NBER WORKING PAPER SERIES

WORKING THEIR WAY UP? US IMMIGRANTS' CHANGING LABOR MARKET ASSIMILATION IN THE AGE OF MASS MIGRATION

William J. Collins
Ariell Zimran

Working Paper 26414
http://www.nber.org/papers/w26414

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2019

For helpful comments, the authors thank Ran Abramitzky, Leah Boustan, Joseph Ferrie, Joel Mokyr, Zach Ward, and seminar participants at Northwestern University and Vanderbilt University. Work on this paper was completed while Ariell Zimran was a W. Glenn Campbell and Rita Ricardo-Campbell National Fellow and the William C. Bark National Fellow at the Hoover Institution, Stanford University; funding from the Hoover Institution is gratefully acknowledged. Collins is the Terence E. Adderley Jr. Professor of Economics at Vanderbilt University and a Research Associate of the NBER. Zimran is an Assistant Professor of Economics at Vanderbilt University and a Faculty Research Fellow of the NBER. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Working Their Way Up? US Immigrants' Changing Labor Market Assimilation in the Age of Mass Migration

William J. Collins and Ariell Zimran

NBER Working Paper No. 26414
October 2019, Revised June 2020
JEL No. J61,J62,N11,N12,N13,N14

ABSTRACT

Whether immigrants advance in labor markets relative to natives is a fundamental question in immigration economics. It is difficult to answer this question for the Age of Mass Migration, when US immigration was at its peak. New datasets of linked census records show that immigrants experienced substantial "catching up" relative to natives' occupational status from 1850 to 1880, but not from 1900 to 1930. This change was not due to the shift in immigrant source countries. Instead, it was rooted in a sizable change in natives' occupations. The results revise the influential view that European immigrants "worked their way up".

William J. Collins
Department of Economics
Vanderbilt University
VU Station B #351819
2301 Vanderbilt Place
Nashville, TN 37235-1819
and NBER
william.collins@vanderbilt.edu

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/w26414

1. Introduction

The 30 million European immigrants who entered the United States in the century before 1921's closing of the "Golden Door" powerfully shaped American economic, demographic, and political development.[1] Their experiences form a cornerstone of broad interpretations of US history, especially regarding the opportunity for and realization of upward economic mobility. The idea that the United States is (or was) the "land of opportunity" goes hand-in-hand with the idea that the United States is "a nation of immigrants." These ideas, in turn, continue to inform contemporary views on immigration and social policy. A widely held and influential view is that European immigrants entered the US economy in low-paying occupations but experienced rapid economic and cultural assimilation despite facing discrimination and receiving little aid (Kennedy 1964; Smith et al. 2018).[2] Against this idealized historical benchmark, some question the capacity or willingness of many of today's immigrants to assimilate, and see this as a rationale for more restrictive and selective policies (e.g., Kelly 2018; Martin 2019).[3] In this paper, we test the soundness of this view of historical immigrants' experiences. We show that European immigrants *did* narrow gaps in economic status relative to US natives in the mid-nineteenth century, whereas later European immigrants did not. We also investigate *why* this pattern changed and find, surprisingly, that the dramatic shift of countries-of-origin within Europe (toward the South and East) had little to do with it. The answer instead lies in the deep transformation of the US economy between the middle and late nineteenth century.

Many scholars have studied immigrants' labor market outcomes and how they change with time spent in the United States in both historical settings and in recent decades.[4] Indeed, whether and how quickly immigrants advance in labor markets relative to natives is one of the core questions in the economics of immigration (e.g., Borjas 2014). The value of historical perspective on this

[1] This refers to the 1921 Emergency Quota Act, the first quantitative restriction on European immigration to the United States. Other dates often discussed as the end of the Age of Mass Migration include 1914 (the outbreak of World War I, which effectively stopped transatlantic migration), 1917 (the imposition of the literacy test) and 1924 (the National Origins Act, which tightened the quotas imposed in 1921).

[2] In the 1994 General Social Survey, 80 percent of respondents agreed or strongly agreed with the statement that, "The Irish, Italians, Jews, and many other minorities overcame prejudice and worked their way up. Today's immigrants should do the same without any special favors" (Smith et al. 2018).

[3] Ken Cuccinelli, the acting head of US Citizenship and Immigration Services, stated that he would welcome immigrants "who can stand on their own two feet, be self-sufficient, pull themselves up by their bootstraps—again, as in the American tradition. My Italian-Irish heritage looks back at that. Most people in America look back at that. And that's what we expect going forward" (Martin 2019).

[4] See Blau (1980), Hannon (1982), Eichengreen and Gemery (1986), Hanes (1996), Hatton (1997, 2000), Ferrie (1999), Minns (2000), and Abramitzky, Boustan, and Eriksson (2014) for historical studies. See Chiswick (1978), Borjas (1985, 2015), Card (2005), Lubotsky (2007), and Cassidy (2019) for studies of more recent decades.

question is widely recognized in this literature, but data limitations have obscured scholars' view of what actually happened between the 1840s and the 1920s, during the "Age of Mass Migration." This period, on which we focus, predates the design and implementation of national longitudinal surveys or large-scale and representative administrative records. Instead, we conduct our analyses with individual-level census records that we link over time. To construct the datasets, we start with all white men ages 18 to 40 in the "complete-count" Census of Population records for 1850, the first year in which birthplace is recorded in the census; then, we use automated methods to search for the same men in the complete-count 1880 records by using information on their birthplace, year of birth, and name.[5] We repeat this approach with men in the complete-count records of the 1900 census, searching for their matches in 1930. The panel structure of the linked data avoids biases that arise in assimilation studies that use cross-sectional data, such as changing cohort quality and selective return migration (Chiswick 1978; Borjas 1985; Lubotsky 2007; Abramitzky, Boustan, and Eriksson 2014). In addition, the datasets are sufficiently large that we can study outcomes for immigrants from specific source countries within the early and later immigrant cohorts. We show that our results are not sensitive to the linkage method we employ, which gives us confidence that we can reliably compare immigrants' and natives' outcomes over the full span of the Age of Mass Migration for the first time.

A challenge to studying historical labor market outcomes is that individual wage information is not available in nationally representative data sources. The census did not collect wage data until 1940 and even then omitted earnings for farmers and other self-employed workers. Therefore, we focus on workers' occupational status. We start by characterizing occupational status according to broad and simple categories—unskilled, operative, craft, farmer, and white collar. Then, we make full use of the censuses' detailed (three-digit) occupational information to rank workers using two different approaches. One approach relies on the complete-count records of the 1870 census to calculate the average level of wealth for men in each detailed occupational category; this "occupational score" is then assigned to each worker in our linked data according to his current occupation and transformed into a ranking.[6] A second approach relies on average income circa 1900

[5] As described below, subsequent analyses focus on non-southern whites. We focus on men because name changes at marriage make it difficult to link women and historical occupational codes for women have serious shortcomings (Goldin 1990, pp. 222-227). We focus on non-southern men because few immigrants resided in the South, because the Civil War's impact on the South would confound comparisons over this period (e.g., the application of occupational rankings that rely on post-Civil War data to the pre-Civil War South), and to maintain comparability to prior studies.

[6] In 1860 and 1870, the census collected personal and real wealth information; in 1880 and later censuses, there are no questions about wealth.

for occupations reported in Preston and Haines (1991), which in turn relies heavily on Commissioner of Labor (1903), Douglas (1930), and Lebergott (1964). We supplement that information with a new estimate of white farmers' average income using the Census of Agriculture in 1900.

As with other metrics of occupational status, which are commonly used in studies of historical economic mobility and immigrant assimilation, such rankings will capture changes in status associated with changes in occupation but not changes in economic status *within* occupation. However, later in the paper, we examine whether the main results are sensitive to using age-group-specific occupational scores, which allow for differential changes in status within occupations over the lifecycle. This addresses the concern that farmers (or others) might have accumulated wealth and improved their relative status without changing occupations, which would confound results based on a fixed occupational score. Our results and interpretations are not sensitive to this alternative.

Our first main finding is that there was a substantial change in immigrants' occupational assimilation patterns over the Age of Mass Migration. The cohort of immigrants observed in 1850 and 1880 experienced substantial occupational upgrading relative to native-born men. This pattern runs contrary to influential conclusions based on the experiences of immigrants in the early twentieth-century (Abramitzky, Boustan, and Eriksson 2014). Measured according to a ranking that averages over the two approaches mentioned above, the mid-nineteenth-century immigrant cohort reduced the gap in occupational status relative to natives from 14 to 7 percentiles. In stark contrast, when we examine a later cohort of immigrants observed in 1900 and 1930, we find little or no occupational upgrading relative to natives. This finding for the twentieth century is in line with Abramitzky, Boustan, and Eriksson's (2014) conclusions and is reassuring since we employ different methods for linking census records and assessing occupational status, and we build our linked data from larger census datasets. In sum, whereas the nineteenth-century patterns support the commonly held view that immigrants in the Age of Mass Migration started behind natives in labor markets but gained ground with time, the early twentieth-century experience was quite different.

Our second main finding is that the difference in immigrants' upgrading relative to natives in the nineteenth versus twentieth century was *not* the product of changing immigrant source countries. This finding is important in light of influential claims by social scientists, commentators, and policymakers circa 1900 that immigrants from the "new sources" were "inferior" to those who had previously come from the "old sources."[7] Such claims helped to motivate the design and implementation of immigration restrictions in the 1920s that were severely biased against potential

[7] The "old sources" were primarily Germany, Britain, and Ireland. The "new sources" were primarily Italy, the Russian Empire, and the Austro-Hungarian Empire.

immigrants from eastern and southern Europe and remained in place for most of the twentieth century (Walker 1896; Commissioner-General of Immigration 1903; Zolberg 2006; Benton-Cohen 2018). This is the first paper to use panel data to provide comparative perspective on the relationship between assimilation and country of origin over the Age of Mass Migration.

Instead of changes in country of origin, we show that the difference in immigrants' occupational upgrading between the early and later parts of the Age of Mass Migration was rooted in sizable changes in the occupational distribution of *natives* between 1850 and 1900. In 1850, natives in our sample were concentrated in farming whereas immigrants were largely engaged in unskilled labor. Men engaged in unskilled labor, whether immigrant or native, experienced high rates of occupational upgrading on average, whereas those in farming did not. By 1900, the US economy's structural transformation away from agricultural employment had collapsed differences in occupations between immigrant and native workers under age 40. That is, there were far more young natives in "unskilled" jobs in 1900 than before. We show that because natives upgraded at least as much as immigrants did *conditional* on the initial year's occupation in both the early and later cohorts, the smaller immigrant-native differences in initial-year occupations (in 1900 compared to 1850) were strongly associated with diminished upgrading for immigrants relative to natives. In this sense, the "catching up" of immigrants in the nineteenth century—the only basis that we find for the idealized historical assimilation benchmark—is largely attributable to their arrival at a time when natives were primarily employed in farming.

Our finding that immigrants in both periods (early and late in the Age of Mass Migration) failed to keep pace with natives who started in similar occupations is novel and implies an important qualification of the widely held view of historical assimilation. It contradicts the notion that European immigrants were at first under-placed in labor markets relative to natives with similar productive abilities and then gained ground rapidly with the accumulation of US-specific human capital (e.g., language skills). In fact, we show that even though immigrants in the early twentieth century did substantially improve their language skills, this did not translate into overall occupational gains relative to natives.

This paper advances the economics and economic history of immigration in several ways. It does so by first building large new datasets that provide a clear and wide window on immigrants' and natives' labor market outcomes, spanning both their individual careers and the entire Age of Mass Migration. Most crucially, the paper revises our knowledge and understanding of the facts of economic assimilation in American history. It provides the first study of immigrants' assimilation patterns over a long period with consistent methods and panel data, and the first evidence based on

such data of a change in immigrant assimilation patterns in the Age of Mass Migration. It also goes beyond other papers in this literature by providing the first explanation for *why* these patterns prevailed and changed over time.[8] At a time when an oversimplified view of historical immigrants' experiences is held up as justification for policy, these insights are timely and instructive. They challenge prevalent narratives about European immigrants' economic mobility in America—myths that continue to influence modern debates over immigration policy.

