NBER WORKING PAPER SERIES

US IMMIGRANTS' SECONDARY MIGRATION AND GEOGRAPHIC ASSIMILATION 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, Revised March 2022

For helpful comments, I thank the editor, Marianne Wanamaker, as well as two anonymous referees, 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. 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 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. This paper previously circulated under the title "Immigrant Distribution in the United States during the Age of Mass Migration." The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

US Immigrants' Secondary Migration and Geographic Assimilation during the Age of Mass Migration

Ariell Zimran

NBER Working Paper No. 28812
May 2021, Revised March 2022
JEL No. F22,J11,J15,J61,N31,N32,R23

ABSTRACT

I study the rates of, selection into, and sorting of European immigrants' secondary migration within the United States and their geographic assimilation during the Age of Mass Migration. These phenomena are recognized as important components of the economics of immigration, but data constraints have limited prior study of them in this context. As part of the debate over immigrant distribution, they were also major issues in the broader twentieth-century immigration policy debate, which were influenced by the widely held view that immigrants in the early twentieth century were less geographically mobile and specifically more attached to urban areas than were natives and earlier immigrants. I find that immigrants throughout the Age of Mass Migration were at least as likely as natives to make inter-county moves, were more attached to urban areas, were more likely to move to urban destinations, and shared natives' increasing attachment to urban areas over time. In spite of their mobility, immigrants experienced relatively little assimilation in their place-of-residence distributions relative to natives with time in the United States, though they did experience somewhat more convergence on natives in terms of urbanization. These results help to better understand immigrant assimilation and the effects of immigration during the Age of Mass Migration and imply that the contemporary views of immigrant immobility were either false, oversimplified, or the product of changes in the US economy.

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

Published in 1911, the report of the Immigration Commission, commonly called the Dillingham Commission, was a landmark analysis of immigration, and contained a number of policy recommendations intended to address the so-called "immigration problem" (US Congress 1911a). These proposals guided immigration policy for the next decade, and most of them eventually became law (Benton-Cohen 2018), including the imposition of a literacy test and national origin quotas. Another of the commission's proposals was that efforts be made to achieve *immigrant distribution*—that is, that immigrants be encouraged to leave coastal cities and move inland. This recommendation, as well as the calls of many influential observers and policymakers that it echoed, was based on the widespread belief that the *new immigrants* of the early twentieth century were generally immobile, and specifically reluctant to leave urban areas, exhibiting a greater attachment than natives and the earlier waves of *old immigrants*.[1] While some argued that immigrants' supposed geographic immobility strengthened the case for immigration restriction, many supporters of immigration also accepted that immigrants were immobile. They saw immigrant 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.[2] As these were among the reasons cited as justification for immigration restriction (e.g., Benton-Cohen 2018; Commissioner-General of Immigration 1903; Goldin 1994; Hall 1904; US Congress 1911a; Willcox 1906), successful distribution was seen as a way to keep the Golden Door open to immigration in the face of rising calls to close it (Benton-Cohen 2018).

This historical focus on immigrants' *secondary migration* within the destination country mirrors the economics of migration's recognition of the important role of this force in any episode of immigration. Just as immigrants may converge economically and culturally on natives with time in the destination (Abramitzky and Boustan 2017; Borjas 2014), their secondary migration can result in *geographic assimilation*—convergence of immigrants' and natives' place-of-residence characteristics and distributions with time in the destination (e.g., Bartel and Koch 1991; Beenstock 1997, 1999; Belanger and Rogers 1992; Hall 2009; Kritz and Nogle 1994; Newbold 1999). Secondary migration is thus an important determinant of immigrant assimilation. Moreover, by determining the final location distributions of immigrants in the destination (Bartel 1989; Bartel and Koch 1991; Kritz and Gurak 2001), and specifically by enabling a group of potentially highly mobile

[1] 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.

[2] 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.

individuals to respond to differences in labor market opportunities over space (Basso, D'Amuri, and Peri 2019; Basso and Peri 2020; Borjas 2001; Cadena and Kovak 2016; Özgüzel 2021; Schündeln 2014), secondary migration contributes to determining the effect of immigration on the receiving economy.

Given its role in the economics of immigration and its position in the early twentieth-century immigration policy debate, immigrants' secondary migration is an important component of a complete economic history of the Age of Mass Migration. But difficulties in measuring internal migration in this period in general and for immigrants in particular have constrained prior scholars' ability to study secondary migration and geographic assimilation in general and specifically to interrogate the perceptions of immigrant immobility that motivated the contemporary focus on immigrant distribution. In this paper, I bring these important but understudied aspects of the economics of the Age of Mass Migration into focus for the first time, and shed light on contemporary claims of immigrant immobility, investigating them in greater detail than has previously been possible. In particular, I take advantage of recent data advances to document the rates of, selection into, and sorting of immigrants' secondary migration and the degree of immigrants' geographic assimilation in the Age of Mass Migration, with coverage of the complete period and of European immigrants of all origins.

My analysis is based on 10 datasets linking European- and native-born white men aged 18–40 from the complete-count US censuses of 1850–1910 to those taken 10 and 20 years later. Linked data are necessary to study secondary migration because no other source provides systematic information on immigrants' place of residence in the United States at two points in time. Moreover, 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 (Abramitzky, Boustan, and Eriksson 2014; Lubotsky 2007). But the potential pitfalls of these data, especially the danger of errors in linking (Abramitzky et al. 2021b; Anbinder et al. 2021; Bailey et al. 2020), pose the main challenge to my analysis. To ensure that my results are not driven by problems inherent to linked data, I test the robustness of my results to linkage methods of varying 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 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.[3]

[3]Notwithstanding any of their limitations, the linked datasets that I construct provide a view of immigrants' secondary 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.

My first focus is on the frequency of immigrants' secondary migration, with a focus on inter-county moves. I document a large decline in immigrants' inter-county migration rates by about 15 percentage points over the Age of Mass Migration from an initial level of about 50–60 percent, depending on the linkage method and length of period. This decline occurred almost entirely at the beginning of the Age of Mass Migration, from the cohort first observed in 1850 to that first observed in 1860. From this point to the end of the Age of Mass Migration, immigrants' secondary migration rates were largely constant or even slightly rising. I also compare immigrants' secondary migration rates to natives' internal migration rates, finding no evidence that immigrants were ever meaningfully less mobile than natives, with most estimates indicating the opposite.

Next, I measure immigrants' selection into secondary migration, and contextualize these results by comparing them to natives' selection into internal migration. I find evidence of negative selection into secondary migration on the basis of literacy and occupational rank. I also study selection into inter-county migration on the basis of urban status. I find that both immigrants' and natives' attachment to urban areas increased over the Age of Mass Migration, as measured by the relationship between initial urban status and the probability of inter-county migration. I also find that the foreign born were more attached to urban areas than were natives through the whole of the Age of Mass Migration

Such greater attachment of immigrants to urban areas was also evident in the destination choices of internal 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, and that this was also true throughout the Age of Mass Migration. More generally, I find that 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.

Finally, despite the relatively high mobility of immigrants in the United States after arrival, I find that the geographic assimilation of immigrants relative to natives was limited. Immigrants closed at most 5 percent of the gap in their county-of-residence distributions relative to natives over 10years, and at most 8 percent over 20 years. But the combination of immigrants' secondary migration and natives' internal migration did lead immigrants to experience somewhat more convergence on natives in terms of urbanization with time in the United States, albeit incompletely, closing up to 14 percent of the initial gap in urban status over 10 years, and up to 21 percent over 20 years.

This paper advances the economics of immigration and the economic history of the Age of Mass Migration in a number of ways. It constitutes the first systematic analysis of immigrants' secondary migration in the Age of Mass Migration, contributing to the broader literature on secondary migration by extending

it to cover the period in which US immigration was at its peak and borders were open. Moreover, by documenting, for the first time, the rates of, selection into, and sorting of immigrants' secondary migration for European immigrants of all origins throughout the whole Age of Mass Migration, this paper sheds light on a phenomenon that the modern literature has identified as a crucial component of any episode of migration. It thus contributes to a more complete economic history of this period. More specifically, by showing that immigrants were at least as internally mobile as natives, and in an urbanizing economy shared natives' increasing attachment and attraction to urban areas—and indeed were more attached and attracted to such areas—the results of this paper are consistent with the immigrants of the Age of Mass Migration sharing modern immigrants' greater responsiveness to spatial differences in economic opportunity. This paper thus contributes to better understanding the effects of immigration in this period (e.g., Basso and Peri 2020; Cadena and Kovak 2016; Schündeln 2014).

This paper also adds to the literature on immigrant assimilation in the Age of Mass Migration.[4] The focus of this literature has largely been on assimilation along economic (e.g., Abramitzky, Boustan, and Eriksson 2014; Collins and Zimran 2021) and cultural (e.g., Abramitzky, Boustan, and Eriksson 2020; Ward 2019) lines. Assimilation in the dimension of place of residence has also been recognized as important, leading to studies of immigrants' spatial assimilation within cities and of immigrant enclaves (e.g., Eriksson 2019; Eriksson and Ward 2019, 2020). This paper's evidence on immigrants' geographic assimilation thus expands the literature's understanding of immigrant assimilation during the Age of Mass Migration. Immigrants' limited convergence on natives' place-of-residence distribution contrasts with findings of substantial cultural assimilation (e.g., Abramitzky, Boustan, and Eriksson 2020) and economic assimilation or similarity (Abramitzky, Boustan, and Eriksson 2014; Collins and Zimran 2021) in this period. Together with Eriksson and Ward's (2019) results on limited spatial assimilation of immigrants relative to natives *within* cities, this finding shows that immigrants' convergence on the experience of natives with time in the United States was limited in one important dimension.

The paper also complements the literature on internal migration in the United States, and especially 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 (c.f.,

[4]Through this contribution, the paper also contributes more broadly to the literature on modern immigrants' assimilation (e.g., Borjas 1985; Card 2005; Lubotsky 2007), to the modern literature on spatial aspects of immigrant assimilation (Bartel and Koch 1991; Beenstock 1997, 1999; Belanger and Rogers 1992; Hall 2009; Kritz and Nogle 1994; Newbold 1999) and, given its relationship to these literatures, to the literature on immigrant enclaves (e.g., Cascio and Lewis 2012; Cutler, Glaeser, and Vigdor 2008a,b; Damm 2009; Edin, Frederiksson, and Åslund 2003).

