this approach uses the estimated rate of false matching from the subsample for whom inter-state migration could be determined by linkage and by household composition and applies it to the complete sample, distinguishing between immigrants and natives. As a result of the different sample used to compute

[58] So, for instance, I use the estimated false match rate for 1850–1860 to correct the estimated inter-county migration rates for the 1850–1860 and 1850–1870 spans.

[59] In principle I could estimate this value as described in Online Appendix E.1, but the estimate is so close to one for inter-county migration that it is not necessary to do so.

equations (A.3) and (A.4), it is possible that the correction is incomplete, but the method certainly represents an improvement over the uncorrected estimates, and the similarity between the estimated migration rates for natives across all methods after correction helps to point to the value of the method. Indeed, when applied to natives in Online Appendix C and by Zimran (2021), the estimated inter-county migration rates of natives become very similar across linkage methods.

B Contemporary Quotations Regarding Immigrant Distribution

[i] [on] the matter of distribution, we see a marked difference between the immigration prior to 1870, which built up the Northwest, and the races which now come to us (Hall 1904, p. 176)

it is clear that the tendency of the new immigration is toward industrial and city pursuits rather than toward agriculture. (US Congress 1911a, p. 40)

In making the larger cities and industrial communities their place of residence, aliens composing the new immigration movement have continued to follow a tendency which originated with the advent of such immigrants in considerable numbers. (US Congress 1911a, p. 40)

The great bulk of the present immigration settles in four of the Eastern States, and most of it in the large cities of those States. Notwithstanding the well-known demand for agricultural labor in the Western States, thousands of foreigners keep pouring into our cities, declining to go where they might be wanted . . .

(Commissioner-General of Immigration 1903, p. 70)

Landing in the large cities, [immigrants] seldom move on to the greater air spaces of the country.

(Whelpley 1905, p. 319)

[ii] taken as a whole, the element of foreign birth seeks the cities with far greater avidity than does the element of native birth. (Department of the Interior 1895, p. lxxxix)

[iii] A large part of the immigrants were agricultural laborers at home, and their immigration is due to a desire to escape the low economic conditions which attend agricultural pursuits in the countries from which they come. With no knowledge of other conditions it is natural, therefore, that they should seek another line of activity in this country. (US Congress 1911a, p. 40)

The destination of these immigrants in the United States on arrival is controlled by the fact that they almost invariably join relatives or friends, and few of these, even among earlier immigrants of the class, are engaged in agricultural pursuits. (US Congress 1911a, p. 40)

[iv] [Immigrants had] no choice of location; they came to the only spot in America with which they were familiar by report. They knew nothing of opportunities elsewhere in the United States, and in many cases they still know nothing of the possibilities outside of the section of the city where they live. (US Congress 1911b, p. 146)

[v] [The new immigrants] are neither physically nor mentally fitted to go to these undeveloped parts of our country and do as did the early settlers from northern Europe . . . Past immigration was good because most of it was of the right kind and went to the right place. (Commissioner-General of Immigration 1903, p. 70)

[vi] [Have immigrants] come with the purpose of making the United States their residence? . . . Have they brought their wives? Have they brought their children? Are they actuated by a desire to enter into our life and to find their development under American institutions? In other words, have they sought the soil, or are they the denizens of the cities, belonging to a floating population unrecognized by the permanent population except as so many labor units? . . . [The old immigrants] sought the soil and open[ed] up farms. What about the new?

(Senator William Dillingham, 1914, in Benton-Cohen 2018, p. 233)

[vii]

True Americanism opposes the further use of Western lands to invite immigrants

(Senator Henry Cabot Lodge, in Benton-Cohen 2018, p. 11)

viii some pressure must be brought to bear upon the immigrants to secure distribution because under the present system they do not voluntarily distribute themselves. (Hall 1906, p. 300)