2. Immigrants in the US Labor Market during the Age of Mass Migration

Figures 1(a) and 1(b) show the volume of European immigration to the United States (relative to the population) and its composition by source country during the Age of Mass Migration, respectively.[9] During this time, there were minimal restrictions on immigration from Europe. This openness allowed for a surge in immigration during the 1840s, reflecting reduced costs of transatlantic travel, the Irish Famine, and the beginning of large-scale migration from Germany. Later in the nineteenth century, the main source countries shifted from northern and western Europe to the south and east.[10] World War I interrupted the flow of immigrants, but only temporarily. After adding a literacy test for immigrants in 1917, the United States implemented restrictive immigration quotas in 1921 and 1924, culminating decades of intense debate over immigration policy (Hutchinson 1981; Goldin 1994; Benton-Cohen 2018). In addition to imposing restrictions on annual immigration totals, the country-specific quotas were severely biased against potential migrants from southern and eastern Europe and effectively barred immigration by Asians. With minor modifications, this system governed entry to the United States until the Hart-Celler Act of 1965.

Figure 2 shows the share of white male natives and European immigrants, ages 18 to 65, working in two broad occupational categories—unskilled labor and farming—from 1850 to 1930.[11] The predominance of immigrants in unskilled work and natives in farming in 1850 is immediately apparent, as is the secular decline in farming as a share of the native labor force (and of the

[8] In advancing our understanding of assimilation in the Age of Mass Migration, this paper complements research by Hatton and Williamson (1998), Ferrie (1999), Abramitzky, Boustan, and Eriksson (2014), Abramitzky et al. (2019b), Collins and Zimran (2019), and Pérez (2019), among others. Abramitzky and Boustan (2017) and Hatton and Ward (2019) survey this literature.

[9] We focus on European migration because it accounts for nearly 90 percent of the total inflows to the US prior to 1914 (Abramitzky and Boustan 2017, Figure 2, p. 1316), because it was the subject of the main policy debates in the nineteenth and twentieth centuries, and because it was effectively unconstrained prior to World War I (as opposed to Asian immigration, which was effectively banned by the Chinese Exclusion Act of 1882 and the "Gentlemen's Agreement" of 1907).

[10] See Gould (1980), Hatton and Williamson (1998), and Spitzer and Zimran (2020) for a discussion of the timing, rapidity, and causes of this shift.

[11] We define these occupational categories as described in footnote 29 below.

immigrant labor force after 1880). From 1850 to 1880, there is convergence in the shares of each group working in farming, and by 1930, the gaps between immigrants' and natives' shares in farming and unskilled jobs were smaller than in the mid-nineteenth century. But these patterns of convergence are difficult, and possibly misleading, to interpret as evidence of labor market assimilation. Due to the continuous inflow and outflow (i.e., return migration) of immigrants, cross-sectional data sources provide little useful insight into how immigrants who stayed in the United States fared over their lifecycle. This has been a central challenge to studying immigrants' labor market outcomes during the Age of Mass Migration.

Data shortcomings, of course, did not forestall policy debates and decisions centered on immigrants' assimilation in the labor market and in other social dimensions. US natives' concerns over immigrants' poverty, lack of skills, insularity, and religion are at least as old as the onset of mass immigration. For instance, the influx of poor Irish immigrants in the 1840s and 1850s triggered an intensification of nativism and anti-Catholicism, epitomized by the rise of the "Know-Nothing" Party (Higham 1955; Anbinder 1992; Alsan, Eriksson, and Niemesh 2019). Later in the nineteenth century, the shift in source countries to southern and eastern Europe renewed vigorous debates over immigration policy (Goldin 1994; Williamson 1998; Benton-Cohen 2018). By this time, a common argument in favor of restricting European immigration involved unfavorable comparisons of the "new immigrants" to the "old immigrants," where "new" and "old" referred to place of origin. In 1896, Francis A. Walker, the first president of the American Economic Association, characterized the "new" immigrants as "beaten men from beaten races; representing the worst failures in the struggle for existence" (Walker 1896).[12] The Commissioner-General of Immigration, William Williams, bemoaned the "new" immigrants' urban settlement patterns, viewing them as evidence of the immigrants' low quality: "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 because they 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" (Commissioner-General of Immigration 1903, p. 70).[13]

In 1907, Congress authorized the era's most influential study of immigrants—the Dillingham Commission Reports. The Commission's investigation was unprecedented in scope and detail,

[12] Later, Irving Fisher, another president of the AEA, who had come to embrace eugenics and promote immigration restriction, wrote, "The core problem of immigration is one of race and eugenics…" (1921, p. 226; cited by Benton-Cohen 2018, p. 115).
[13] See Willcox (1906) for further discussion of this issue, including a discussion of the difficulties in drawing conclusions from cross-sectional data.

including the collection of data on immigrant and native workers in a variety of industries. The Commission's policy recommendations included a literacy test, source-specific quotas, and continuing bans on Asian immigration (US Congress 1911). Hourwich (1912) and Douglas (1919) published critiques of the reports and of popular summaries of the Commission's work, most notably Jenks and Lauck's (1912) book, *The Immigration Problem*.[14] They highlighted the limits of interpretation from cross-sectional data, especially in comparing "new" and "old" immigrants who had been in the United States for different lengths of time. Nonetheless, several of the Commission's main recommendations eventually became law (Benton-Cohen 2018).

Modern econometric analyses have produced mixed evidence as to whether nineteenth-century immigrants improved their status relative to natives as they gained experience in the US labor market. Blau (1980), for instance, drew on cross-sectional data tabulated and published by the Dillingham Commission and concluded that immigrants, especially from "new" sources, initially earned less than natives, but caught up over time.[15] Later studies shifted attention to cross-sectional, micro-level datasets, which include wage information for workers in specific industries and states circa 1890. Hannon (1982), for instance, concluded that immigrants in Michigan's mining industry slowly caught up to those of natives, whereas Eichengreen and Gemery (1986) and Hanes (1996) concluded that immigrants did not gain in earnings relative to natives. In contrast, after allowing more flexible functional forms in age-earnings profiles for workers in Michigan and California, Hatton (1997) concluded that immigrants who arrived as adults started at a lower wage than similar natives, gained substantially in relative pay with age, but did not fully catch up to natives. Minns (2000) shifted attention to census microdata for 1900 and 1910, which provided a more geographically comprehensive view of immigrants and natives in the labor market, albeit without wage information. He found evidence consistent with occupational gains for immigrants relative to natives, at least outside the farm sector.

More recently and most closely related to our work, Abramitzky, Boustan, and Eriksson (2014) extended methods pioneered by Ferrie (1994, 1995, 1997, 1999) to construct and study a panel dataset of immigrants and natives between 1900 and 1920.[16] They found almost no differential gains for immigrants in terms of occupational status, attributing findings of immigrant assimilation in

[14] Jenks, an economist, was an appointed member of the Commission; Lauck was a key staff member. Their book summarized their view of the Commission's findings and was more accessible to readers than the 41 volumes published by the Commission (Benton-Cohen 2008).
[15] Higgs (1971) and McGouldrick and Tannen (1977) also examined data tabulated by the Commission.
[16] Ferrie focused on immigrants arriving in New York City in the mid-nineteenth century and documented their occupational upgrading and wealth accumulation. He did not directly compare their upgrading to that of natives.

this period by prior studies to data deficiencies that are addressed by the use of panel data. Relative to their work, we have extended coverage to the mid-nineteenth century, linked the data using updated methods (but also linked it using their method to check robustness), worked from the complete-count census records in the base year (resulting in larger samples), and focused on occupational ranks derived from different approaches to measuring occupational status.[17] This enables us to provide consistent and comprehensive (all industries, all non-southern states) comparisons of immigrants' and natives' labor market outcomes over both the early decades of the Age of Mass Migration, which were dominated by immigration from "old" source countries, and the later decades, which were dominated by immigration from "new" sources, and to show that the patterns of immigrant assimilation differed dramatically between them. This is the first systematic comparison of immigrants' labor market assimilation using panel data in both the early and later eras within the Age of Mass Migration and the first analysis of the assimilation of nineteenth-century immigrants based on panel data. These data also enable us to make an important advance over such broad descriptions of assimilation patterns by explaining *why* patterns of occupational upgrading appear substantially different across the two eras within the Age of Mass Migration. The first step toward accomplishing these goals entails building new datasets from historical census records.

3. Building new datasets

Linking census records

We began by extracting all US- or European-born men from the complete-count records of the 1850 US census (Ruggles et al. 2015; Minnesota Population Center 2017). We then searched for the same men in the complete-count records of the 1880 census (Ruggles et al. 2015; Minnesota Population Center 2017), using linkage methods described below and in more detail in Appendix A. For men found at both points in time, we can observe and compare their occupations; this forms the basis for our analyses of the occupational status of immigrants relative to natives. The 1850 census was the first to record place of birth (thus identifying immigrants) and was implemented several years after the onset of mass migration from Europe. The 30-year span (to 1880) allows sufficient time to observe occupational upgrading over the lifecycle. We focus our subsequent analyses on non-southern white men because few immigrants located in the South in this period and because the Civil War's impact on the South may confound comparisons that span the war (e.g., applying an occupational ranking based on post-Civil War data to the South in 1850 would be tenuous). Our

[17] We use occupational ranks instead of scores so that cardinal differences in our scores, which may be a source of inaccuracy, are de-emphasized.

second dataset covers the early twentieth century, linking men from the 1900 to the 1930 census records using the same method (Ruggles et al. 2019). The 1900 start point allows us to include a large number of "new" immigrants, and the 1930 end point observes outcomes in the first census taken after the imposition of quotas but just before the full effects of the Great Depression.

The panel structure of the linked datasets is essential to study changes in immigrants' occupational status. Abramitzky, Boustan, and Eriksson (2014) show that using either cross-section or repeated cross-section data can generate spurious evidence of labor market assimilation due to changing arrival-cohort quality and negatively selected return migration. As in their work, our panel datasets avoid potential confounds by limiting attention to a fixed sample of men who remained in the United States. This design is appropriate for studies that are primarily interested in whether immigrants who stayed in the United States experienced relative improvements in labor market status or other aspects of assimilation. It is worth bearing in mind, however, that the linked census data might not capture the lifetime experience of the "average immigrant" given that many returned to their country of origin or died before being observed for a second time (Bandiera, Rasul, and Viarengo 2013); we return to this issue below.

Our record linkage procedure follows that of Collins and Zimran (2019), which belongs to a class of methods derived from Ferrie's (1996) pioneering work.[18] In brief, we began by eliminating from consideration men who are not unique in the base-year census (1850 or 1900) on their combination of name, birthplace, and birthyear (+/- 4 years); for example, a match would not be attempted for a John Smith born in New York in 1820. Removing non-unique individuals reduces the likelihood of making false matches. We then searched for each remaining man in the later-year census (1880 or 1930), allowing for inexact matches on the basis of name and birth year. We retained only unique matches for the analyses of occupations (i.e., matches in which an 1850 record and an 1880 record have a unique match in the opposite census, and similarly for 1900 and 1930). This approach incorporates several features highlighted by Bailey et al. (2019) and Abramitzky et al. (2019a) as useful in improving match quality, such as the use of string distance measures to compare names,[19] the elimination of non-unique individuals in the base year before linkage, and the restriction to cases in which a unique match is made between the initial and final censuses.

Table 1 presents rates of successful linkage separately for immigrants and natives. For 1850-

[18] Our approach is most similar to the methods used by Ferrie and his coauthors in recent work such as Aizer et al. (2016), Beach et al. (2016), and Long and Ferrie (2018). This class also includes the linkage methods of Abramitzky, Boustan, and Eriksson (2012, 2014).

[19] Specifically, we use the SAS function SPEDIS, which is described in detail by Gershteyn (2000). This string distance measure is similar to the Jaro-Winkler distance but treats different types of edits differently.