Basso and Peri 2020). This paper thus adds to 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 most closely 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 historic US internal migration (e.g., Collins and Wanamaker 2015; Hornbeck 2020; Sichko 2021; Stewart 2006).

Most directly, this paper speaks to the historic debate over immigrant distribution. It builds on a small set of 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). Despite these studies, a complete picture, such as this paper provides, of immigrant distribution over the entirety of the Age of Mass Migration, is lacking. Based on the results described above, I draw the following conclusions regarding the views of immigrant immobility that pervaded the early twentieth-century immigration policy debate. Immigrants' high geographic mobility relative to natives contradicts claims that immigrants were less mobile than natives. And while it was technically true that secondary migration rates had declined over time, the decline was almost entirely between the first and second large cohorts of immigrants, meaning that the characterization of twentieth-century immigrants as less mobile than earlier immigrants was an oversimplification—they were no less mobile than any previous cohort of immigrants in the Age of Mass Migration other than the very first. Immigrants' negative selection into secondary migration also contradicts claims that it was the poor, illiterate, and unskilled immigrants who remained in arrival cities.

Finally, my findings on selection and destination choice give credence, on their face to claims that immigrants in the twentieth century were more attached to urban areas than were natives and earlier immigrants. But, again, the contemporary interpretation of these patterns was oversimplified. Immigrants were more attached and attracted to urban areas than were natives *throughout* the Age of Mass Migration, not just in the twentieth century. Moreover, immigrants' increasing attachment and attraction to urban areas over time was not unique to them, but was in fact shared with natives, suggesting that immigrants' growing urbanization was driven by the changing US economy rather than the changing character of immigrants. Indeed, the difference in immigrant and native internal migrants' probability of choosing an urban destination was at its lowest in the early twentieth century. As a result, these patterns also shed light on the failure and eventual abandonment of distribution efforts—economic opportunities were most abundant in urban areas, and the success of distribution efforts hinged on immigrants defying the incentive to move to and remain in cities. Immigrants were not concentrated in urban areas because they were immobile, but because these areas offered them greater opportunities.

The corpus of evidence gathered in this paper 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.

On the whole, given that immigrant distribution, geographic assimilation, and secondary migration were important in both the economics and the politics of the Age of Mass Migration, this paper contributes to a more complete understanding of the economics of the Age of Mass Migration and 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.

2 Background

2.1 Immigrant Distribution in the Age of Mass Migration

Immigrant distribution was a central issue in the early twentieth-century immigration policy debate.[5] 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. In particular, it was widely held that the new immigrants arriving from southern and eastern Europe in the late nineteenth and early twentieth centuries were immobile, especially in comparison with the old immigrants arriving from northern and western Europe earlier in the nineteenth century. 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]

Immigrant distribution also impacted the debate in the normative sense, with the broader set of views on immigration including whether immigrants should have moved within the United States, whether they should be encouraged to move, and what impacts their moving would have. In particular, a number of policy prescriptions grew out of the positive views of immigrant distribution described above, combined with normative views on the matter. The strongest advocates of immigration restriction in the public and in government, including President Theodore Roosevelt, saw immigrant immobility as yet more evidence

[5]Contemporary quotations illustrating the role of distribution and related issues in the policy debates are provided in Appendix B, and are indicated by roman numeral endnotes. See also Willcox (1906, pp. 523–524).

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 it.[xii] Many even viewed 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). There were even those who believed that the problems of immigration could be solved *entirely* by distribution.[xiv] Distribution was also favored by regions, particularly the South and West, that did not attract many immigrants but sought to do so.[xv]

For opponents and proponents of immigration restriction alike, concerns about immigrant distribution were intimately linked to concerns about the contemporaneous growth of cities in the United States. Indeed, many of the progressives who were concerned about immigration were also noted anti-urbanists and urban reformers (Benton-Cohen 2018). In particular, 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.

On the whole, distribution efforts were popular "because they appealed to urban immigrant constituencies [because they might dampen calls for restriction] and to western and southern business interests" and "offered a palatable solution to the inseparable concerns about immigration and rapid urbanization" (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), 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). These programs included the Industrial Removal Office and the Galveston Movement among Jews and the Sunnyside Plantation among Italians. But all of these failed, in some cases leading some of the Dillingham Commission's members to give up on distribution and favor restriction (Benton-Cohen 2018, p. 233).

Willcox (1906), however, realized that the complaints about immigrant mobility and the resultant calls

for immigrant distribution were based on a fundamental misunderstanding and miscontextualization of immigrants' internal migration patterns in the United States.[6] Using a variety of data from the published volumes of the 1890 and 1900 censuses, Willcox (1906) challenged the perceptions of immigrant immobility motivating the advocates of distribution efforts. 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.

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 published volumes of the census, as the constraints of the time required, Willcox (1906) was not actually able to directly observe migration, 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. Thus, while Willcox's (1906) analysis casts doubt on contemporary views of immigrant distribution, it is not dispositive.[7]

Modern studies have used census microdata to delve more deeply into historical immigrants' secondary 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.[8] Rosenbloom and Sundstrom (2004), using census information on children's ages and birthplaces to infer parents' inter-state migration, find a negligible difference in the migration rates of the native and foreign born. But studying the secondary migration of a more expansive group of immigrants requires linked census data, as explained in section 3 below.

Ferrie (1995, 1999) pioneered the use of such data to study immigrants' internal mobility, giving evidence

[6]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 support distribution efforts. 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.

[7]Indeed, I will show that some of his claims, such as that there was "no evidence on [immigrants'] 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.

[8]Eriksson (2019) and Spitzer (2019) have used census microdata to study the settlement patterns of Norwegian and Russian Jewish immigrants, respectively, and Abramitzky and Boustan (2017) have studied immigrant settlement patterns more broadly. But none of these papers study secondary migration.

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 linked census data to compare the rates of inter-state migration of the native and foreign born in the early twentieth century.[9] 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). But 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' secondary migration during the Age of Mass Migration, nor is there an existing study comparing immigrants' secondary migration to natives' internal migration or a systematic study of the distribution question in this period. Given the importance of this phenomenon in the economics and politics of the period, the lack of evidence on it leaves an important gap in economic historians' understanding of the Age of Mass Migration, which this paper aims to fill.

2.2 Secondary Migration and the Economics of Immigration

Immigrants' secondary migration is recognized as an important aspect of the economics of migration. One strand of literature links secondary migration to immigrants' assimilation more broadly (e.g., Bartel and Koch 1991; Beenstock 1997, 1999; Belanger and Rogers 1992; Hall 2009; Kritz and Nogle 1994; Newbold 1999). Insofar as convergence of immigrants' and natives' place-of-residence distributions is a form of immigrant assimilation, secondary migration enables immigrants to become more similar to natives in this regard, though it is also possible that secondary migration can lead to divergence (Kritz and Nogle 1994). Therefore, beyond the link to the literature on immigrants' assimilation (Borjas 1985; Card 2005; Lubotsky 2007), there is also a clear relationship between immigrants' geographic assimilation driven by secondary migration and the literature on immigrant enclaves and residential segregation within cities (Cascio and Lewis 2012; Cutler, Glaeser, and Vigdor 2008a,b; Damm 2009; Edin, Frederiksson, and Åslund 2003).[10] In an effort to better understand the drivers of immigrant assimilation, and more generally, building on the argument that the effects of immigration are determined by immigrants' final settlement patterns in the destination after secondary migration, this literature has also been interested in the determinants of secondary migration (e.g., Bartel 1989; Bartel and Koch 1991; Basso and Peri 2020; Beenstock 1997, 1999; Belanger and Rogers

[9] 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.

[10] I refer to the convergence of place-of-residence distributions between immigrants and natives as *geographic assimilation* and, following Eriksson and Ward (2019) to a reduction in within-city segregation as *spatial assimilation*. This is different from Hall (2009), who defines what I call geographic assimilation as spatial assimilation.

1992; Kritz and Gurak 2001; Kritz, Gurak, and Lee 2011; Kritz and Nogle 1994; Rashid 2009). For instance, some studies have investigated the relationship between education and secondary migration (Bartel 1989) to determine the potential impacts of selective immigration policy on immigrants' geographic assimilation and to better understand the potential impacts on natives of various skill groups (Gurak and Kritz 2000; Kritz, Gurak, and Lee 2011; Rashid 2009).

Beyond simply determining immigrants' settlement patterns, a more specific mechanism by which immigrants' secondary migration contributes to determining the effects of immigration has been recognized as a result of Borjas (2001), who argued that immigrants "grease the wheels" of the labor market in the destination. In particular, Borjas (2001) argued that immigrants, having already incurred the fixed cost of migration, were particularly responsive (and specifically more responsive than were natives) to spatial differences in economic opportunity within the destination country. This greater responsiveness would, in turn, lead to a faster arbitrage of regional differences, and thus to a more rapid elimination of misallocation than would occur in the absence of immigrants.

Differences in immigrants' and natives' responsiveness to the incentive to move within the destination country could arise from differences in observable characteristics, but could also be the product of immigrants simply being more mobile "types" than natives (e.g., Friedberg and Hunt 1995), either because of a lack of attachment to their place of residence in the destination or because of selection into immigration on the basis of risk aversion, taste for mobility, or some other factor (Schündeln 2014). As a result, part of this literature has sought to determine whether differences in immigrants' and natives' propensity to move can be explained by differences in their observables (e.g., Gurak and Kritz 2000; Schündeln 2014).

Given Borjas's (2001) emphasis on it, a literature developed documenting the greater responsiveness of immigrants to differences in economic conditions across potential destinations in the host country (Basso and Peri 2020). One line of research studies the responsiveness of immigrants' initial destination choice to differences in economic opportunities (e.g., Cadena 2013, 2014). More relevant to the present paper, a second line of research focuses on the responsiveness of immigrants' secondary migration to economic conditions, in general and specifically in comparison to that of natives' internal migration (Basso and Peri 2020; Gurak and Kritz 2000; Gutiérrez-Portilla, Maza, and Hierro 2018; Kritz and Gurak 2001; Kritz, Gurak, and Lee 2011; Rashid 2009; Schündeln 2014). A robust finding of this literature is that immigrants, and in particular recent immigrants, are more responsive in their secondary migration to differences in economic conditions than are natives in their internal migration, supporting Borjas's (2001) argument.