Of course, it is desirable that even the right kind of immigration should be properly distributed in this country . . . As much as possible should be done to distribute the immigrants upon the land and keep them away from the contested tenement-house districts of the great cities. (Roosevelt 1905)

ix Change of location from the Eastern States to other States will not change the character and tendencies of an immigrant even as much as the change from Europe. Unless the distribution is very wisely done, and possibly even then, the result will be the spreading of big slums over the country in the form of little slums . . . distribution from the large seaports acts as a force pump to draw in even larger numbers, so that the situation might be even worse than at presence. (Statement of the Immigration Restriction League, in US Congress 1911c, p. 109)

x we cannot have too much immigration of the right sort and we should have none whatever of the wrong sort. Of course, it is desirable that even the right kind of immigration should be properly distributed in this country. . . . But distribution is a palliative, not a cure. The prime need is to keep out all immigrants who will not make good American citizens. (Roosevelt 1905)

xi The systematic distribution of the incoming masses of immigrants tends to make immigration healthy and desirable. There is enough room in this country for millions more inhabitants, provided they are fairly distributed and are not allowed to clog up any one particular point. By judicious distribution only the benefits of immigration will be felt and at the same time conditions in the port of entry will be greatly relieved.
(Statement of David M. Bressler, General Manager of the Industrial Removal Office, in US Congress 1911c, p. 305)

There is no cause for fear on the part of the South or Southwest that they will have "aliens dumped on them," as has been intimated, and I may say in passing that, while alarmists and bigots are protesting against certain people being sent to the South, the agriculturists of the South are clamoring for industrious, sober, honest workingmen regardless of their race or religion.

(Terence Powderly, head of the Division of Information, in US Congress 1911c, p. 389)

Congestion of immigrants in any one locality is not desirable; to keep the stream moving in steady current to places where the labor field is not crowded will be good for immigrant and country.

(Terence Powderly, head of the Division of Information, in US Congress 1911c, p. 416)

xii Since we cannot depend on the immigrants to scatter, means must be taken to diffuse them throughout the country and to localize them away from the great cities. (Norton 1904, p. 161)

xiii If this work of informing the immigrants [of distribution opportunities] is worth doing—and no one knowing the problem doubts it—let it be done on a scale worthy of the dignity of the United States and commensurate to the need. . . . It is a reflection upon our civilization to have a thousand men idle in Pittsburg and a crying need for their labor in Minnesota. The Government could obviate this anomaly and be of great service to immigrants
(Statement of the International Committee of Young Men's Christian Associations, in US Congress 1911c, p. 86)

xiv

The problem of immigration with us is essentially one of distribution. (Speranza 1904, p. 930)

It [was] . . . not a question of too many immigrants, but of improper distribution of them.
(Senator LeRoy Percy, 1907, in Benton-Cohen 2018, p. 203)

This so-called white immigration problem [is] a question largely of proper distribution
(Congressman Anthony Michalek, 1906, in Benton-Cohen 2018, p. 83)

[xv] We need more of such [right kind of] immigration for the South; and special effort should be made to secure it. (Roosevelt 1905)

[xvi] Since the distribution of the thrifty immigrant to sections of the country where he may secure a permanent residence to the best advantage, and especially where he may invest his savings in farms or engage in agricultural pursuits, is most desirable, the Division of Information, in the Bureau of Immigration and Naturalization, should be conducted as to cooperate with States desiring immigrant settlers; and information concerning the opportunities for settlement should be brought to the attention of immigrants in industrial centers who have been here for some time and who thus might be induced to invest their savings in this country and become permanent agricultural settlers. The division might also secure and furnish to all laborers alike information showing opportunities for permanent employment in various sections of the country, together with economic conditions in such places.
(US Congress 1911a, p. 46)

[xvii] the old immigration came to be a part of the country, while the new, in a large measure, comes with the intention of profiting, in a pecuniary way, by the superior advantages of the new world and then returning to the old country. (US Congress 1911a, p. 14)

As far as possible, the aliens excluded should be those who come to this country with no intention to become American citizens or even to maintain a permanent residence here. (US Congress 1911a, p. 47)

[xviii] No satisfactory or permanent distribution of immigrants can be effected through any federal employment system, no matter how widespread, because the individual will seek such social and economic conditions as best suit him, no matter where sent. (US Congress 1911a, p. 40)