80, linkage rates are expressed relative to the base population of non-southern white males aged 18-40 in 1850.[20] Columns (2) and (5) indicate the number and fraction of men (in parentheses) who remained after our removal of those who were not unique on linkage characteristics. Columns (3) and (6) present the number of successful links, with the fraction of successful links expressed relative to the unique sub-sample in parentheses (i.e., those we attempted to match) and expressed relative to the entire base population (columns (1) and (4)) in square brackets. The linkage rates relative to the unique sub-sample are 25.7 percent for natives and 13.7 percent for immigrants. The former is approximately comparable to the rate of linkage relative to unique men achieved by Ferrie (1996).[21] The latter is somewhat lower. The lower rate for immigrants likely reflects the large number of immigrant arrivals after 1850, which may prevent successful links even for those who were unique in 1850. The linkage rates relative to the entire base population are 11.5 percent for natives and 5.7 percent for immigrants. These are lower than those achieved by other algorithms derived from Ferrie (1996) in other contexts (e.g., Beach et al. 2016). But as shown in Appendix F, the linkage rates are similar to those achieved using alternative methods described in detail by Abramitzky et al. (2019a).[22] Subsequent rows in Table 1 describe the linkage rates for 1900 to 1930, which are somewhat higher than in the 1850-80 data.[23] Two possible reasons for this improvement are the increased diversity of birthplaces over time and, perhaps, improvement in the quality of census data.

Several concerns arise in the use of linked census data. First, non-random selection into linkage is common in such studies. For instance, individuals with higher human capital may be more likely to survive and report consistent information to census enumerators, and name uniqueness may vary by country of origin or other characteristics.[24] Figure 3 compares observable characteristics in the linked and full samples of census data. Panel 3(a) (for natives) and 3(b) (for immigrants) focus on 1850-80 and reveal some evidence of positive selection into the linked sample, such as the over-representation of individuals with greater 1850 wealth holdings. To address this in subsequent analyses, we construct inverse probability weights based on probit regressions of a dummy variable

[20] We attempt to link all males regardless of age and make the restriction to 18-40 year olds afterwards.
[21] Note that Ferrie (1996) defines uniqueness relative to the target census rather than the base census.
[22] Our slightly lower link rates as compared to those using these alternative methods likely reflect the relatively strict definition of uniqueness that we use.
[23] The second row, which describes linkage rates relative to the 1900 population, provides statistics most comparable to those for 1850-1880, which compare to the 1850 population. The rates relative to the 1930 population, however, may be more important to consider, as we discuss below.
[24] Consistent with this concern, Appendix Tables A.1, A.2, and A.3 report rates of successful linkage for a variety of European source countries and reveal considerable variation in the rates of successful linkage.

for successful linkage on flexible functions of the 1850 covariates (separately by country of birth).[25] The effect of these weights is to make the distribution of 1850 observables in the linked sample match (as closely as possible) their distribution in the sample of men in 1850 who we attempted to link. Panels 3(c) and 3(d) report similar information for the 1900-30 samples. Given the higher rates of return migration in this period (Bandiera, Rasul, and Viarengo 2013), we weight the linked 1900-30 immigrant sample to look like the 1930 population of immigrants, aged 44-74, who reported arriving in the US before 1900 (rather than to resemble the 1900 sample that we attempted to link) and similarly reweight the 1900-30 native sample to match the 1930 native population aged 44-74.[26]

Second, some of the matches in the linked samples may be false positives, perhaps leading to spuriously high estimates of the probability of changing occupations over time. From the start, we adopted relatively strict match criteria to limit false positives with this concern in mind. Nonetheless, in Appendix E, we repeat our analysis using subsets of the linked samples that are limited to exact matches only; the results are qualitatively similar. Finally, results could be sensitive to some idiosyncrasy of the linkage method. To verify that this is not the case, Appendix F repeats our results using alternate linkage methods; again, the results are qualitatively similar.

Studying occupations to learn about labor market assimilation

We focus on the occupational status of immigrants relative to natives. This is primarily because the US census did not inquire about individuals' incomes until 1940, and even then it did not cover proprietors' or farmers' incomes.[27] Occupational information, on the other hand, was collected in each census beginning in 1850, enabling our analysis of individual-level occupational changes in both the nineteenth and the twentieth centuries in a fairly consistent manner.[28] We rely on the detailed three-digit occupational coding provided by Ruggles et al. (2015, 2019) and the Minnesota

[25] For birth countries with few observations, we estimate one regression with birth-country fixed effects.
[26] Weighting the linked 1900-30 sample to match the 1900 population of immigrants would make the (linked) sample of stayers match the average characteristics of the combined groups of stayers and return migrants. This is undesirable because our focus is on stayers, who are likely different from return migrants. Because 1880's census did not ascertain year of arrival, we cannot weight the 1850-80 sample in the same way as the 1900-30 sample; but the low rates of return migration in the early period mitigate our concern. We discuss the use of 1900-based weights for the 1900-30 sample later in the paper.
[27] For this reason, the recent literature on immigrants' economic assimilation during the Age of Mass Migration, as well as of historical socioeconomic mobility more generally, have focused on occupational status rather than on some other measure of economic status, such as income.
[28] Similarly, the census reported wealth holdings for 1850-1870 and these data formed the basis of studies of immigrant assimilation such as Ferrie's (1999). But because wealth is not reported in twentieth-century censuses, we cannot use wealth-based assimilation to draw comparisons between the experiences of immigrants in the early and late Age of Mass Migration.

Population Center (2017). We begin by using the codes to define five broad classes of occupations—farmer, white collar, craft, operative, and unskilled.[29] Our preliminary analysis entails simple comparisons (controlling only for age) of the shares of immigrants and natives in each occupational category and of changes in those shares over the men's lifecycle. This approach has the advantage of not requiring any assumptions about the relative standing of occupations beyond this broad categorization.

We then introduce rank measures of occupational status. The literature has often relied on an occupational income score based on information from the 1950 census. We prefer not to use this measure because of its temporal distance from our study period, particularly in the early portions. Instead, we construct several alternative occupational scores. The first is based on the 1870 census questions about real and personal wealth (Collins and Zimran 2019; Ager, Boustan, and Eriksson 2019; Craig, Eriksson, and Niemesh 2019). We use the complete-count sample of the 1870 census (Ruggles et al. 2020) to compute average total wealth (the sum of real and personal wealth) within each detailed (three-digit) occupational cell for men aged 30 to 65.[30] This produces an occupational wealth score that is similar in spirit to the occupational income score commonly used in studies of historical labor markets, including immigrants' occupational assimilation. We use this score in two ways—directly (in logarithmic form) and to compute occupational ranks.[31] The ranks are expressed relative to all non-southern white men in the US labor force (ages 18-74 with reported occupations). Thus, for any given man, a change in rank over time depends on whether he changed occupations and on how the occupational distribution of all men in the labor force changed.[32] We prefer the rank measure over the score itself because it is less sensitive to the specific cardinal differences of the particular score assigned. As a robustness check, we also derive wealth scores based on age groups *within* detailed occupational categories in 1870; this allows changes in scores over the lifecycle even

[29] The broad occupational categories are defined using IPUMS occ1950 codes: white collar [1,100) and [200, 500); farmer [100, 200); craft [500, 600); operative [600, 700); unskilled [700, 970]. The 1900 and 1930 data include a significant fraction of individuals whose occupation codes are "Not Yet Classified" (occ1950 code 979). Our baseline analysis omits these individuals. In Appendix D, we classify these individuals by using the modal classification of the NYSIIS code of their occupational string (similar to Collins and Wanamaker 2020) and repeat the main analysis. Our results are qualitatively unaffected by the inclusion of these individuals.
[30] These wealth statistics are calculated separately for southern and non-southern men. In cases where a region-occupation cell has fewer than 100 observations, the score is based on a broad occupational category (e.g., white collar) rather than on the specific occupation. We depart from Collins and Zimran (2019), who were the first to use these scores, by basing them on the complete 1870 sample rather than the 1% sample.
[31] In the case of ties (which are common due to the invariance of the score within occupational cells), we assign each individual in a category the average rank of the group. For example, if farmers are 20 percent of the work force and range between the 50th and 70th percentile, we assign them a rank of 0.6.
[32] This property of occupational scores and rankings has led to criticism by Inwood, Minns, and Summerfield (2019) and Saavedra and Twinam (2020).

for those who do not change occupations. Specifically, we compute separate scores for individuals aged 18-40 and those aged 41-70. Since we observe individuals once when they are aged 18-40 and once when they are aged 48-70 (setting aside allowances for differences in age in the linkage), this approach allows changes in rank between the two censuses to also reflect life-cycle changes in wealth holdings.[33]

We create another occupational income score using information that Preston and Haines (1991) compiled from Commissioner of Labor (1903), Douglas (1930), and Lebergott (1964). This provides estimates of annual income by detailed occupation circa 1900.[34] Preston and Haines do not provide an estimate for farmers, so we used the 1900 Census of Agriculture to estimate average net income for farmers based on the reported value of farm production, cost of inputs, and estimates of in-kind income. Again, we use this measure directly (in logarithmic form) as well as to compute a ranking, as with the 1870-based score. Unlike the 1870 wealth-based score, data are not available to modify the wage-based score to capture changes over the lifecycle.

Finally, we also use the two ranking schemes to create an average occupational rank, which is the simple average of ranks based on the two occupational scores discussed above. This is our preferred measure for making consistent comparisons across the 1850-80 and 1900-30 periods. The main results are qualitatively insensitive to which ranking scheme is used.[35]

In our analysis, our initial focus is on immigrants' "unconditional assimilation"—that is, we ask whether the average occupational status of immigrants improved relative to that of natives over the 30 years in which we observe each cohort, adjusting only for differences in age. This answers a simple but fundamental question about the Age of Mass Migration, and it shows how immigrants' and natives' relative occupational status changed as the US economy transformed into an urban and industrial powerhouse.

---

[33] Dividing into these "young" and "old" age cells is a simple but effective way of addressing concerns over different upgrading over time within occupational cells. In principle, we could create even finer cells. For instance, we could create birthplace-specific cells. But doing so would prevent us from distinguishing between immigrants who had been in the United States for some time from recent arrivals, thus eliminating one of the gains from the use of panel data.

[34] The Preston and Haines (1991) occupation codes do not correspond directly to those in the current IPUMS files (occ1950). Therefore, we merged the "old" Preston-Haines public use microdata sample (https://usa.ipums.org/usa/samples.shtml), which includes both the Preston-Haines occupation codes and the occ1950 census codes, with the Preston-Haines income data, and then collapsed the data to calculate averages for occ1950 cells. We thank Laura Salisbury for providing helpful insight on this issue.

[35] The correlation of the two ranking schemes depends on the year in which it is calculated. In our linked samples, the correlations between the ranks of the two measures are (in chronological order) 0.49, 0.44, 0.72, and 0.56. These correlations increase when excluding farmers and become (again in chronological order) 0.74, 0.80, 0.85, and 0.79. The difficulty in finding the appropriate ranking for farmers is common to the use of occupational scores, and it is unsurprising that the precise ranking of this large group is important.

Specifically, we estimate a series of regressions of the form

$$y_{it} = \beta_t F_i + g(a_{it}) + \varepsilon_{it} \quad (1)$$

where y_{it} is an outcome for individual *i* in year *t*, F_i is an indicator for foreign birth, and $g(a_{it})$ is a quartic function in individual *i*'s age in year *t*. We first describe differences in occupational categories, then in occupational scores and ranks, and finally differences in some rough measures of human capital. Although human capital is an input into labor market outcomes, we provide the separate analysis for a direct view on differences in basic skills that workers bring to the market, including the ability to speak English, which has long been central to discussions of immigrants' assimilation (e.g., US Congress 1911, Borjas 2015).

We estimate equation (1) separately for each outcome for each census year *t*. The coefficient β_t represents the difference in occupational status between natives and immigrants in that census year, conditional on age. To measure changes between census years for each cohort, we estimate equation (1) twice by seemingly unrelated regression, once using data from the early census year and once using data from the later census year. A measure of assimilation is then given by computing $\beta_t - \beta_{t-30}$.[36] The difference will be positive in cases where immigrants increased average levels of *y* relative to natives over time, and indicates assimilation in cases where immigrants were initially behind natives.[37]

We also use estimates of equation (1) for the five occupational category measures to compute dissimilarity indices between the two groups according to

$$D_t = \sum_k \frac{1}{2} \left| \hat{\beta}_{tk} \right|, \quad (2)$$

where $\hat{\beta}_{tk}$ is the estimated coefficient from estimating equation (1) with occupational category indicator *k* as the dependent variable. This statistic is analogous to the standard dissimilarity index (Duncan and Duncan 1955), but the difference in the fraction of each nativity group in each occupational category is estimated while controlling for the different age structures of the two groups.[38]

After characterizing assimilation patterns using these measures, we compare immigrants'

[36] We use seemingly unrelated regression because the estimates of immigrant-native differences will be correlated for a linked sample across its two years (i.e., ε_{it} and ε_{it-30} will be correlated). Seemingly unrelated regression adjusts standard errors for this and enables proper hypothesis testing of $\beta_t - \beta_{t-30}$.