Cadena and Kovak (2016) have recently provided an explicit test of the role of secondary migration in

equilibrating shocks in the United States, as have Basso, D'Amuri, and Peri (2019) and Özgüzel (2021) for Europe. These papers show that, not only did the location choices of new and secondary immigrants respond to differences in these conditions, but that their responses are sufficient to reduce the incidence of local labor demand shocks on natives.[11] Indeed, Cadena and Kovak (2016) show that 75 percent of the geographic responsiveness of immigration to shocks comes from secondary and return migration rather than from new migration. Together, these findings point to a potentially important role for immigrants' secondary migration in determining the effects of migration.

This literature demonstrates the importance of secondary migration in the economics of immigration. Despite this, there has been, as discussed above, no inquiry into the secondary migration of immigrants in the Age of Mass Migration, save a few studies of limited scope. But it is an aspect of the Age of Mass Migration that must be understood in order to better understand the economics of immigration in this period in general, and specifically to shed light on immigrant assimilation and the effects of immigration. This paper contributes to incorporating secondary migration into the study of the Age of Mass Migration.

This paper also adds to the literature on the economics of secondary migration. Given the prevalence of restrictions on immigration in modern contexts, it is unclear how much of what is known about secondary migration is the product of these restrictions as opposed to the underlying economics of interest. By studying secondary migration in a context without meaningful immigration restrictions, this paper expands the economics of immigration (Abramitzky and Boustan 2017; Spitzer and Zimran 2018).

3 Data

My analysis relies on complete-count census data for the US censuses of 1850–1930 provided by Ruggles et al. (2021). Although none of these censuses inquired directly about internal migration,[12] 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

[11] Amior (2021) argues that immigrants are indeed highly responsive to differences in economic status, but argues that this mobility is the product of network effects rather than a responsiveness to conditions per se, and often leads to crowd-out of natives, so that the net effects are minimal. Nonetheless, Basso and Peri (2020) point out that in either case, immigrants as a whole, through their initial, secondary, and return migration, benefit the native-born as a group, either by smoothing shocks as Borjas (2001) and Cadena and Kovak (2016) suggest, or by enabling natives to avoid the costs of migrating to do so, if, as Amior (2021) suggests, the impacts are largely through displacement.

[12] It is also not possible to use birth state-residence state comparisons to determine internal migration, as has been done for natives (e.g., Ferrie 1997, 2006a,b; Hall and Ruggles 2004), since birthplace is not informative of immigrants' prior place of residence in the United States. It is also tempting to presume that any immigrant not living in a port of arrival is an internal migrant; but it is not clear whether the immigrant went to this place immediately after arrival or first settled somewhere else before moving again. Moreover, immigrants living in arrival cities may have settled there only after a period of residence elsewhere in the United States.

the two.[13] Linkage across censuses is also necessary to measure geographic assimilation—as in any study of immigrant assimilation in the Age of Mass Migration (e.g., Abramitzky, Boustan, and Eriksson 2014; Collins and Zimran 2021)—in order to avoid biases arising from return migration and from differences across cohorts in place of residence.[14] The individual's observables in the first census also provide data on pre-(secondary) migration characteristics. The exclusion of eventual return migrants, though necessary, 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, migration, and geographic assimilation of persistent immigrants.

I construct 10 distinct census-to-census linked datasets encompassing all possible 10- and 20-year links from initial censuses 1850–1910 to final censuses 1860–1930. I do this by merging the complete-count census files provided by Ruggles et al. (2021) with linkage crosswalks that I created according to the method described in Appendix A and which have been used also by Collins and Zimran (2019, 2021) and Zimran (2021).[15] In essence, this method follows the example of Ferrie's (1996) pioneering matching approach and best practices outlined by Abramitzky et al. (2021b) and Bailey et al. (2020) to make links on the basis of fixed and pre-determined rules, using string distance measures to make comparisons of names (to address potential differences in spelling across censuses), and removing cases where individual records are not unique.

Throughout the analysis, 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

[13] My focus is on inter-county migration, since the county is the finest unit of place of residence that can be observed in all censuses. 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). I address this by defining migration as a move from one county in the initial census to one in the final census that has no geographic overlap with it. An example is given in Online Appendix D.

[14] The danger posed by return migration is that, if immigrants living in urban areas were more likely to return to their countries of origin than were immigrants living in rural areas, immigrants would appear to become less urban over time even if they had not engaged in any secondary migration; more generally, such a movement could result in an appearance of geographic assimilation. The danger coming from differences across immigrant cohorts is that it is necessary to distinguish between individuals living in cities because they had just arrived there and those who had remained there for some time.

[15] Online Appendix Table C.1 presents linkage rates for the native and foreign born for each sample.

(Abramitzky and Boustan 2017; Hatton and Ward 2019).[16] 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.

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.[17] The 1850–1910 coverage of initial censuses concentrates on the period of essentially unrestricted migration.

Recent research has recognized two primary difficulties in the use of linked data. The first, highlighted by Bailey et al. (2020) in general and by Anbinder et al. (2021) for immigrants in particular,[18] 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. Besides causing me to overstate internal migration rates generally, this issue has specific potential impacts on each part of my analysis. Falling rates of false linkage over time and greater rates of false linkage among immigrants than natives (both consistent with my findings below) can lead to spurious indications of falling migration rates over time and greater mobility by immigrants than natives, respectively. Negative selection into false matches can also lead to spurious indications of negative selection into internal migration. Similarly, over-representation of urbanites among the incorrectly linked can lead to spurious indications that urbanites were more mobile. Finally, since false matches are likely to generate spurious links to areas where there are more potential links, they are likely to result in urban destinations appearing more popular than they truly were. Since my results are in many cases—but not always—consistent with those that would be expected to arise as a result of false matches, it is important to be particularly vigilant in ensuring that my results are not simply the product of false matches.

I use two strategies to address the challenge of false matches. First, in Online Appendix E, I repeat the analysis using four alternate linkage methods, and in the main text I draw only conclusions that are qualitatively robust to the choice of linkage method. Two of the alternate linkage methods are the ABE-

[16]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

[17]Of course, the 1890 census is excluded because its data have not survived.

[18]It is reassuring that my estimated false match rates are similar to those estimated by Anbinder et al. (2021).

Exact and ABE-NYSIIS conservative matches, with linkage crosswalks provided by Abramitzky et al. (2020). These are conceptually similar to the matches that I use for the main results. The other two are stricter methods,[19] one of which uses only the intersection of the set of matches by my method and the two ABE methods, and the last is a subsection of these in which the match is not characterized by a conflict across censuses in any purportedly time-invariant observables not used to make the match.[20] I refer to these as the intersection-of-matches and the intersection-of-matches-plus-corroboration methods, respectively.

In studying trends in internal migration rates over time and in making comparisons across nativities, I use an additional method laid out in detail in Appendix A to protect against spurious conclusions. In particular, I use a 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 use Bayes's Rule to estimate a rate of false matching in linkage and then to use this estimate to adjust my estimated rates of inter-county migration.[21]

The second difficulty in using linked data is that the linked samples may not be representative of the population of interest.[22] As is typical in studies using linked data, I address this issue by reweighting each linked sample to match (as closely as possible) the sample at risk for linkage on the basis of the available observables.[23] Following Collins and Zimran (2021), where possible (linkage spans ending 1900–1930), I use the latter census year of a span to determine the sample at risk for linkage, which enables me to exclude from this sample immigrants who had returned to their home countries before the final year of the span.[24]

These linked datasets provide two measures of migrant selection 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.[25] Following

[19] In principle, it would be possible to reduce the danger of false matches by only using the stricter matching methods. But these come at the cost of sample size, and more importantly, of representativeness of the linked samples.

[20] For instance, I require that the reported birthplaces of parents agree across censuses for a match to be retained.

[21] For 20-year linkage spans, I use the false match rate computed for the 10-year span beginning in the same year.

[22] Online Appendix Figure C.1 compares the observable characteristics of the linked sample to the sample at risk for linkage.

[23] That is, I estimate for each sample a series of probit regressions of the form $P(y_i = 1|\mathbf{x}_i) = \Phi(\mathbf{x}_i'\beta)$, where y_i is an indicator taking a value of 1 if individual i is successfully linked and 0 otherwise and $\mathbf{x}_i$ are all available observables in that census (see Online Appendix Table C.2). I then compute $\hat{p}_i = \Phi(\mathbf{x}_i'\hat{\beta})$ and weight linked observations according to $\frac{1}{\hat{p}}$.

[24] For instance, in the 1900–1910 linked sample, I re-weight the sample of linked immigrants to match, as closely as possible, the observables of all immigrants in 1910 who reported arriving in the United States by 1900. For linkage spans ending before 1900, year of arrival data are unavailable, and so I re-weight the sample based on the first year of the span (e.g., 1850 for the 1850–1860 span). Lower rates of return migration earlier in the Age of Mass Migration (Hatton and Williamson 1998) reduce the danger of not being able to address this issue in the earlier part of the Age of Mass Migration.

[25] Following Collins and Zimran (2021), I probabilistically re-categorize the occupational category of farmers in the 1850 census to address changes in the definition of farmers over time. I also exclude individuals in the census of 1900–1930 whose occupations are listed as "Not Yet Classified" (occ1950=979). Online Appendix F shows that my results are robust to imputing these individuals' occupations and including them.

the literature on immigrant assimilation in the Age of Mass Migration (Abramitzky, Boustan, and Eriksson 2014; Collins and Zimran 2019, 2021), I use the 3-digit occupational codes created by Ruggles et al. (2021) to create occupational scores. 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, 2021). The second is based on Preston and Haines's (1991) estimated average income by occupation in 1900 with accommodations for farmers (Collins and Wanamaker 2021; Collins and Zimran 2021). The last is the 1950-based occscore variable provided by Ruggles et al. (2021).[26] 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 occupational status measure is the simple average of these three ranks.[27] 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 G, I explore the consequences for my results of using alternative definitions of urban residence.

4 Results

Figure 1 presents my estimates of the rates of and changes over time in immigrants' secondary migration, corrected for false matches using the procedure described in Appendix A.[28] Panel (a) focuses on the estimated migration rates and panel (b) focuses on changes over time, normalizing the migration rate of the last cohort of the nineteenth century to zero. Despite the application of the correction detailed in Appendix A, the estimated secondary migration rates in panel (a) are not robust to the strictness of the linkage method.[29] But the estimated changes over time largely are,[30] so I focus my interpretation on these. The estimated 10-year migration rate declined by about 14 percentage points from the 1850–1860 cohort to the 1860–1870

[26]Following Collins and Zimran (2021), for all three methods I use the midpoint of the scores of farmers and farm laborers for men living on their family's farm who are listed as farm laborers.

[27]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.

[28]The 20-year linkage spans do not include corrected migration rates for the 1880–1900 cohort because it is not possible to estimate 10-year inter-state migration rates beginning in 1880 due to the absence of the 1890 census, and I thus cannot apply the correction of Appendix A. I present uncorrected estimates of secondary migration rates in Online Appendix Figure C.2, estimated false match rates in Appendix Figure A.1, and corrected secondary migration rates by all methods in Online Appendix Figure C.3. These figures cover all 5 linkage methods described in section 3. The main takeaway from these figures is that rates of false linkage were higher for the more permissive linkage methods than for the stricter ones, and declined over time.

[29]See Online Appendix Figure C.3 and Online Appendix E. The estimated corrected migration rates are lower according to the stricter methods.

[30]Online Appendix Figure C.4 shows the estimated changes over time in corrected secondary migration rates by all five methods.

cohort, before declining a further 4 percentage points to the 1870–1880 cohort. With an estimated secondary migration rate of about 55 percent for the 1850–1860 cohort (panel a), this is a sizable decline. From the nineteenth to the twentieth centuries, there was a slight (approximately 2 percentage point) increase followed by an additional 2 percentage points from the 1900–1910 to the 1910–1920 cohorts. The estimated changes over time for 20-year spans are similar, though the base rates are higher. The nineteenth-century decline in internal mobility appears to have been 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.[31]

Figure 2 contextualizes immigrants' secondary migration within the broader US economy, comparing the estimated inter-county migration rates of immigrants and natives. It contains three sets of estimates—differences in uncorrected migration rates,[32] differences in these rates after correcting for errors in linkage,[33] and differences after controlling for all covariates available in each census.[34] Because it is not possible to determine which individuals are falsely matched, the latter set of estimates is not corrected for false matches. But insofar as the probability of making a false match varies according to the covariates for which I control, the danger is reduced relative to the unconditional uncorrected estimates. My interpretation focuses on the corrected estimates, since these are the most conservative relative to the conclusions that I draw.

After correction, a clear premium in internal migration rates of about 22 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 at about 5–6 percentage points relative to an estimated secondary migration rate at the end of the nineteenth century of about 37–44 percent, depending on the length of the span. The premium them rose somewhat for spans beginning in 1910. These results must be interpreted cautiously, however, as the magnitude of the estimated difference is affected by the strictness of the linking method.[35] The following results are robust, however. For the nineteenth century, there is no evidence of immigrants being less mobile than natives. For the cohort first observed in 1910, all but the strictest linkage

[31] See also Online Appendix H, 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.

[32] These are the differences between the estimates of Online Appendix Figures C.2 and H.2.

[33] These are differences between the estimates of Figure 1 and Online Appendix Figure H.4.

[34] That is, I estimate an equation of the form

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

where y_{it} is an indicator equal to one if individual i made an inter-county move in span t and zero otherwise, 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. The coefficients of interest are the β_t. In addition to these fixed effects, $\mathbf{x}_{it}$ includes all the observables listed in Online Appendix Table C.2 that are observed for both immigrants and natives.

[35] See Online Appendix Figure C.5.

methods show an internal migration advantage for the foreign born, and even the strictest method shows a migration premium for natives of only 1.7 percentage points relative to an estimated 20-year secondary migration rate of 38 percent. The result is most tenuous for the cohort first observed in 1900. In this case, the strict linkage methods show a migration premium of natives of 2 to 5 percentage points relative to a base migration rate for immigrants of about 27–31 percent for 10-year spans and 35–38 percent for 20-year spans. Such a difference is certainly non-negligible, but still implies that for this single cohort whose relative mobility compared to natives is questionable, all estimates indicate that their migration rate was at least 84 percent that of natives.[36] I thus conclude that there is no evidence that immigrants were meaningfully less mobile than natives in the Age of Mass Migration, and point out that most estimates indicate the reverse.

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, household size, and an indicator for being the head of a household in the initial year, which are the variables available in all 6 initial censuses, and which are unlikely to confound the interpretation of results.[37] 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 3 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 natives alongside those for immigrants in order to contextualize the results.

Panels (a) and (b) of Figure 3 focus on the relationship between initial urban status and migration probability. The vector $\mathbf{x}_{it}$ includes initial-state fixed effects.[38] These estimates show that foreign-born urbanites were slightly more likely to move than foreign-born ruralists in the nineteenth century, and about

[36] It is important to note that this is an unconditional difference. Conditional differences after correcting for false matches cannot be computed, but it is possible that doing so would close the gap further. Moreover, consistent with results in the modern literature (e.g., Bartel and Koch 1991; Basso and Peri 2020), the immigrant migration premium is considerably larger for recent arrivals (Online Appendix Figure J.2.

[37] 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.

[38] 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.

5 percentage points less likely in the twentieth.[39] While the greater migration probability for urbanites in the nineteenth century is not robust to 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.[40] 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.[41]

Panels (c)–(f) of Figure 3 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 magnitudes of the estimates are reduced considerably by the use of stricter matching methods. The coefficient of -0.14 for the 1900–1910 span implies that a one-standard deviation increase in occupational rank (about 0.25, or 25 percentiles) was associated with a 3.4-percentage point decline in the probability of secondary migration—small but nit negligible.[42] Panels (e) and (f) focus on literacy. Again, there is clear evidence of negative selection—for the 1900–1910 span, literate immigrants are estimated to have been 7.7 percentage points less likely to move than illiterate ones. This result, however, is not robust to the strictness of the linkage method. There is, however, no evidence of *positive* selection of immigrants into secondary migration on the basis of literacy.[43] There is also clear and robust evidence of more negative selection into internal migration by the foreign born than by the native born on the basis of occupational rank.

Next, I study the destination choices of foreign-born internal migrants, restricting attention to movers. But rather than focusing solely on the characteristics of the chosen destinations, such that changes over time

[39]Since the outcome variable in this regression is an uncorrected measure of secondary migration, the proper comparison of this result is to the uncorrected secondary migration rate of 53–66 percent.

[40]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 slight urban migration premium of the nineteenth century in moving to stricter matching methods supports this.

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

[42]Again, the appropriate contextualization is from a comparison to the uncorrected secondary migration rates of 53–66 percent.

[43]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 for both occupational rank and literacy 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.

would largely reflect changes in the economy,[44] I compare the destination choices of foreign- and native-born inter-county migrants. The simplest way to do so is to compute an index of dissimilarity between them (Cutler, Glaeser, and Vigdor 2008b; Duncan and Duncan 1955). Figure 4 plots this index for each linkage span. The dissimilarity index between immigrants' and natives' destination choices hovered at about 0.45 throughout the study period, which can be interpreted as indicating that at any time about 45 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 4 also presents dissimilarity indices that are computed conditional on initial characteristics, including fixed effects for county of initial residence.[45] When comparing immigrants to observationally similar natives from the same initial county, the dissimilarity in destination choices is substantially reduced to about 0.3. Moreover, I find that the conditional dissimilarity in the destination choices of observationally similar native- and foreign-born internal migrants declined slightly over time, reaching its lowest level for the cohort first observed in 1900.

Figure 5 focuses on differences in the characteristics of destinations chosen by native- and foreign-born 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.[46] The coefficients β_t capture the difference in the means of each destination characteristic between immigrants and natives. Figure 5 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. The unconditional results are useful in exhibiting the aggregate impacts of immigrants' secondary migration on the economy.

[44] 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.

[45] Following Collins and Zimran (2021), I compute this by estimating, for each county j, an equation of the form

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

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.

[46] The controls include all variables in Online Appendix Table C.2.

Panels (a) and (b) of Figure 5 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 attracted 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.[47] 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. They then rose slightly, but never exceeded 15 percentage points.

Panels (c) and (d) of Figure 5 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,[48] where industries are defined according to Ruggles et al.'s (2021) 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.

Panels (e) and (f) of Figure 5 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).

Figure 6 measures immigrants' geographic assimilation in terms of their county-of-residence distributions. It 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

[47] Results are similar if the regression conditions only on initial county of residence and urban status.

[48] To address changing boundaries, I apply initial-year employment shares to final-year county boundaries using Hornbeck's (2010) method.

each case, there is a decline in this index from the initial to the final year, showing that some geographic assimilation did take place. But these changes were universally small, indicating that immigrants' geographic assimilation was limited: for 10-year spans, immigrants closed between 2.7 and 4.9 percent of the initial gap; for 20-year spans, immigrants closed between 4.0 and 7.9 percent of the initial gap.

Convergence was greater, however, along one specific dimension of geography. Panels (a)–(d) of Figure 7 plot 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. Panels (e) and (f) formally compare the changes over time in immigrants' and natives urbanization within each linked sample.[49] 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. The regressions of panels (e) and (f) confirm that immigrants did indeed urbanize more slowly than natives, though panels (a)–(d) show that their convergence was far from complete. In particular, over 10-year spans, immigrants closed between 2.4 (1870–1880 span) and 13.8 (1900–1910 span) percent of the gap in initial urbanization. For 20-year spans, the comparable figures are from 8.9 (1860–1880) to 20.9 (1900–1920) percent.

This notion receives further support from Figure 8, which presents similar destination choice results to Figure 5 but with the change in log population density of the residence place as the outcome.[50] This figure shows that, throughout the study period, immigrants moved to destinations with considerably lower popu-

[49] That is, I estimate a regression of the form

$$\Delta u_{it} = \beta_t f_{it} + f(a_{it}) + \varepsilon_{it},$$

where Δu_{it} is the change in individual i's urbanization (i.e., 1 if he moved from rural to urban, -1 if he moved from urban to rural, and 0 otherwise) over span t, f_{it} is an indicator taking a value of one for immigrants and zero for natives, and $f(a_{it})$ is a non-parametric function of age. Panels (e) and (f) of Figure 7 report the β_t. Note that an individual is considered to have moved from rural to urban even if he does not migrate, but his location grows and transitions from rural to urban. This motivates the separate analysis for movers only.

[50] 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.

lation 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.

5 Discussion

5.1 Immigrant Distribution and Contemporary Politics

The results described above shed new light on the claims of immigrant immobility that were the foundation for the calls for distribution efforts in the twentieth-century immigration policy debate. On the whole, these results show that these claims were either false, oversimplified, or the product of broader national trends that applied to natives as well as to immigrants.