[37] For instance, if immigrants start off with a lower occupational rank than natives, then a positive value of $\beta_t - \beta_{t-30}$ indicates that this gap closed over thirty years, implying that immigrants were more like natives (so long as it is not large enough for immigrants to overtake natives).

[38] Estimating the five coefficients by seemingly unrelated regression enables us to compute the standard error of D_t by the delta method.

occupational upgrading to that of natives who start in occupations with similar economic status by estimating equations of the form

$$R_{it} - R_{it-30} = \delta F_i + \gamma R_{it-30} + g(a_{it-30}) + \eta_i \quad (3)$$

where $R_{it} - R_{it-30}$ is the change in worker's average rank measure between the two periods, R_{it-30} is the worker's rank in the initial year based on the detailed occupational coding and scoring described above, and all other notation is the same as in equation (1). The coefficient δ measures the average upgrading of immigrants relative to natives, conditional on the worker's initial occupational status and a quartic in age. These "conditional assimilation" results enable us to determine whether changes in assimilation (or lack thereof) documented by estimating equation (1) reflect changes in lifetime mobility or in initial occupational distributions.

Summary statistics

Table 2 presents summary statistics for broad occupational categories and rank, measures of human capital, and indicators for region of residence and urban status. The indicators for residence confirm well known patterns, such as the over-representation of immigrants in urban areas and the Northeast. Immigrants were less literate than natives and, not surprisingly, less likely to speak English in 1900 (this cannot be observed in earlier censuses).

Figure 4 depicts the broad occupational distributions for men in the linked samples for easier comparison. In 1850, immigrants and natives had quite different occupational distributions; the modal immigrant held an unskilled job and the modal native was a farmer. Immigrants were more likely than natives to hold craft or operative occupations, whereas natives were more likely to hold white-collar occupations. In 1880, the occupational distributions retained these properties, but both immigrants and natives experienced considerable movement *into* the "farmer" category over their lifecycles (often upgrading from farm laborer).

In contrast, moving to the later cohorts, in 1900 the modal immigrant and native both held unskilled occupations, though by 1930 both groups had reduced their presence in unskilled work. Immigrants were over-represented in craft and operative labor, while natives remained over-represented in white collar and farming occupations. Comparing the 1850 and 1900 occupational distributions highlights important changes in the American economy that will have implications later in our analysis, especially the decline in the share of natives who were farmers and the rise in unskilled occupations.

4. Unconditional gains in occupational status in the 19th and 20th centuries

Table 3 presents estimates of equation (1) for broad occupational categories. Columns (1) and (2) pertain to 1850 and 1880, respectively, and column (3) presents the measure of assimilation ($\beta_t - \beta_{t-30}$). Columns (4) to (6) present analogous figures for the 1900-30 cohort. The starkest evidence of convergence in occupations between immigrants and natives comes from the first two rows for the 1850-80 cohort. In 1850, the foreign born were 23.8 percentage points more likely to hold an unskilled occupation than were natives; by 1880, the gap had shrunk to only 12.8 percentage points. Meanwhile, natives were 27.8 percentage points more likely than immigrants to be farmers in 1850, but by 1880 the gap had declined to 18.7 percentage points. Gaps in white collar and craft occupations declined only slightly, and the gap slightly widened in operative jobs.

In 1900-30, natives and immigrants began with a statistically indistinguishable probability of holding an unskilled occupation (conditional on age);[39] by 1930, however, the foreign-born were 4.6 percentage points more likely to hold an unskilled occupation. Immigrants narrowed the gap in farming, but only by a small amount (from 10.7 to 8.4 percentage points), and they did not narrow the gap in white-collar work. Overall, Table 3 reinforces the visual impression from Figure 4, in that the 1850-80 period was characterized by wide differences in initial broad occupations and strong convergence, whereas the 1900-30 period was characterized by comparatively smaller initial differences and limited convergence.

In Table 4, we summarize the occupational category differences in the form of the dissimilarity index computed as in equation (2). Specifically, the table reports dissimilarity indices for each year based on the five major categories defined above and on the estimates of Table 3. The statistic 0.31 for 1850 in column (1) indicates that 31 percent of immigrants would have to change their occupational category to match the occupational distribution of natives, or vice versa. The dissimilarity index declined from 0.31 to 0.21 between 1850 and 1880, reflecting the considerable occupational assimilation visible in Figure 4 and Table 3. On the other hand, although there is a slight decline in the index between 1900 and 1930, from 0.15 to 0.13, it is much smaller than in the earlier period, supporting the conclusion from Table 3 of limited assimilation in the twentieth century. It is notable, however, that the index in 1900 is already at a much lower level than for the earlier cohort in 1850 or even in 1880—the differences between natives and immigrants had greatly compressed. Indeed, it is plausible that the rising tide of calls for immigration restriction reflected

[39] Note that the adjustment for age eliminates the 1900 difference in unskilled work that appears in Figure 4.

the more direct contact and competition between immigrants and natives in the labor market (Goldin 1994; Hatton and Williamson 1998).

It is possible that the broad occupational categories examined to this point obscure important variation within each category; moreover, such broad categories provide a limited sense of whether men are moving up the occupational ladder with time. Therefore, we incorporate more detailed occupational definitions into the analysis and rank occupations based on what is known about the average economic status of men holding them. Table 5 repeats the estimation of equation (1) for the occupational income scoring and ranking schemes described above. Column (3) again reports *changes* in immigrants' status relative to natives over time, with positive numbers indicating that immigrants advanced faster than natives. Substantial assimilation for the 1850-80 cohort is again evident, with the immigrant-native gap in rank (averaged over the two ranking methods) closing by nearly 7 percentiles; using the 1870 wealth-based scoring, the gap narrows by 8 percentiles. All methods of scoring and ranking indicate substantial labor market assimilation in the 1850-80 dataset. In contrast, for the 1900-30 cohort, there is evidence of slightly *widening* gaps in status using every metric.

Table 6 estimates equation (1) for various measures of human capital for years in which they are available. The measures include literacy (not available in the 1880 complete count records), numeracy (an indicator for reporting an age not divisible by five), and whether the individual speaks English (available only for 1900 and 1930). Data limitations make it difficult to say much about changes for the 1850-80 cohort, though age heaping becomes slightly more prevalent for immigrants compared to natives. For the 1900-30 cohort, it is notable that the English-speaking and literacy gaps both narrowed; by 1930, the foreign-born in our sample were only 1.6 percentage points less likely than natives to speak English (though, of course, we have no metric for fluency). This is particularly interesting because the acquisition of language skills is often hypothesized to be a mechanism for immigrants' improvement in labor market outcomes (Bleakley and Chin 2010; c.f., Ward 2020); in the early twentieth century, the improvement in immigrants' language skills did not translate into overall gains relative to natives in occupational status.

Robustness

In Appendix B, we rule out several concerns regarding the robustness of the results described above. One possible concern is that the continued assimilation of second-generation immigrants in the 1900-30 sample (i.e., US-born children of immigrants, who would be included as natives) might lead to an understatement of the relative gains of immigrants. We repeat our main results comparing

immigrants to natives with native fathers and find that our results are qualitatively unchanged. Another concern is that lack of relative progress for immigrants in the 1900-30 sample might be due to including immigrants who had been in the United States for a long time and who had already experienced a period of catching up before first being observed in 1900. By using the information in 1900's census on each immigrant's year of arrival, we can limit the sample to those who arrived between 1890 and 1900.[40] Again, our results are qualitatively unchanged. Next, we repeat our main results omitting from the sample men aged 31-40 in the initial census year (1850 or 1900) to address the possibility that selective mortality that differed between immigrants and natives may have an impact on our results. We find a somewhat better starting position for immigrants relative to natives for 1900-1930 and somewhat less assimilation, but the differences are small and our results are qualitatively unchanged. Finally, we use occupational scores that are based on the wealth holdings of younger (18-40) and older (41-70) men within each detailed occupational category in the 1870. For example, if the age-wealth profile were steeper among farmers than other occupational categories, this approach would allow that fact to register in the assignment of scores and ranks. The results are similar to those from our baseline analysis.[41]

5. Conditional gains in occupational status in the 19th and 20th centuries

The changing pattern of unconditional upgrading between the early and later cohorts in the Age of Mass Migration raises the question of whether the results primarily reflect changes in natives' or immigrants' occupational mobility or, alternatively, changes in their initial distribution of occupations, each of which had a different likelihood of leading to upward mobility. To better understand the empirical underpinnings of the unconditional results, we estimate equation (3), presenting results in Table 7.

Columns (1) and (2) of Table 7, which omit the initial rank variable (R_{it}) from the right-hand side, simply reiterate the findings from section 4—that immigrants achieved greater occupational upgrading than natives in the period 1850-80 but somewhat slower upgrading in the period 1900-30—but provide a benchmark to which subsequent estimates can be compared. Columns (3) and (4)

[40] The average immigrant in our 1900-30 sample had been in the United States for 13 years in 1900 (Table 2). It is not possible to do something similar with the 1850-80 sample because neither of those censuses inquired about year of arrival in the US.

[41] Although this changes only the 1870-based scores (and not in the occupational category definition), using these scores might lead to different results for other measures of status, such as occupational category, because these scores factor into the creation of weights to correct for selection into linkage. It is notable that not only the results based on the average occupational rank (which includes the non-age-varying 1900-based rank), but also the ranks based only on 1870 show very similar results to those of the main text.

add the control for initial rank. From this perspective immigrants' upgrading relative to natives was remarkably similar—*and negative*—over the two periods within the Age of Mass Migration. If anything, immigrants in the later period (-0.03) fared slightly better than in the earlier period (-0.04). In both periods immigrants upgraded less than natives who were comparably positioned in the base year. This runs contrary to the notion that immigrants were initially under-placed in the labor market relative to their underlying skills or productivity but gradually erased that disadvantage with the acquisition of US-specific experience. It also implies that immigrant-native differences in upgrading conditional on initial occupation tended to dampen overall immigrant-native convergence in the nineteenth century.

Figure 5 provides a simple visual summary of the immigrant-native differences in upgrading conditional on initial occupational status. In this figure, we group men by broad categories for their base-year job, but changes in status are calculated using the detailed occupational information. The figure plots the average change in rank for men in each (broad) occupational category, expressed relative to that for natives who held white-collar occupations in the base year; that is, zero corresponds to the average upgrading for natives who were initially in a white-collar occupation. Three patterns stand out. First, average upgrading from the "unskilled" category tends to be larger than from other categories. Second, immigrants experienced less occupational upgrading than natives in almost every case, especially in the 1850-80 cohort. Third, relative to white-collar natives, upgrading throughout the occupational distribution was greater in 1850-80 than in 1900-30.

We can also make comparisons across the early and later periods within the Age of Mass Migration. In column (5) of Table 7, we re-weight the 1850-80 sample so that its initial occupational distribution matches that of the 1900-30 sample for natives and immigrants separately, but retaining the 1850-80-specific occupational upgrading. Column (6) weights the 1900-30 sample to match the occupational distribution of the 1850-80 sample. Column (5) shows that, had 1850-80 immigrants and natives had the 1900-30 occupational distributions, the immigrant-native gap would have *widened* by 4.4 percentiles rather than narrowing by 6.9 percentiles between 1850 and 1880. Similarly, if the 1900-30 sample is weighted by the 1850 distributions of occupations (column 6), the immigrant-native gap would have narrowed by 8.2 percentiles after 1900 instead of widening by 1.8 percentiles. In this counterfactual sense, the change in occupational distributions between 1850 and 1900 strongly influenced whether convergence occurred. The changing initial occupational distributions are thus proximately responsible for the different immigrant assimilation patterns that

we observe between the two periods.[42]

The transformation of the US economy over the nineteenth century underpins these results. Most importantly, young native workers experienced a large change in their occupational distribution; the age-adjusted dissimilarity index between natives' 1850 and 1900 occupational distributions is 0.26, a difference almost as large as that between natives and immigrants in 1850 (in Table 4). Immigrants, on the other hand, had a relatively small change in occupational distributions between 1850 and 1900; the age-adjusted dissimilarity index is only 0.08. The dramatic change in the initial occupational distribution of natives was driven primarily by a shift of native labor out of farming and into unskilled labor. Whereas 41.8 percent of natives were farmers in 1850's linked sample, only 18.2 percent were farmers in 1900's linked sample (a shift of 23.6 percentage points); conversely, whereas only 15.2 percent of natives were in unskilled occupations in 1850's sample, 38.9 percent were by 1900 (an almost identical shift of 23.7 percentage points).