One of the specific claims made regarding immigrants' secondary migration was that immigrants were less internally mobile than natives. My comparison of natives' and immigrants' inter-county migration rates (Figure 2), and my finding that immigrants were not meaningfully less mobile than natives, contradict this claim.[51] It was also claimed that immigrants were positively selected into secondary migration in general and specifically on the basis of literacy, leaving only the poorest and least educated immigrants in their initial places of residence. My finding of negative selection on the basis of occupational rank (Figure 3, panels c–d), and that there is no evidence of positive selection into secondary migration on the basis of literacy (Figure 3, panels e–f) contradicts this claim as well.

Other claims in the contemporary debate, though technically true, were oversimplified. One such claim was that the new immigrants of the twentieth century were less mobile than the old immigrants of the nineteenth. In the strictest sense, the decline in the internal migration rates of the foreign born over my study period (Figure 1) confirms the claims that immigrants became increasingly immobile over the Age of Mass Migration. But this 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. Thus, the view of declining immigrant mobility, while technically true, is 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,[52] were

[51] In evaluating the internal mobility of the new immigrants relative to natives, the 1910 cohort is the most relevant, as it is the first to have been composed primarily of immigrants from the new source countries. The cohort first observed in 1900 included substantial portions of immigrants from the old source countries (Collins and Zimran 2021).

[52] 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 2006; 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.

less mobile than the very first cohort that arrived before 1850. Fully explaining the greater geographic mobility of this first cohort of immigrants is beyond the scope of this paper. But it is notable that it was disproportionately composed of refugees from the Great Irish Famine and the German Revolutions (e.g., Collins and Zimran 2019). There is reason to believe that the experiences and outcomes of refugees may differ from those of other migrants (e.g., Benton-Cohen 2012; Cortes 2004; Evans and Fitzgerald 2017; Ó Gráda and O'Rourke 1997). In the case of internal migration, there is reason to believe that these refugees had less choice of initial location in the United States (Ó Gráda 1999),[53] and as a result may have been more mobile as they re-optimized their locations with time in the United States.[54]

Another true but oversimplified claim was that the new immigrants were more attached and attracted to urban areas than were natives. This claim is also, strictly speaking, true, confirmed by the more negative urban selection coefficient for immigrants than for natives (Figure 3, panels a–b) and by the greater likelihood that foreign-born secondary migrants chose urban destinations than did native-born internal migrants, even beyond the extent to which natives and immigrants were observationally different (Figure 5, panels a–b).[55] Indeed, beyond simply being the product of different observable characteristics between immigrants and natives that could be addressed by proposed screening policies, there appear to have been some quality unique to immigrants that drove this attraction.[56] But again, the technical truth of these claims belies a greater complexity. It was immigrants *throughout* the Age of Mass Migration, not only those of the twentieth century, who were both more attached and more likely to move to urban areas than were natives. Insofar as the distribution concerns raised in the twentieth century were based on perceived differences between the new immigrants and the old, these results show that they were misplaced.

Finally, the claim that immigrants grew more attached to urban areas over time, though also technically true, as evidenced by the increasing magnitude over time of the negative relationship between initial urban residence and the probability of secondary migration (Figure 3, panels a–b), was also oversimplified. In particular, rather than being the product of the changing character of immigrants, my results suggest that immigrants' increasing attachment and attraction to urban areas over the Age of Mass Migration was likely the product of the structural transformation and urbanization of the US economy over this period. This is most simply demonstrated by the fact that immigrants' increased attachment to urban areas over time was paralleled by natives, suggesting that these areas were becoming increasingly attractive over time. Put

[53]See also Bartel and Koch (1991) regarding the placement of US refugees and their subsequent secondary migration.

[54]It is also telling that this group was among the most economically mobile groups of immigrants in the Age of Mass Migration (Collins and Zimran 2021), and the two forms of mobility may have been connected.

[55]Notably, Willcox (1906) had declared such claims to be untrue.

[56]That is, the unconditional greater attraction of immigrants to urban areas is robust to controlling for all available observables.

differently, immigrants were not moving to rural areas as the advocates of distribution hoped, not because they were immobile—indeed, they had relatively high secondary migration rates (Figure 2)—but because the urbanization of the economy meant that urban areas were the location of economic opportunity, as evidence by the fact that they were also attracting native internal migrants. This interpretation is consistent with the fact that there is no evidence that immigrants were ever meaningfully less responsive to labor demand growth than natives (Figure 5, panels c–d).

The context provided by natives' internal migration also challenges the contemporary claims in that destination choices of observationally similar immigrants and natives were generally more similar in the twentieth century than at any other point in the Age of Mass Migration (Figure 4), and specifically the probability of moving to urban areas was both conditionally and unconditionally more similar across nativities in the twentieth century than it would ever be (Figure 5, panels a–b).[57] That is, 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 to them, were most similar to natives in this regard.

That immigrants were simply following the incentives of an urbanizing economy can also help to understand the failure and eventual abandonment of distribution: 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] It is also perhaps why the hopes of advocates of distribution policy that immigrants might be moved to the South and West never came to fruition (see also Figure 5, panels e–f).

5.2 The Economics of the Age of Mass Migration

Besides speaking to the contemporary political debate, my results are also informative regarding the economics of the Age of Mass Migration and add to the economics of migration more broadly. In part, this contribution comes from extending analyses of secondary migration that have been conducted for modern immigrants to the period in which US immigration was at its peak and borders were effectively open. For instance, a substantial portion of the modern literature studies the determinants of immigrants' secondary migration (Bartel 1989; Bartel and Koch 1991; Basso and Peri 2020; Beenstock 1997, 1999; Belanger and Rogers 1992; Kritz and Gurak 2001; Kritz, Gurak, and Lee 2011; Kritz and Nogle 1994; Rashid 2009). But while modern studies find positive selection into secondary migration (on the basis of education) it appears that the reverse was true in the Age of Mass Migration (Figure 3, panels c–f).

[57] The gap is even smaller in context given the rise in the base probability of choosing an urban destination.

My results also help to generate a better understanding of the economics of the Age of Mass Migration. This is because the modern literature has identified immigrants' secondary migration as an important aspect of the economics of any episode of migration. Thus, by developing the basic facts regarding the rates, selection, and sorting of secondary migration in this period, this paper illuminates a previously little understood aspect of the period.

More specifically, the results described above shed light on two specific aspects of the economics of the Age of Mass Migration. The first is immigrants' assimilation. Immigrants' limited convergence in terms of their county-of-residence distributions (Figure 6) and urbanization (Figure 7 stands in contrast to the considerable economic assimilation of immigrants in the first and last cohorts of the Age of Mass Migration (Collins and Zimran 2021), the relative similarity of immigrants' and natives' occupational status in the intervening cohorts (Abramitzky, Boustan, and Eriksson 2014; Collins and Zimran 2021), and the considerable cultural assimilation of immigrants in the Age of Mass Migration (Abramitzky, Boustan, and Eriksson 2020). For instance, Collins and Zimran (2021) find that immigrants throughout the Age of Mass Migration closed between 23 and 52 percent of the gap in initial occupational status with respect to natives,[58] and that for several cohorts, the initial gaps were very small to begin with (see also Abramitzky, Boustan, and Eriksson 2014). The geographic assimilation that I document, as well as the persistence of large gaps, is thus more in line with Eriksson and Ward's (2019) findings of persistence in immigrants' segregation from natives within cities.

These results also shed light on the effects of immigration in the Age of Mass Migration. One issue of particular concern, as discussed above, is that immigration was contributing to the growth of cities in general, and to their increasing population by low-skilled and low-educated individuals. To some extent, it was true that immigrants were more urban (Figure 7) and were more likely to be in unskilled occupations than were natives (Collins and Zimran 2021). But the effect was clearly more nuanced than this. On the one hand, the low-skilled and less-educated did not disproportionately avoid secondary migration (Figure 3, panels c–f). On the other, immigrant movers' average decline in population density and unconditionally greater decline relative to native movers (Figure 8) shows that their secondary migration, to some extent, moderated any such impact.

Moreover, in combination with insights from the modern secondary migration literature, my results suggest that, as has been shown to be the case with modern immigration (e.g., Cadena and Kovak 2016),

[58]Part of the greater assimilation documented by Collins and Zimran (2021) may come from the focus on 30-year linkage spans instead of 10- and 20-year linkage spans in this paper. But Collins and Zimran (2021) also provide results for 20-year linkage spans, and these results generate similar conclusions.

immigrants in the Age of Mass Migration may have acted through their secondary migration to equilibrate spatial differences in labor market opportunities, though I do not evaluate directly whether immigrants' secondary migration did, in fact, have this effect. First, my results are consistent with immigrants being more inherently "footloose" than natives—a property identified as potentially contributing to immigrants' response to economic opportunities elsewhere in the destination (Schündeln 2014)—as shown by the conditional comparison of internal migration rates (Figure 2). Second, my finding that, in an economy whose urban centers were rapidly growing and offering more economic opportunities, immigrants were less likely than natives to leave urban areas (Figure 3, panels a–b), more likely to move to them when they did move (Figure 5, panels a–b), and generally were at least as responsive as natives to labor demand growth (Figure 5, panels c–d) are consistent with immigrants being more responsive to geographic differences in economic opportunity. That the differences are robust to controlling for all observable differences between immigrants and natives again suggests that this responsiveness is not simply the product of different demographics, but of some other difference between immigrants and natives such that the effect is truly one of immigration and not simply of the presence of a demographically different group.

6 Summary of Robustness Checks

In addition to verifying, in Online Appendix E, the robustness of my conclusions to the choice of linkage methods, I also verify in Online Appendices F–L, the robustness of my conclusions to a variety of other potential perturbations of the data and analysis.

In Online Appendix F, 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. In Online Appendix G, 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. 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; 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.

In Online Appendix I, I redefine migration so that anyone staying in the same county or moving less than 150 miles is considered a stayer. That is, I shift the focus to longer-distance moves. The main result—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. In Online Appendix J, I also omit from spans beginning 1900–1930 immigrants who reported having been in the United States for more than 10 years. Consistent with modern results, this omission increases the internal migration rates of the foreign born in the twentieth century. This result bears particularly on the comparison of immigrants' secondary migration rates to natives' internal migration rates: the modern literature (e.g., Bartel and Koch 1991; Basso and Peri 2020) finds that recent immigrants in particular are especially mobile, and my results confirm this for the Age of Mass Migration. Next, in Online Appendix K, I 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. Any impact on results is small. In Online Appendix L, 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.[59] When imposing this limitation, the similarity in destination choices between immigrants and natives is increased relative to the baseline results and the difference in the probability of choosing a southern or western destination is also reduced.