Taken together, the results of sections 4 and 5 indicate that prevalent popular (e.g., Smith et al. 2018, Martin 2019) and academic (Abramitzky, Boustan, and Eriksson 2014; Ferrie 1999) views of European immigrants' assimilation during the Age of Mass Migration are incomplete. In section 4, we showed that immigrants' assimilation patterns changed dramatically over the Age of Mass Migration, with immigrants in the nineteenth century experiencing substantial convergence over time in occupational status relative to natives, unlike immigrants in the early twentieth century. But while the "catching up" of immigrants in the nineteenth century was real and contrasts with the conclusions for the Age of Mass Migration based solely on twentieth-century data, it was not the product of immigrants' superior performance relative to comparably placed natives, nor was it the product of "old" source immigrants faring better than "new" source immigrants, as we show below. Rather, the nineteenth-century immigrants were disproportionately concentrated in jobs from which many white workers (whether native, immigrant, new source or old) were likely to upgrade during their lifetime. Even though immigrants experienced less upgrading than natives in the same initial occupation, immigrants' initial concentration in low-status (but high-upgrading) occupations led to greater average occupational gains. The shift of natives into these higher-upgrading occupations by 1900, combined with natives' greater upgrades conditional on initial status, led to a divergence in occupational status over time for the later cohort. Although it might be tempting to suppose that being concentrated toward the bottom of the occupational distribution left nineteenth-century

[42] The results of columns (5) and (6) also imply that if the initial occupation-specific upgrading were interchanged between periods, this would be insufficient to reverse the results of assimilation in the nineteenth century and slight divergence in the twentieth century.

immigrants with upward mobility relative to white natives as the only possibility, the findings are not the result of processes that are mechanical or inevitable. In particular, the experience of black Americans after the Civil War shows that upward mobility from the bottom of the occupational distribution could be very slow (Collins and Wanamaker 2020).

Robustness

As with the results of section 4, there is concern that these results might be affected by the continued assimilation of second-generation immigrants, by different time spent in the United States, or by differences in selective mortality. As with the results of section 4, we verify robustness to these concerns in Appendix B.

6. Alternative explanations for differences in immigrant assimilation over cohorts

In this section, we evaluate several alternative explanations for the changing patterns of occupational upgrading that we documented above and interpreted in light of changes in the US economy during its structural transformation. Specifically, we consider whether changing source countries, changing immigrant selection, changing patterns of return migration, or changing correlates of upward mobility (e.g., literacy or internal migration) might have driven the differences in upgrading between the 1850-80 cohort and the 1900-30 cohort. We find that none of these can account for the changes observed.

Changing source countries

Influential commentators and policymakers in the early twentieth century lamented the arrival of "new" immigrants and questioned their ability to adapt to life in the United States. This view framed the "immigration problem" that the literacy test and then the quota system were intended to solve, but at the time (and until now) data did not exist to study the issue carefully. With linked census data, a simple way to test whether "old" and "new" source immigrants fared differently in terms of occupational upgrading in the early twentieth century is to include an indicator variable for "old source" in the upgrading regression, as shown in column (7) of Table 7.[43] The coefficient is small, negative, and statistically insignificant. This implies that on average there was no meaningful difference between the groups in average upgrading and, therefore, that changes in immigrants' source countries cannot account for the lack of upgrading relative to natives in the 1900-30 period.

[43] Table 2 shows that "old source" immigrants made up 42 percent of all immigrants in our 1900-30 sample.

A subtler question is whether "old" immigrants fared better than "new" immigrants conditional on initial occupational rank. In column (8) of Table 7, there is some evidence of better performance for the "old" immigrants, but it is a small difference (approximately 1 percentile), and the sum of the coefficients on *immigrant* and *old source* still implies that on average "old" source immigrants fell behind natives during the 1900-30 period.

We provide more detailed information on the relative outcomes for immigrants from different source countries in Figures 6 (occupational rank) and 7 (unskilled occupations), which present estimates of equation (1) with separate β_t coefficients for selected countries. The Irish are of particular interest given the magnitude of their migration and their relatively low initial occupational status (see also Collins and Zimran 2019). Importantly, whereas the Irish in 1850-80 showed strong evidence of "catching up" to natives, the Irish in 1900-30 did not. The patterns for German and British immigrants, though less stark, are consistent: they upgraded relative to natives in 1850-80, but they fell behind slightly in 1900-30. Again, this undermines the view that the main difference between the early and later period was the source of immigrants.

While we conclude that changes in source country cannot account for the differences in average occupational upgrading over the two immigrant cohorts we analyze, there is some important heterogeneity by source country that merits closer examination in future research. Two of the largest "new" immigrant groups were from Italy and Russia (60 percent of whom reported Yiddish or Hebrew as their mother tongue). In Figures 6(b) and 7(b) (for 1900-30), the Italians have the among the lowest starting occupational status and the Russians have the highest. In Figure 6(b), however, neither group manages to keep pace with natives in terms of average upgrading (i.e., the 1930 bar declines relative to the 1900 bar),[44] with a similar result in Figure 7(b).

Given section 5's emphasis on the quantitative importance of upgrading by men who started in the unskilled category, it may seem surprising that the Irish and Italians, who were relatively concentrated in unskilled work in 1900, failed to narrow the gap in average occupational status by 1930. Instead, the gap widened by 2 percentiles for the Irish and by 5 percentiles for the Italians (Appendix Table C.1). In part, this is because natives upgraded more than the Irish and Italians from all initial job categories after 1900,[45] but it is also because the Irish and Italian relative concentration in unskilled work was small by 1900 (around 6 percentage points more than natives). The

[44] Appendix Figure C.1 shows that differences are also apparent conditional on initial occupational group, with Russians upgrading more than natives who start in the same broad occupation (and better than other immigrant groups) and Italians upgrading less (and worse than other immigrant groups).
[45] This can be seen in Appendix Figure C.1.

importance of the changes in occupational distributions between 1850 and 1900 can be seen in a counterfactual that is specific to Irish and Italian immigrants (Appendix Table C.1): if the 1900-30 Irish and Italian immigrants are re-weighted to match the 1850 occupational distribution of all immigrants, while the 1900-30 natives are re-weighted to match the 1850 distribution of natives, the Irish would have closed the gap relative to natives by 5.6 percentiles and the Italians by 2.2 percentiles.

Changing return migration

Because we focus on individuals who could be linked between US censuses, return migrants are not included in our analysis. Return migration likely increased as transportation costs fell,[46] and therefore selective return migration might complicate comparisons of the earlier and later samples (Greenwood and Ward 2015; Ward 2017; Abramitzky, Boustan, and Eriksson 2019). For instance, it may be the case that the share of immigrant *stayers* (in the linked sample) who were in unskilled jobs in 1900 was much lower than the share of all immigrants in such jobs in 1900. Because upgrading from unskilled jobs tended to be greater than from other categories (for stayers and natives), this might cause us to miss some of the immigrants' upgrading because they left the United States before we observed them a second time.

To test whether differential rates of return migration between the two periods might be responsible for our results, we reweighted the 1900-30 sample based on the 1900 immigrant population (rather than the 1930 immigrant population, as described in section 3). In practice, this tends to "upweight" immigrant men in the linked sample who started in unskilled occupations relative to our baseline analyses. This implies (by construction) a greater initial gap in occupational status between natives and immigrants in 1900—the coefficient of estimating equation (1) for the average occupational rank using 1900 weights is -0.052 (Appendix Table C.2), as opposed to -0.015 when using the 1930 weights (Table 5). Of course, we cannot see how much the return migrants upgraded while they were in the United States, let alone how much they would have upgraded if they had stayed until 1930. But if we assume that they would have fared the same as (or not much better than) immigrant stayers who held similar occupations in 1900, then we can use the reweighted sample to estimate counterfactual upgrading. From this perspective, by all metrics, immigrants' upgrading relative to natives from 1900 to 1930 would have been very small even when positive—a narrowing by 0.7 percentiles (Appendix Table C.2) as opposed to our benchmark estimate of a

[46] There is evidence of high rates of return migration in the early twentieth century (Bandiera, Rasul, and Viarengo 2013). There is less evidence on return migration in the nineteenth century, but it was likely lower.

widening by 2.0 percentiles (Table 5) and the 1850-80 narrowing of 6.9 percentiles.

Changing migrant selection

Another potential explanation for the difference in upgrading between 1850-80 and 1900-30 is that there was deterioration in migrant selection over time, independent of the source country hypothesis discussed above. In other words, the greater upward mobility of immigrants in the nineteenth century may have reflected superior unobservable quality relative to the later period. Consistent with this hypothesis, Collins and Zimran (2019) provide suggestive evidence of positive selection among Irish migrants prior to 1850, while Connor (2019) documents neutral selection in the later nineteenth century. In general, a Roy-Borjas model would predict that in countries that initially sent positively selected migrants, declining migration costs or growing migrant networks would tend to select less skilled workers into migration at the margin.[47]

It is difficult to measure changes in migrant selection from source countries' historical census data (Spitzer and Zimran 2018). Nonetheless, there is strong evidence against the hypothesis that changing migrant selection drives our results. More negative selection of immigrants in the twentieth century could affect their observed upgrading in two potential ways: (1) the initial occupational distribution could be more unskilled, and (2) upgrading from specific occupations could be weaker. First, if more negative selection in the twentieth century worsened the initial occupational status of immigrants relative to natives (compared to what it would have been with constant migrant selection), this would tend to drive immigrants into occupations from which workers experienced *greater* upgrading. That is, we would expect more, not less, upgrading in 1900-30 if this mechanism were important. Second, we can assess a counterfactual to rule out the possibility that changes in occupational upgrading conditional on initial status are quantitatively important to the results. As shown in column (5) of Table 7, weighting the nineteenth-century conditional upgrading performance by the 1900 occupational distributions effectively gives the 1900-30 cohort the same conditional upgrading performance as the earlier and perhaps more positively selected cohort. This counterfactual estimate still predicts a divergence in occupational status between immigrants and natives in the 1900-30 period.

[47] In principle, declining transport costs might have led to such a change, but there is little evidence that such a change took place after the middle of the nineteenth century (e.g., Keeling 2012). Another reason for such a change might be expanding networks of prior migrants (e.g., McKenzie and Rapoport 2010).

Changing correlates of upward mobility

Finally, it is possible that personal or location-specific characteristics that were associated with upward mobility changed between the earlier and later cohorts. Columns (1) to (4) of Table 8 estimate augmented versions of equation (3), including individual-level variables for 1850 literacy, urban residence, and county-of-residence fixed effects. These initial conditions are not random, and so we make no causal claims regarding their relationship with occupational changes. Instead, we are interested in whether including the additional controls diminishes the difference in immigrants' average upgrading between 1850-80 and 1900-30 cohorts.

Columns (1) and (2) measure immigrants' gains relative to natives without controlling for initial occupational status but with the additional covariates mentioned above. As before, the 1850-80 period is characterized by a sizable amount of upgrading for immigrants (nearly 8 percentiles), whereas 1900-30 is characterized by a smaller (but now positive) amount of immigrant upgrading (just under 2 percentiles). Also as before, columns (3) and (4), which control for initial occupational rank, indicate that immigrants upgraded less than natives who started in similarly ranked occupations. In sum, the upgrading patterns described in sections 4 and 5 are fairly robust to the controls for base-year observables: the 1850-80 cohort still has significantly stronger upgrading than the 1900-30 cohort, and immigrants in both cohorts fail to keep pace with similarly situated natives.