7 Conclusion

Immigrants' secondary migration, in its role as a determinant of immigrant distribution, was a major issue in the early twentieth-century US immigration policy debate. 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 that immigrants were not mobile while in the United States, and were particularly unlikely to leave urban areas. Immigrants' secondary migration is also

[59]This is a limitation often made in other studies of immigrant assimilation (e.g., Abramitzky, Boustan, and Eriksson 2014; Collins and Zimran 2021).

an important issue in the modern economics of immigration, where it has been identified as a contributor to determining immigrant assimilation and the effects of immigration. I use linked census datasets of immigrants and natives in the United States spanning the Age of Mass Migration to develop a clearer view of immigrants' secondary migration than has ever before been possible, contributing to a clearer understanding of the Age of Mass Migration and testing the contemporary claims that helped to shape the contemporary debate. 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.

My results add to the literature on the economics of the Age of Mass Migration. They also lead me to conclude 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 the contemporaries 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. 2021a; 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.

Abramitzky, Ran, Philipp Ager, Leah Platt Boustan, Elior Cohen, and Casper Worm Hansen (2021a). "The Effects of Immigration on the Economy: Lessons from the 1920s Border Closure." *American Economic Journal: Applied Economics* Forthcoming.

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 (2021b). "Automated Linking of Historical Data." *Journal of Economic Literature* 59:3, pp. 865–918.

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, Santiago Pérez, and Myera Rashid (2020). *Census Linking Project: Version 2.0 [dataset]*. `https://censuslinkingproject.org`.

Amior, Michael (2021). "The Contribution of Immigration to Local Labor Market Adjustment." CEP Discussion Paper 1678.

Anbinder, Tyler, Dylan Connor, Cormac Ó Gráda, and Simone Wegge (2021). "The Problem of False Positives in Automated Census Linking: Evidence from New York's Irish Immigrants." UCD Centre for Economic Research Working Paper 2021/14.

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.

Barde, Robert, Susan B. Carter, and Richard Sutch (2006). "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.

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.

Basso, Gaetano, Francesco D'Amuri, and Giovanni Peri (2019). "Immigrants, Labor Market Dynamics and Adjustment to Shocks in the Euro Area." *IMF Economic Review* 67, pp. 528–572.

Basso, Gaetano and Giovanni Peri (2020). "Internal Mobility: The Greater Responsiveness of Foreign-Born to Economic Conditions." *Journal of Economic Perspectives* 34:3, pp. 77–98.

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.

Belanger, Alain and Andrei Rogers (1992). "The Internal Migration and Spatial Redistribution of the Foreign-Born Population in the United States; 1965–70 and 1975–80." *International Migration Review* 26:4, pp. 1342–1369.

Benton-Cohen, Katherine (2012). "'Strangers in the Land': A View from Western History." *Journal of the Gilded Age and Progressive Era* 11:2, pp. 263–269.

——— (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.

Borjas, George J. (1985). "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics* 3:4, pp. 463–489.

——— (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.

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.

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.

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.

——— (2021). "Up from Slavery? African American Intergenerational Mobility since 1880." *American Economic Journal: Applied Economics* Forthcoming.

Collins, William J. and Ariell Zimran (2019). "The Economic Assimilation of Irish Famine Migrants to the United States." *Explorations in Economic History* 74, 101302.

——— (2021). "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.

Cortes, Kalena E. (2004). "Are Refugees Different from Economic Immigrants? Some Empirical Evidence on the Heterogeneity of Immigrant Groups in the United States." *Review of Economics and Statistics* 86:2, pp. 465–480.

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.

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.

Evans, William N. and Daniel Fitzgerald (2017). "The Economic Success of Refugees in the US: Evidence from the ACS." NBER Working Paper 23498.

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.

——— (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.

Friedberg, Rachel M. and Jennifer Hunt (1995). "The Impact of Immigrants on Host Country Wages, Employment and Growth." *Journal of Economic Perspectives* 9:2, pp. 23–44.

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.

Gutiérrez-Portilla, María, Adolfo Maza, and María Hierro (2018). "Foreigners Versus Natives in Spain: Different Migration Patterns? Any Changes in the Aftermath of the Crisis?" *Annals of Regional Science* 61, pp. 139–159.

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.

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.

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.

Kritz, Mary M. and June Marie Nogle (1994). "Nativity Concentration and Internal Migration among the Foreign-Born." *Demography* 31:3, pp. 509–524.

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.

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.

Ó Gráda, Cormac (1999). *Black '47 and Beyond: The Great Irish Famine in History, Economy, and Memory*. Princeton: Princeton University Press.

Ó Gráda, Cormac and Kevin H. O'Rourke (1997). "Migration as Disaster Relief: Lessons from the Great Irish Famine." *European Review of Economic History* 1, pp. 3–25.

Özgüzel, Cem (2021). "The Cushioning Effect of Immigrant Mobility." CESifo Working Paper 9268.

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, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek (2021). *Integrated Public Use Microdata Series: Version 11.0* [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 (2021). "Migrant Selection and Sorting during the Great American Drought." Mimeo., USDA ERS.

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.

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 (2019). "The Low Return to English Fluency during the Age of Mass Migration." *European Review of Economic History* 24, pp. 219–242.

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.

Figures

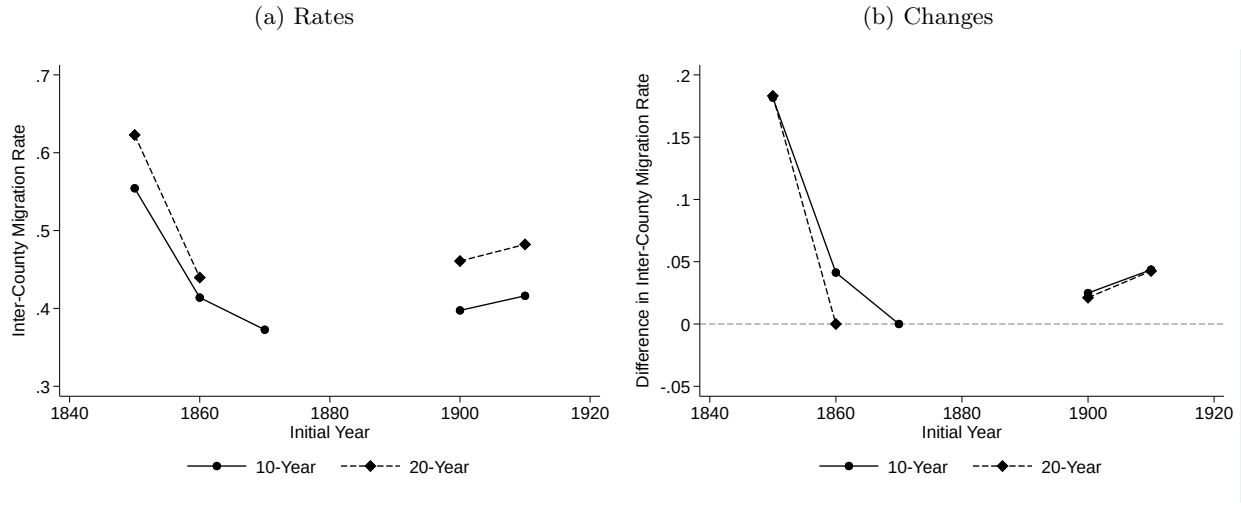


Figure 1: Inter-county migration rates and changes, corrected for false matches

Note: Panel (a) 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, corrected for false matches according to the method described in Appendix A. Panel (b) normalizes the estimates of panel (a) to zero for the last cohort of the nineteenth century. All observations are weighted by inverse linkage probability.

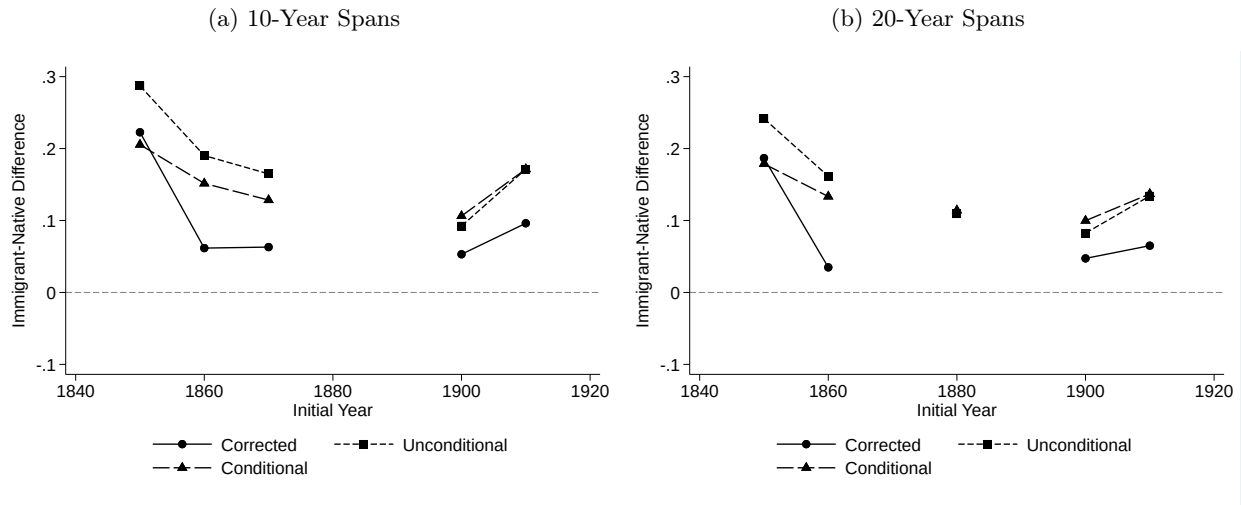


Figure 2: Differences in inter-county migration rates by nativity and span

Note: EEH-21-00058" "Replication" "-" "ExecuteThese figures present estimated differences in inter-county migration rates by nativity. Each panel presents unconditional differences (*Unconditional*), differences according to regressions including all controls available for both immigrants and natives in the census of the initial year of the span (*Conditional*), and unconditional differences after correcting for false matches as described in Appendix A (*Corrected*). All observations are weighted by inverse linkage probability.