In columns (5) to (8), we add covariates for whether an individual moved to a new county between the initial and final year, for whether such a move was to an urban area, and indicators for the county of residence in the final year. These are "bad controls" in that they may go hand-in-hand with occupational changes, but their inclusion may reveal whether changing patterns of geographic mobility underpin the results of sections 4 and 5 (e.g., if immigrants' gains in the 1850-80 period were strongly tied to internal migration, the coefficient on *immigrant* in column 5 may diminish).[48] In practice, however, the main findings regarding immigrant assimilation are essentially unchanged in columns (5) to (8); the coefficients on *immigrant* are very similar to those in columns (1) to (4).

7. Conclusions

The availability of complete-count census datasets and the ability to link records across those datasets have enabled scholars to revisit fundamental questions about the first era of mass immigration to the United States. One of those questions pertains to immigrants' labor market outcomes, specifically whether immigrants improved their economic status relative to natives as they

[48] Including indicators for destination county is done in the intergenerational mobility literature to determine whether characteristics of the destination are responsible for findings regarding upgrading (Tan 2019).

gained experience in the US labor market. Although a clearer picture of migrant assimilation in the twentieth century has recently emerged as a result of these data advances, the experience of earlier migrant cohorts is comparatively uncharted territory. In this paper, we provide a more complete view of the Age of Mass Migration by encompassing both the era's early and later decades and providing consistent comparisons of immigrants who arrived at different times and from different sources. We do this by creating new datasets of linked census records for men between the censuses of 1850 and 1880 and between the censuses of 1900 and 1930.

From a contemporary perspective, at a time when immigration policy is again at the forefront of political discourse, this paper's investigation is valuable because many Americans have strongly held views about the experiences of European immigrants in the Age of Mass Migration and what they imply about appropriate policy today (Smith et al. 2018; Martin 2019). Anecdotes, family lore, and wishful thinking have tended to predominate over systematic evidence in forming and sustaining those views of American history. We offer here new data and analyses that illuminate the Age of Mass Migration and challenge the notion that European immigrants simply "worked their way up" the economic ladder in America.

Our analysis arrives at two main conclusions, each with important implications for understanding the long history of immigration in the United States. The first is that European immigrants substantially upgraded their occupational status relative to natives in the nineteenth century, but not in the early twentieth century. The second main conclusion is that the difference in immigrants' upgrading between the early and later decades of the Age of Mass Migration was not rooted in the changing composition of immigrant source countries. Rather, we show that the answer lies in the collapsing differences in occupational distributions between immigrants and natives over time, which in turn reflects the structural transformation of the US economy. In 1900, many more young native men held unskilled occupations than in 1850. Upgrading from unskilled occupations was more common than from other occupations for both immigrants and natives; moreover, natives' had a consistent advantage in upgrading conditional on initial occupation. Thus, as young immigrants' and natives' initial occupational distributions became more similar between 1850 and 1900, immigrants' likelihood of advancing relative to natives eroded. In this sense, the mid-nineteenth-century immigrants' "catching up" to natives was due primarily to the early timing of their arrival, not to productive advantages associated with their home countries or personal characteristics.

References

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 Katherine Eriksson (2012). "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102:5, pp. 1832-1856.

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.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson (2019). "To the New World and Back Again: Return Migrants in the Age of Mass Migration." *Industrial and Labor Relations Review* 72:2, pp. 300-322.

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

Abramitzky, Ran, Leah Platt Boustan, Elisa Jácome, and Santiago Pérez (2019b). "Intergenerational Mobility of Immigrants in the US over Two Centuries." NBER Working Paper 26408.

Ager, Philipp, Leah Platt Boustan, and Katherine Eriksson (2019). "The Intergenerational Effects of a Large Wealth Shock: White Southerners After the Civil War." NBER Working Paper 25700.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney (2016). "The Long-Run Impacts of Cash Transfers to Poor Families." *American Economic Review* 106:4, pp. 935-971.

Alsan, Marcella, Katherine Eriksson, and Gregory Niemesh (2019). "The Rise and Fall of the Know-Nothing Party." Mimeo., Miami University.

Anbinder, Tyler (1992). *Nativism and Slavery: The Northern Know Nothings and the Politics of the 1850s.* New York: Oxford University Press.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey (2019). "How Well Do Automated Methods Perform in Historical Samples? Evidence from New Ground Truth." *Journal of Economic Literature* Forthcoming.

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

Barde, Robert, Susan B. Carter, and Richard Sutch (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 Gardner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (eds.). Cambridge: Cambridge University Press, pp. 1.560-1.563.

Beach, Brian, Joseph P. Ferrie, Martin Saavedra, and Werner Troesken (2016). "Typhoid Fever, Water Quality, and Human Capital Formation." *Journal of Economic History* 76:1, pp. 41-75.

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

Blau, Francine (1980). "Immigration and Labor Earnings in Early Twentieth Century America." *Research in Population Economics* 2, pp. 21-41.

Bleakley, Hoyt and Aimee Chin (2010). "Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants." *American Economic Journal: Applied Economics* 2:1, pp. 165-

192.

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

Borjas, George J. (2014). *Immigration Economics*. Cambridge: Harvard University Press.

Borjas, George J. (2015). “The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again.” *Journal of Human Capital* 9:4, pp. 483-517.

Card, David (2005). “Is the New Immigration Really So Bad?” *Economic Journal* 115, pp. F300-F323.

Cassidy, Hugh (2019). “Occupational Attainment of Natives and Immigrants: A Cross-Cohort Analysis.” *Journal of Human Capital*, 13:3, pp. 375-409.

Chiswick, Barry R. (1978). “The Effect of Americanization on the Earnings of Foreign-born Men.” *Journal of Political Economy* 86:5, pp. 897-921.

Collins, William J. and Marianne H. Wanamaker (2020). “African American Intergenerational Mobility since 1880.” NBER Working Paper 23395.

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

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.

Commissioner of Labor (1903). *Eighteenth Annual Report of the Commissioner of Labor.* Washington: Government Printing Office.

Connor, Dylan S. (2019). “The Cream of the Crop? Geography, Networks, and Irish Migrant Selection in the Age of Mass Migration.” *Journal of Economic History* 79:1, pp. 139-175.

Craig, Jacqueline, Katherine Eriksson, and Gregory T. Niemesh (2019). “Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1950-1910.” Mimeo., Miami University.

Douglas, Paul H. (1919). “Is the New Immigration More Unskilled than the Old?” *Publications of the American Statistical Association* 16, pp. 393-403.

Douglas, Paul H. (1930). *Real Wages in the United States, 1890-1926.* Boston: Houghton Mifflin.

Duncan, Otis Dudley and Beverly Duncan (1955). “A Methodological Analysis of Segregation Indexes.” *American Sociological Review* 20:2, pp. 210-217.

Eichengreen, Barry and Henry A. Gemery (1986). “The Earnings of Skilled and Unskilled Immigrants at the End of the Nineteenth Century.” *Journal of Economic History* 46:2, pp. 441-454.

Ferrie, Joseph P. (1994). “The Wealth Accumulation of Antebellum European Immigrants to the US, 1840–60.” *Journal of Economic History* 54:1, pp. 1–33.

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.

Ferrie, Joseph P. (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.

Ferrie, Joseph P. (1997). "The Entry into the US Labor Market of Antebellum European Immigrants, 1840–1860." *Explorations in Economic History* 34, pp. 295–330.

Ferrie, Joseph P. (1999). *Yankeys Now: Immigrants in the Antebellum United States, 1840–1860*. New York: Oxford University Press.

Fisher, Irving (1921). "Impending Problems of Eugenics." *Scientific Monthly* XIII (September): 214-231.

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 (1990). *Understanding the Gender Gap: An Economic History of American Women*. New York: Oxford University Press.

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 (eds.). 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.

Greenwood, Michael J. and Zach Ward (2015). "Immigration Quotas, World War I, and Emigrant Flows from the United States in the Early 20th Century." *Explorations in Economic History* 55, pp. 76-96.

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 Gardner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright (eds.). Cambridge: Cambridge University Press, pp. 1.28-1.30.

Hanes, Christopher (1996). "Immigrants' Relative Rate of Wage Growth in the Late 19th Century." *Explorations in Economic History* 33, pp. 35-64.

Hannon, Joan U. (1982). "Ethnic Discrimination in a Nineteenth Century Mining District: Michigan Copper Mines 1888." *Explorations in Economic History* 19, pp. 28-50.

Hatton, Timothy J. (1997). "The Immigrant Assimilation Puzzle in Late Nineteenth-Century America." *Journal of Economic History* 57:1, pp. 34-62.

Hatton, Timothy J. (2000). "How Much Did Immigrant 'Quality' Decline in late Nineteenth Century America?" *Journal of Population Economics* 13, pp. 509-525.

Hatton, Timothy J. and Zachary Ward (2019). "International Migration in the Atlantic Economy 1880-1940." In *Handbook of Cliometrics*. Claude Diebolt and Michael Haupert (eds.). Springer.

Hatton, Timothy J. and Jeffrey G. Williamson (1998). *The Age of Mass Migration: Causes and Economic Impact*. New York: Oxford University Press.

Higgs, Robert (1971). "Race, Skills, and Earnings: American Immigrants in 1909." *Journal of Economic History* 31, pp. 420-428.

Higham, John (1955). *Strangers in the Land: Patterns of American Nativism, 1860-1925*. New Brunswick: Rutgers University Press.

Hourwich, Isaac (1912). *Immigration and Labor: The Economic Aspects of European Immigration to the United States*. New York: G.P. Putnam and Sons.

Hutchinson, E.P. (1981). *Legislative History of American Immigration Policy 1798-1965*. Philadelphia: University of Pennsylvania Press.

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.

Jenks, Jeremiah W. and W. Jett Lauck (1912). *The Immigration Problem*. New York: Funk & Wagnalls.

Keeling, Drew (2012). *The Business of Transatlantic Migration Between Europe and the United States, 1900-1914.* Zurich: Chronos Verlag.

Kelly, John (2018). "Transcript: White House Chief of Staff John Kelly's Interview with NPR" [Interview]. NPR.org. URL: https://www.npr.org/2018/05/11/610116389/transcript-white-house-chief-of-staff-john-kellys-interview-with-npr. Accessed September 20, 2019.

Kennedy, John F. (1964). *A Nation of Immigrants: Revised and Enlarged Version*. New York: Harper & Row.

Lebergott, Stanley (1964). *Manpower in Economic Growth: The American Record since 1800*. New York: McGraw-Hill.

Long, Jason and Joseph P. Ferrie (2018). "Grandfathers Matter(ed): Occupational Mobility across Three Generations in the US and Britain, 1850-1911." *Economic Journal* 128, pp. F422-F445.

Lubotsky, Darren (2007). "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings." *Journal of Political Economy* 115:5, pp. 820-867.

Martin, Rachel (2019). "Rule Would Penalize Immigrants to US for Needing Benefits" [Interview]. NPR.org. URL: https://www.npr.org/2019/08/13/750727515/rule-would-penalize-immigrants-to-u-s-for-needing-benefits. Accessed September 20, 2019.

McGouldrick, Paul F. and Michael B. Tannen. 1977. "Did American Manufacturers Discriminate Against Immigrants before 1914?" *Journal of Economic History* 37, pp. 723-746.

McKenzie, David and Hillel Rapoport (2010). "Self-Selection Patterns in Mexico-US Migration: The Role of Migration Networks." *Review of Economics and Statistics* 92:4, pp. 811-821.

Minnesota Population Center (2017). *North Atlantic Population Project: Complete Count Microdata: Version 2.3* [Machine-readable dataset]. Minneapolis: University of Minnesota.

Minns, Chris (2000). "Income, Cohort Effects, and Occupational Mobility: A New Look at Immigration to the United States at the Turn of the 20th Century." *Explorations in Economic History* 37, pp. 326-350.

Pérez, Santiago (2019). "Southern (American) Hospitality: Italians in Argentina and the US during the Age of Mass Migration." NBER Working Paper 26127.

Preston, Samuel H. and Michael R. Haines (1991). *Fatal Years: Child Mortality in Late Nineteenth Century America*. Princeton: Princeton University Press.

Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek (2015). *Integrated Public Use Microdata Series: Version 6.0* [Machine-readable database]. Minneapolis: University of Minnesota.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek (2019). *IPUMS USA: Version 9.0* [Machine-readable database]. Minneapolis: University of

Minnesota.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek (2020). *IPUMS USA: Version 10.0* [Machine-readable database]. Minneapolis: University of Minnesota.

Saavedra, Martin H. and Tate Twinam (2020). "A Machine Learning Approach to Improving Occupational Income Scores." *Explorations in Economic History* 75, 101304.

Smith, Tom W., Michael Davern, Jeremy Freese, and Stephen Morgan (2018). *General Social Surveys, 1972-2018* [Machine-readable data file]. Chicago: NORC.

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.

Spitzer, Yannay and Ariell Zimran (2020). "Like an Ink Blot on Paper: Testing the Diffusion Hypothesis of Mass Migration, Italy 1876-1920." Mimeo., Vanderbilt University.

Tan, Huiren (2019). "A Different Land of Opportunity: The Geography of Intergenerational Mobility in the Early 20th-Century US." Mimeo., National University of Singapore.

US Congress (1911). *Reports of the Immigration Commission, vol. 1.* Washington: Government Printing Office, 61st Congress, 3rd Session, Document No. 747.

Walker, Francis A. (1896). "Restriction of Immigration." *Atlantic* June 1896.

Ward, Zachary (2017). "Birds of Passage: Return Migration, Self-Selection, and Immigration Quotas." *Explorations in Economic History* 64, pp. 37-52.

Ward, Zachary (2020). "The Low Return to English Fluency during the Age of Mass Migration." *European Review of Economic History* 24:2, pp. 219-242.

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

Williamson, Jeffrey G. (1998). "Labor Markets and Policy Backlash in the Past." *Journal of Economic Perspectives* 12, pp. 51-72.

Zolberg, Aristide R. (2006). *A Nation by Design: Immigration Policy in the Fashioning of America*. Cambridge: Harvard University Press.

Figures

Figure 1: Descriptive statistics on US immigration history

Figure 1(a): European immigration as a share of population

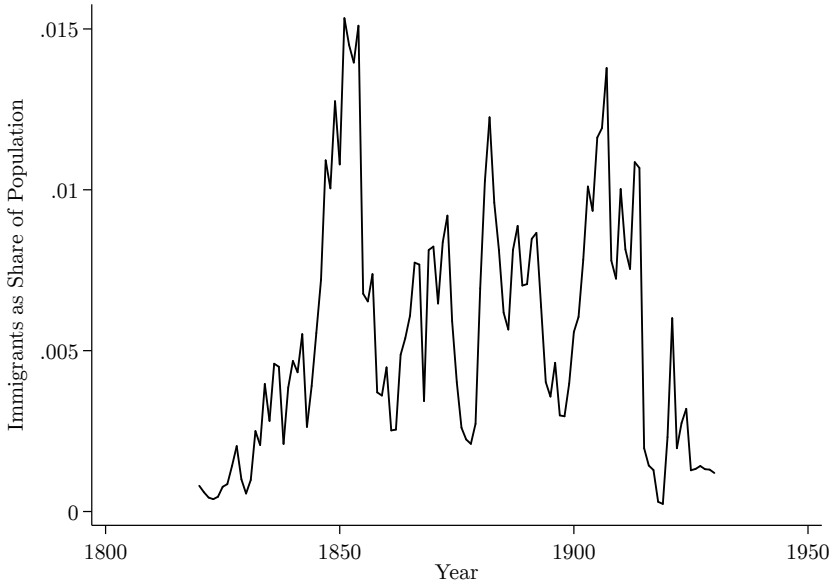
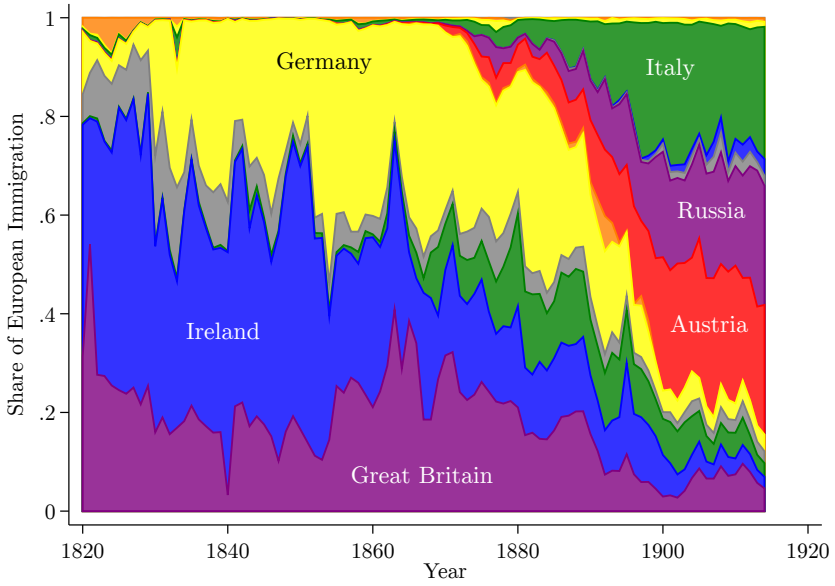
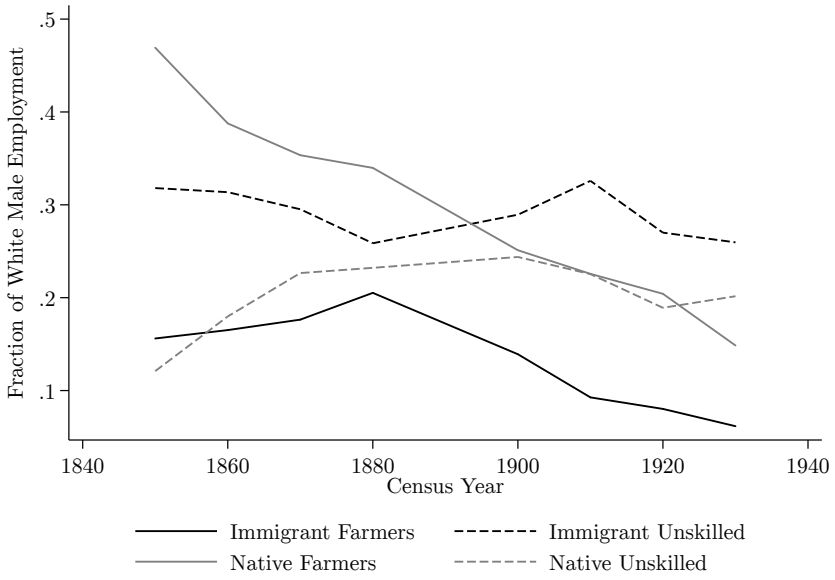


Figure 1(b): European immigration by source country



Notes: Population counts are from Haines and Sutch (2006). Immigration counts are from Barde, Carter, and Sutch (2006). Counts for Austria are from the Barde, Carter, and Sutch (2006) column "other Central Europe," which is central Europe other than Germany and Poland.

Figure 2: Immigrant and native occupation shares



Notes: Data are from 1% samples of the 1850-1930 censuses (excluding 1890) provided by Ruggles et al. (2019). Farming and unskilled occupations defined as in footnote 29. Immigrants restricted to those from European countries. Both natives and immigrants limited to white males ages 18-65.

Figure 3: Selection into successful linkage

Figure 3(a): Natives, 1850-1880

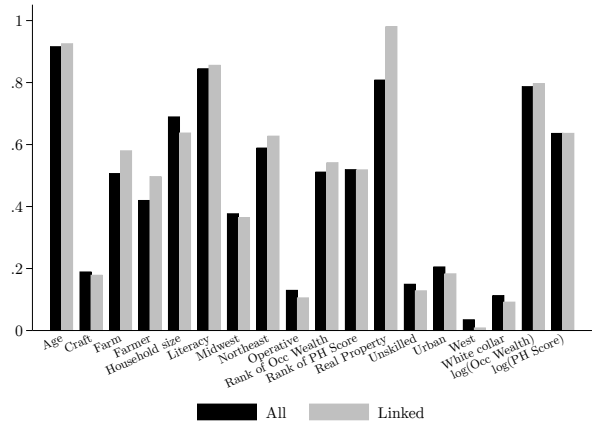


Figure 3(b): Immigrants, 1850-1880

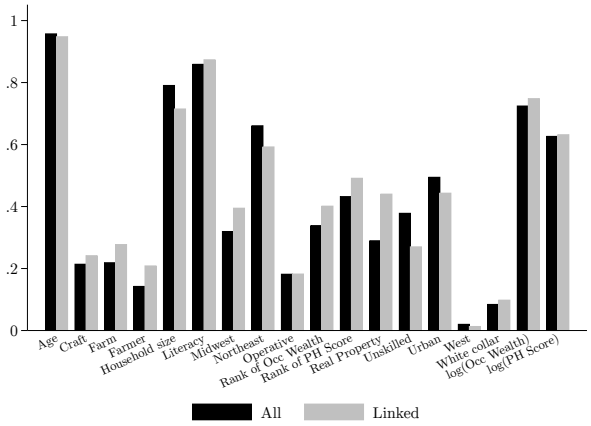


Figure 3(c): Natives, 1900-1930

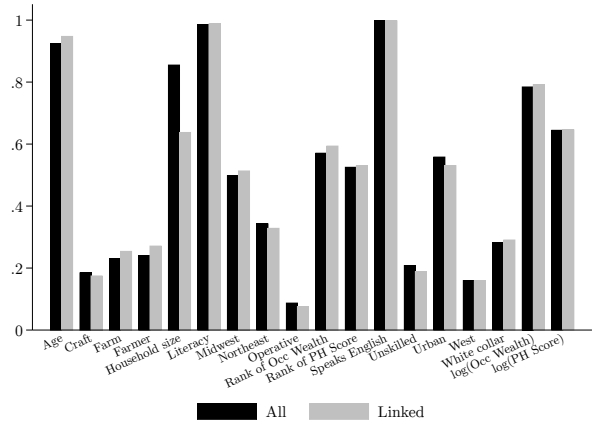
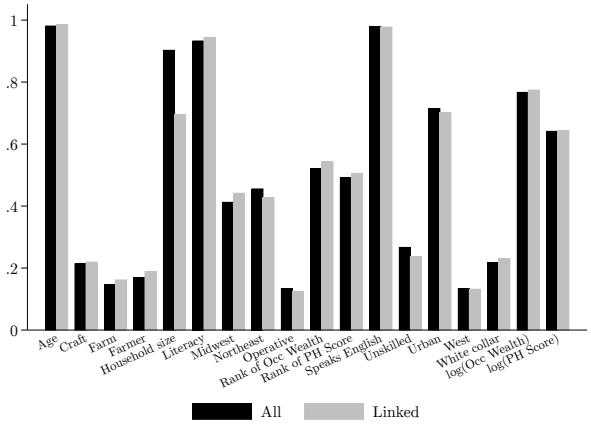


Figure 3(d): Immigrants, 1900-1930



Notes: The bars in each figure reflect the average value of each variable for the successfully linked sample ("Linked") and for the entire sample that we attempted to link, including both successful and failed links ("All"). For the 1850-1880 data the baseline sample and the data are from 1850. For the 1900-1930 data the baseline sample and the data are from 1930. In cases where the mean of a variable is greater than one, we divide that variable by the smallest multiple of 10 needed to generate a mean smaller than one (e.g., age is divided by 30 for the 1850-1880 data).