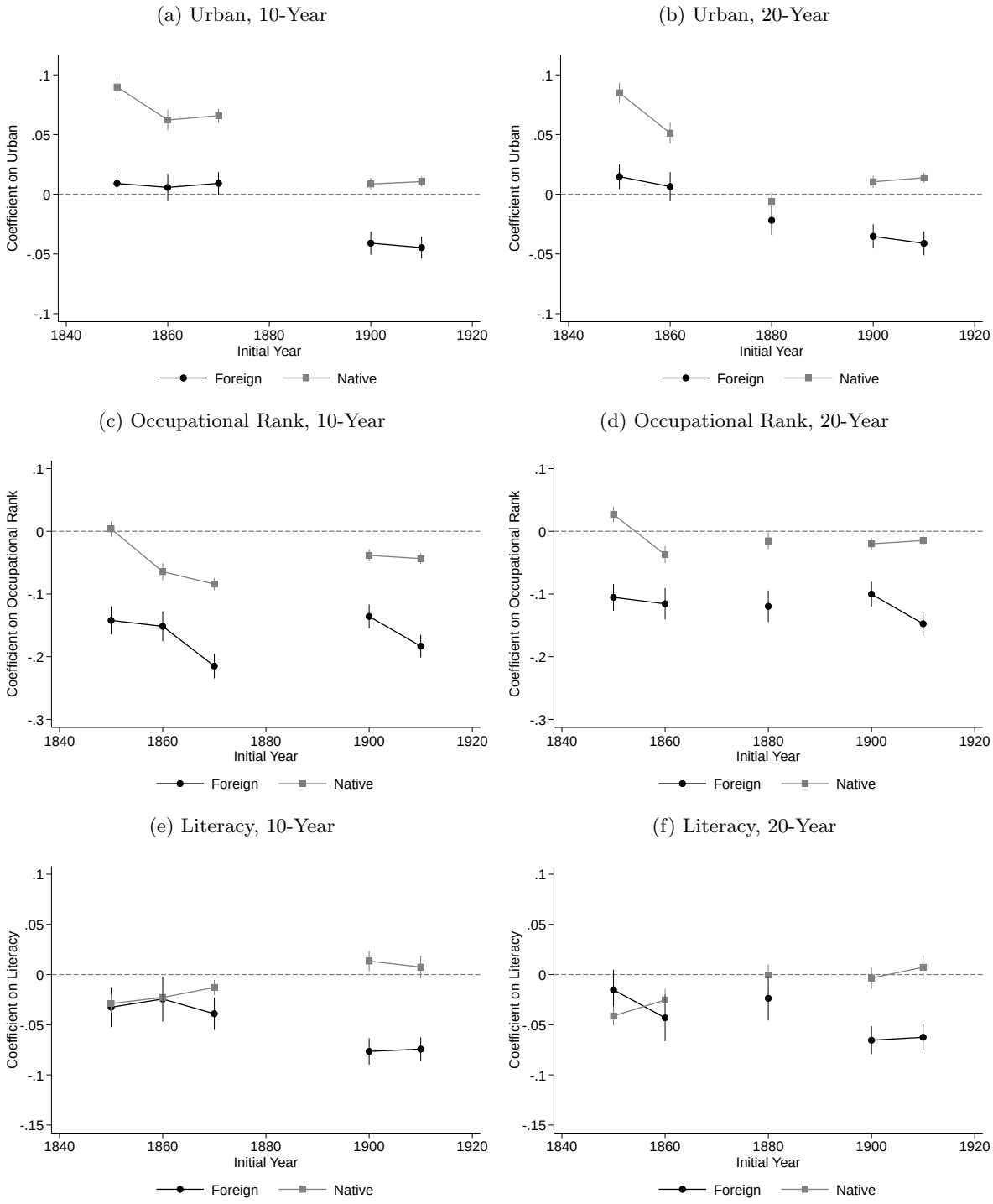


Figure 3: Selection into migration

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 linkage span. Observations weighted by inverse linkage probability. Vertical lines represent 95-percent robust confidence intervals.

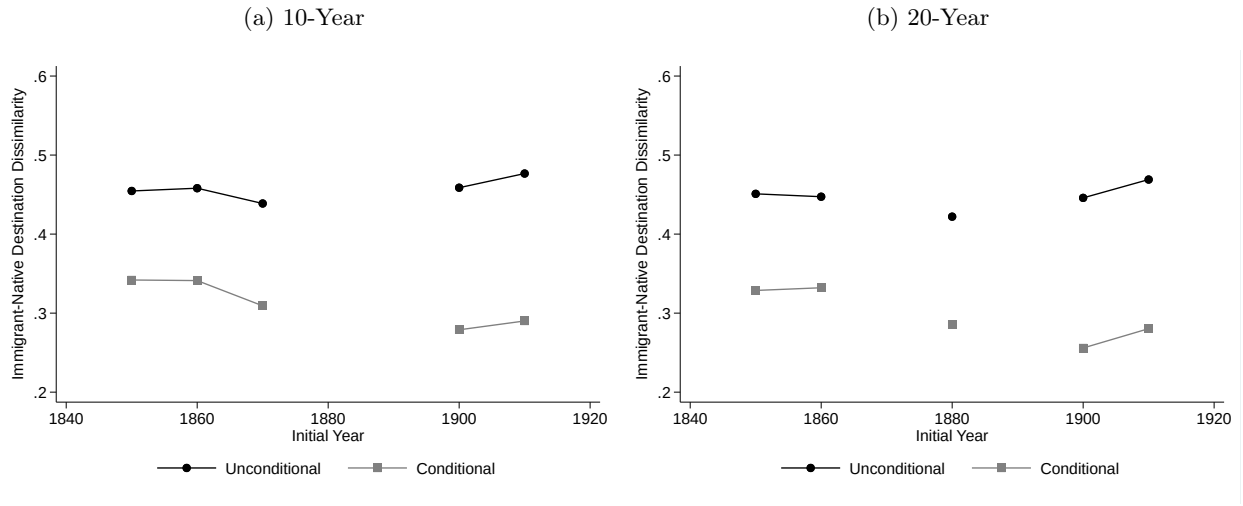
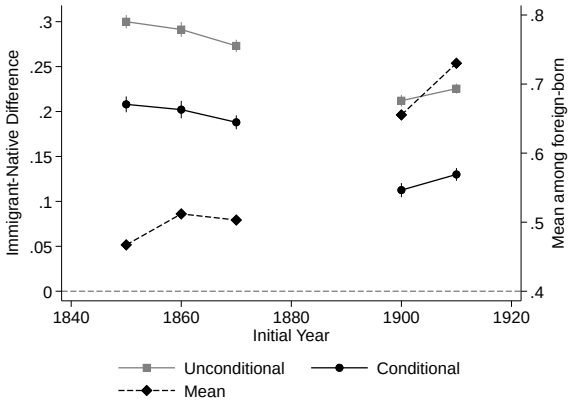


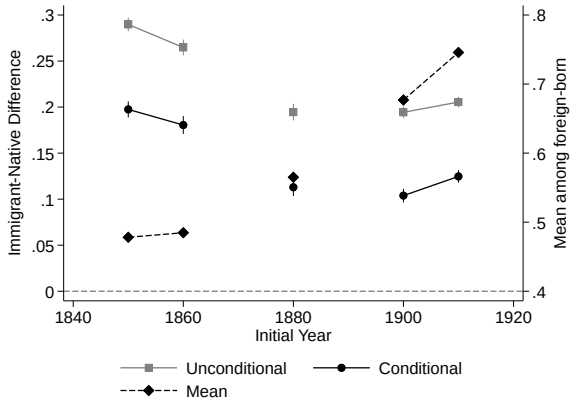
Figure 4: Dissimilarity in destination choices

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 45.

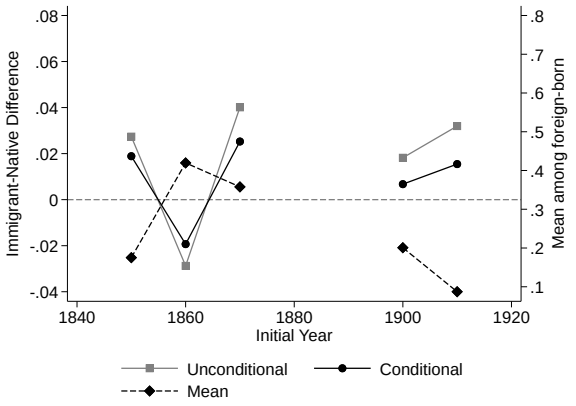
(a) Urban, 10-Year



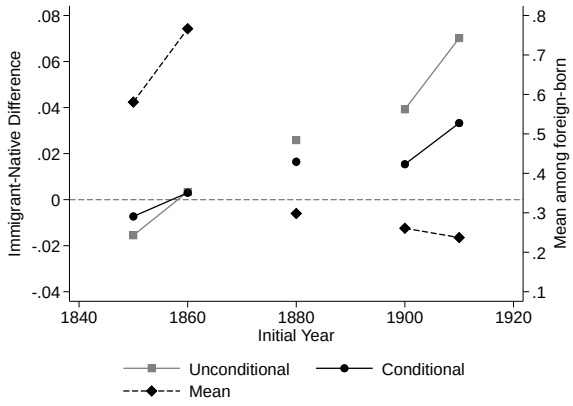
(b) Urban, 20-Year



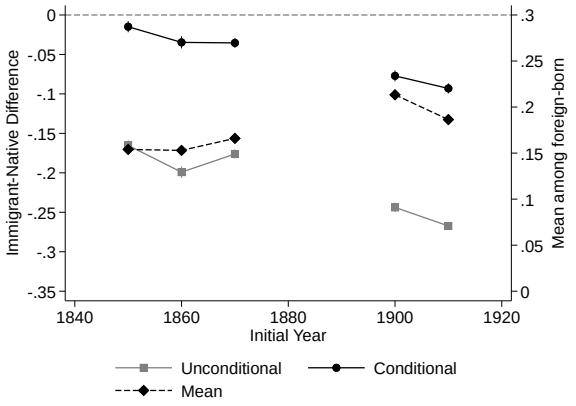
(c) Labor Demand Growth, 10-Year



(d) Labor Demand Growth, 20-Year



(e) South or West, 10-Year



(f) South or West, 20-Year

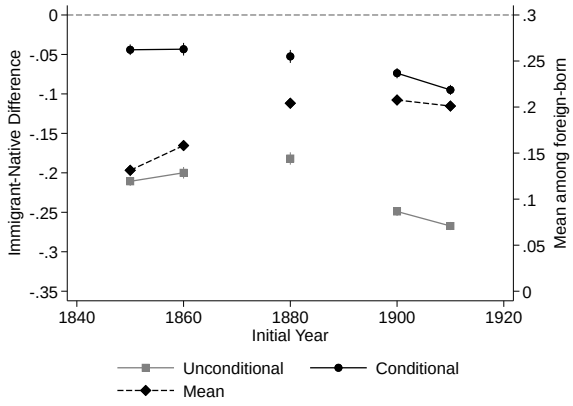


Figure 5: Immigrant-native differences in destination choice

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 destination characteristic 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 robust 95-percent confidence intervals.

(a) 10-Year

(b) 20-Year

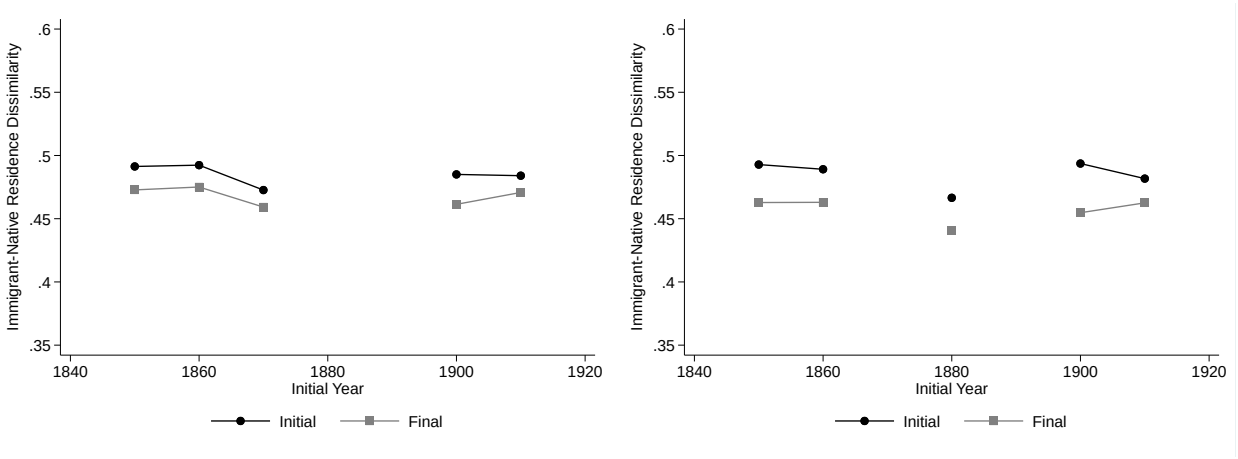


Figure 6: Dissimilarity in counties of residence

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.

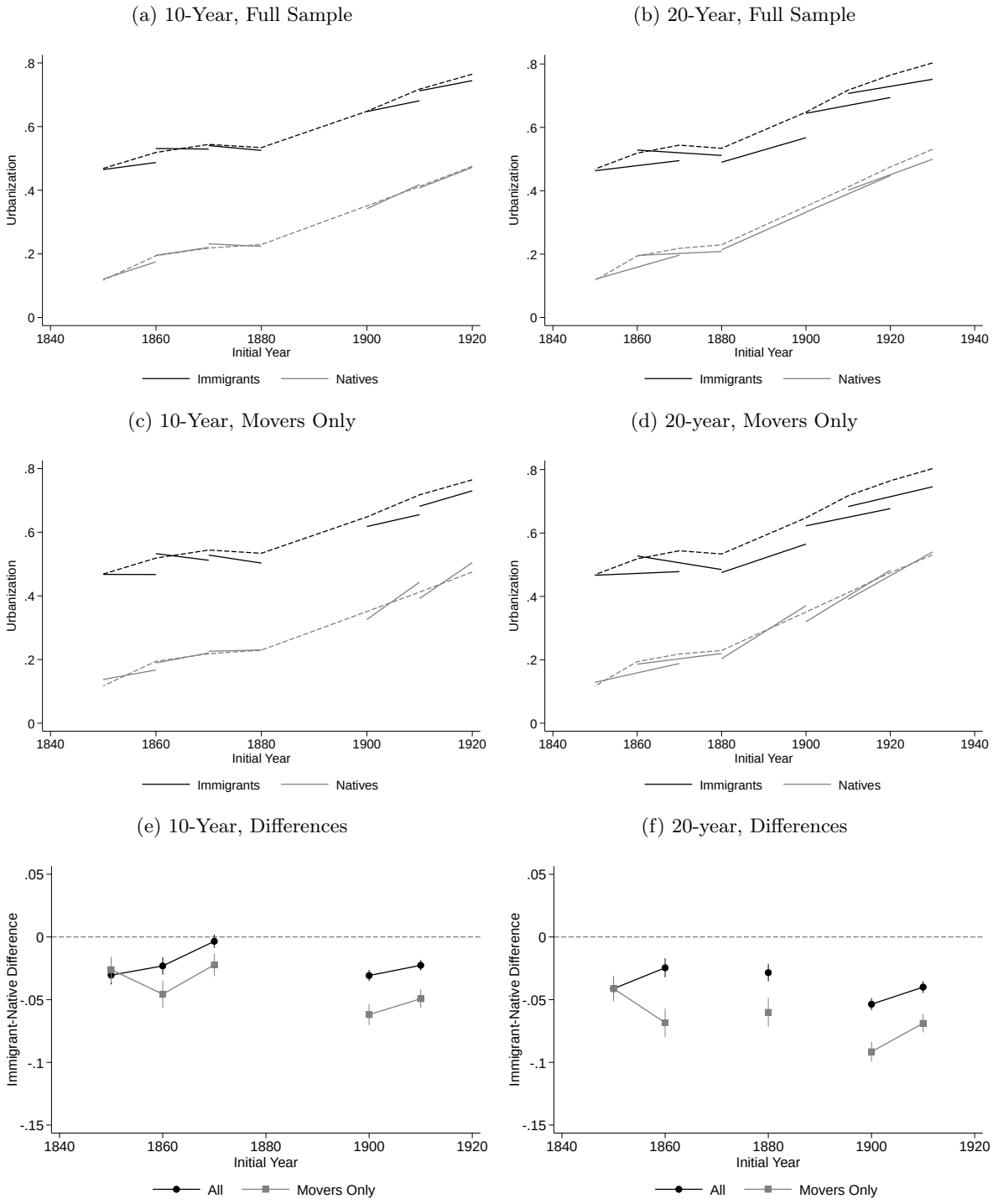
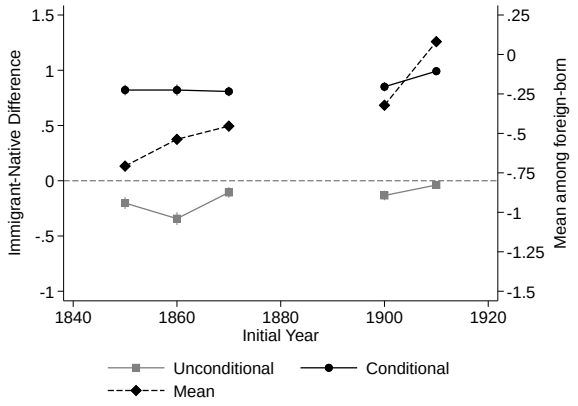


Figure 7: Urbanization of natives and immigrants

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. Each line represents a cohort, and links its initial urbanization to its final. 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. Panels (e) and (f) show regression coefficients comparing the change in immigrants' and natives' urbanization. A negative coefficient implies that immigrants' urbanization increased less than that of natives. Vertical lines are 95-percent robust confidence intervals.

(a) Change in Population Density, 10-Year



(b) Change in Population Density, 20-Year

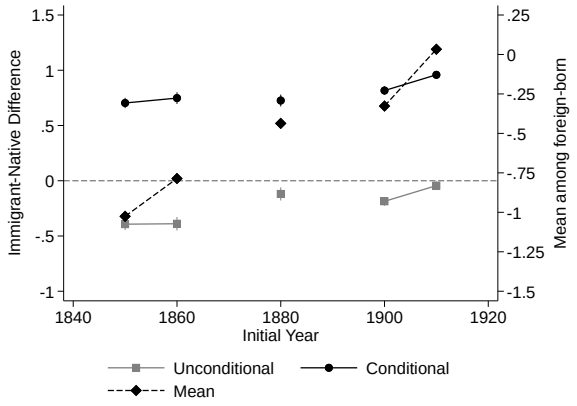


Figure 8: Immigrant-native differences in population density change

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 method used to create the linkage crosswalks underlying the main results is as follows:[60]

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 of the 1850–1910 censuses 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.[61]

      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 (but not 3) above:

   (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 1920 or 1930.

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

[61] 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).

(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

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

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.1) can be written as

$$P(f) = \frac{P(m) - P(m|t)}{P(m|f) - P(m|t)}. \tag{A.2}$$

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 both 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.2) 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. The estimated false match rates for immigrants by each method are presented in Figure A.1.

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.2) in hand, the next step is to correct the estimated migration rates. An alternative rearrangement of equation (A.1) yields

$$P(m|t) = \frac{P(m) - P(m|f)P(f)}{1 - P(f)}, \tag{A.3}$$

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.2) and Figure A.1 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.[62] 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.[63] The exception is in Online Appendix I—focusing on moves of at least 150 miles—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 equations (A.2) and (A.3), it is possible that the correction is incomplete, but the method certainly represents

[62] 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.

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

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 H and by Zimran (2021), the estimated inter-county migration rates of natives become very similar across linkage methods (Online Appendix Figure H.4).

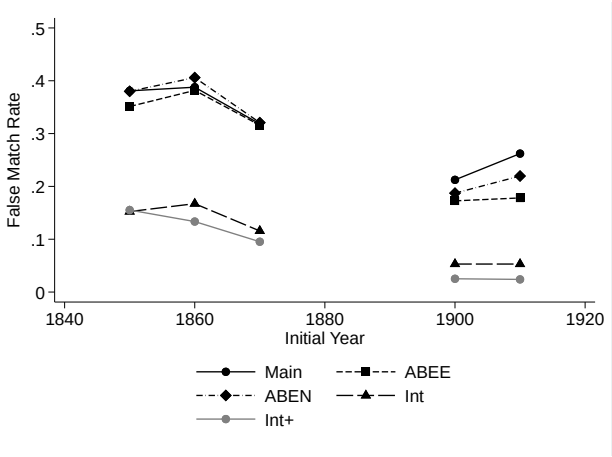


Figure A.1: Estimated rates of false linkage for immigrants

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.

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)