Figure 4: Occupational distributions

Figure 4(a): 1850

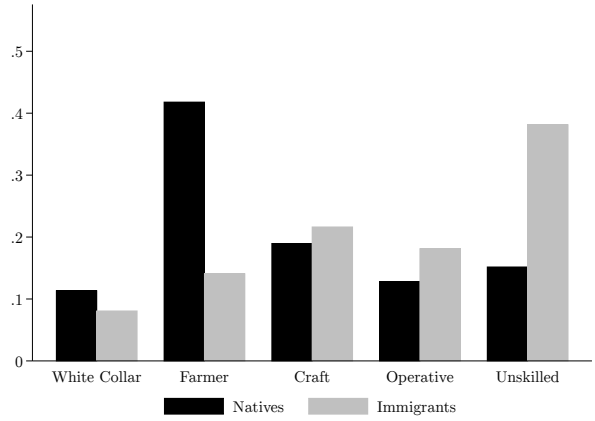


Figure 4(b): 1880

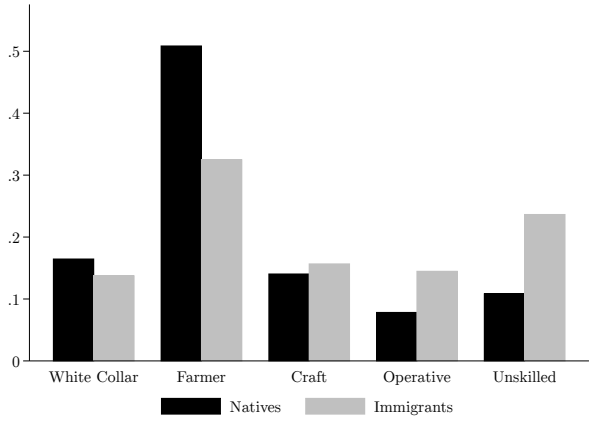


Figure 4(c): 1900

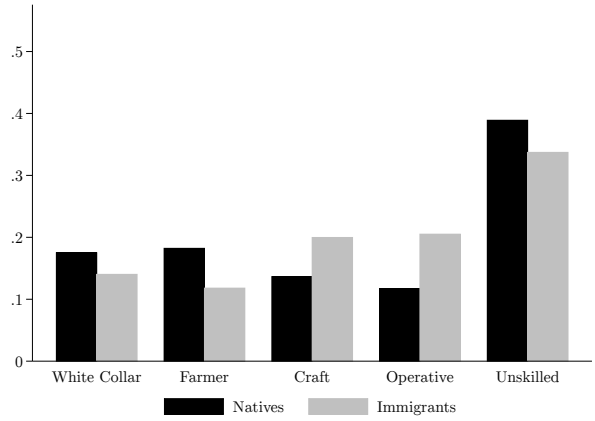
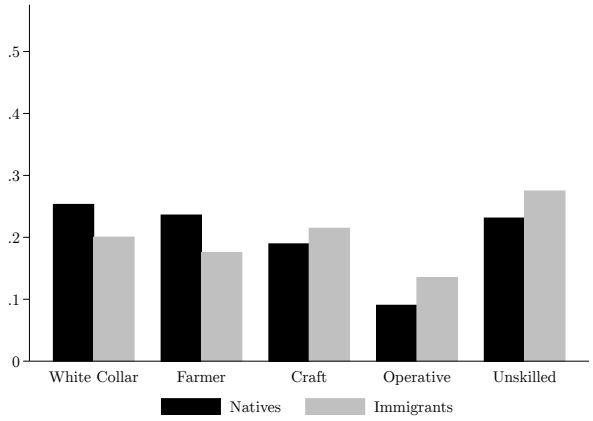


Figure 4(d): 1930



Notes: These graphs show the occupational distribution of natives and immigrants in each year using inverse-probability weights to correct for selection into linkage, but making no other corrections. Occupational categories are defined as in text. Sample limited to individuals with occupations in both years.

Figure 5: Change in occupational rank by initial occupation

Figure 5(a): 1850-1880

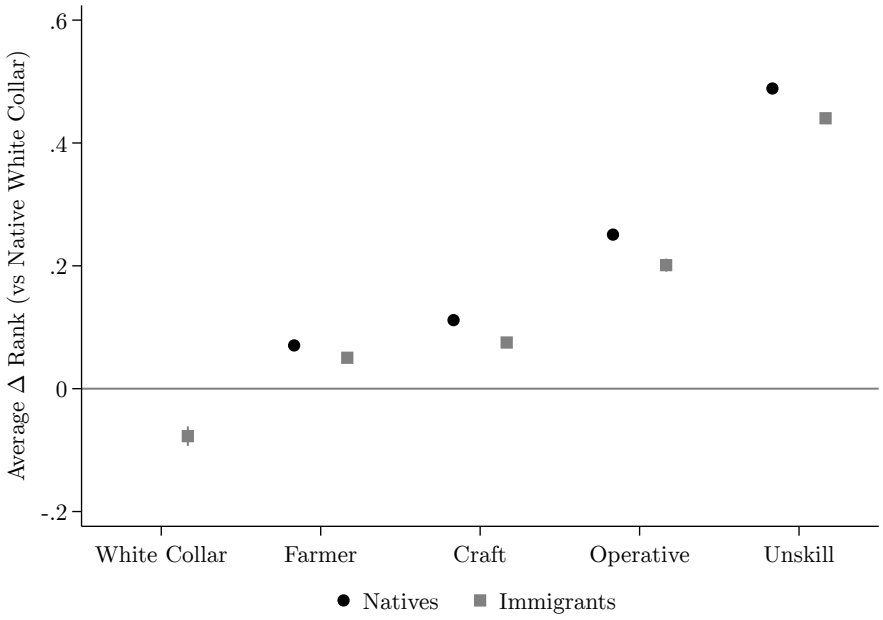
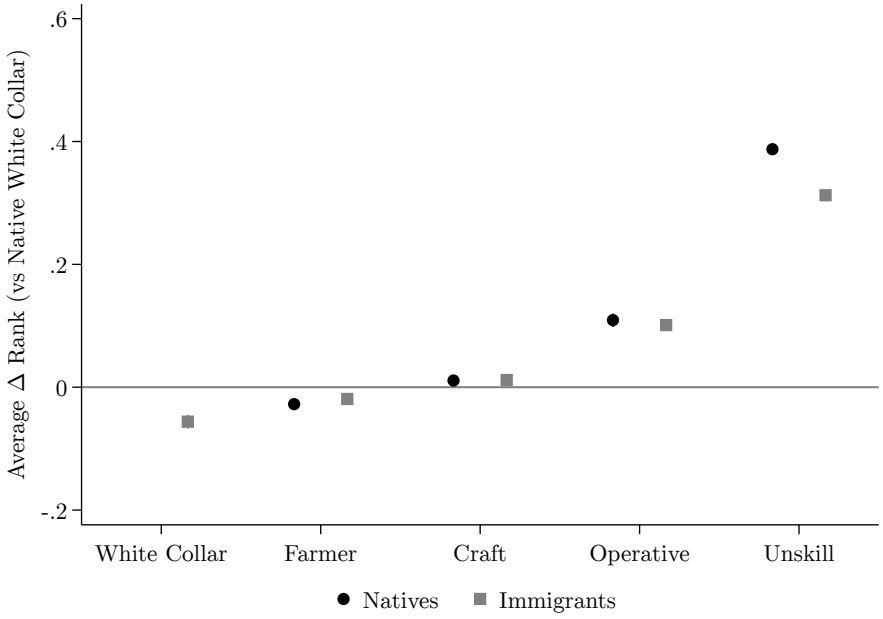


Figure 5(b): 1900-1930



Notes: Each figure presents coefficients from regressing the change in the average occupational rank on nativity-initial occupational category indicators, with native-white collar as the excluded group and controlling for a quartic in age. Robust 95 percent confidence intervals reported (but are so small that they do not exceed width of the point estimate markers). Observations weighted to correct for selection into linkage. Sample limited to individuals with an occupation on both years.

Figure 6: Average occupational rank by country of origin

Figure 6(a): 1850-1880

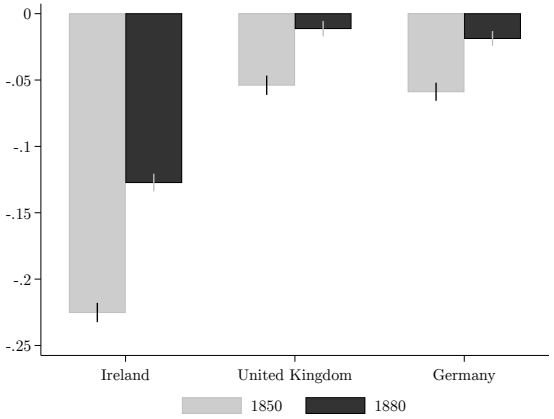
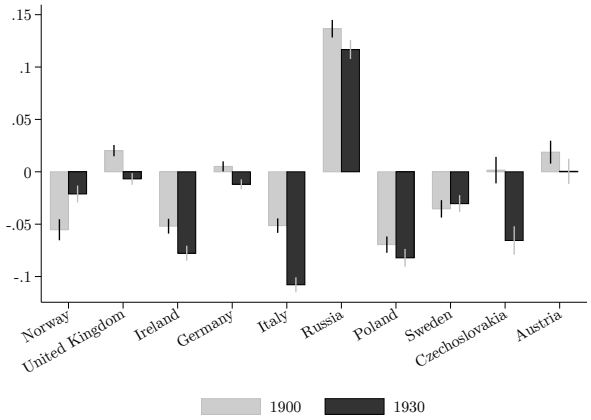


Figure 6(b): 1900-1930



Notes: Each bar represents a coefficient from estimating equation (1) with separate β_t coefficients for each ethnicity and with the average occupational rank measure as the dependent variable. The excluded group is natives. Robust 95 percent confidence intervals reported. Observations weighted to correct for selection into linkage. Sample limited to individuals with occupations in both years

Figure 7: Unskilled occupation by country of origin

Figure 7(a): 1850-1880

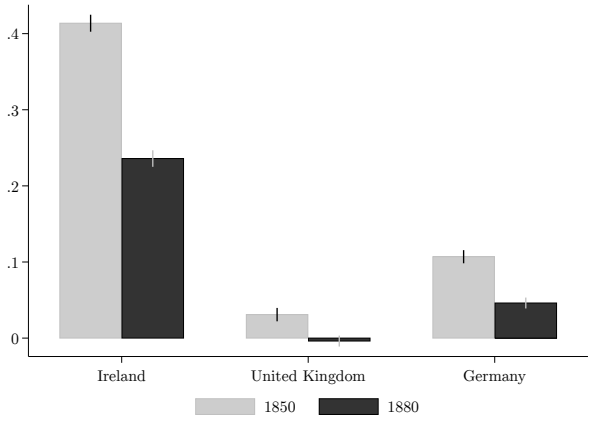
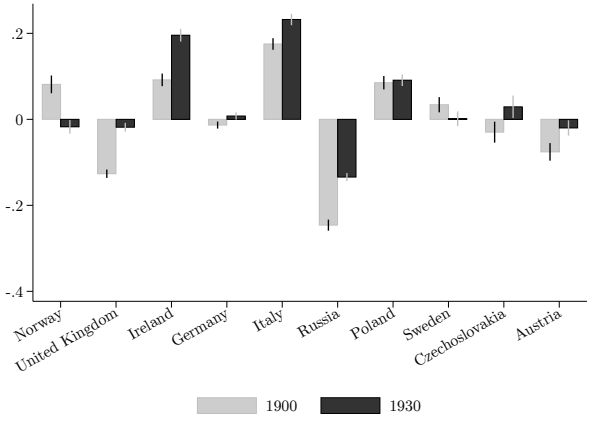


Figure 7(b): 1900-1930



Notes: Each bar represents a coefficient from estimating equation (1) with separate β_t coefficients for each ethnicity and with an indicator for holding an unskilled occupation as the dependent variable. The excluded group is natives. Robust 95 percent confidence intervals reported. Observations weighted to correct for selection into linkage. Sample limited to individuals with occupations in both years.

Tables

Table 1: Rates of successful linkage

	Natives			Immigrants		
	(1)	(2)	(3)	(4)	(5)	(6)
Link	Start	Searched	Linked	Start	Searched	Linked
1850–1880	2,064,491	927,431	238,269	626,320	260,713	35,819
		(0.449)	(0.257)		(0.416)	(0.137)
			[0.115]			[0.057]
1900–1930 (Base 1900)	7,447,320	3,237,275	987,619	2,235,798	1,153,700	178,203
		(0.435)	(0.305)		(0.516)	(0.154)
			[0.133]			[0.080]
1900–1930 (Base 1930)	7,105,156		1,005,051	1,642,477		178,166
			[0.141]			[0.108]

Notes: Numbers in parentheses indicate the fraction of the column that was advanced from the previous step. Numbers in square brackets indicate match rates relative to the full sample in columns (1) and (4). For 1850-1880, numbers are relative to the base year 1850. For 1900-1930, numbers are relative to the indicate base year. For 1850-1880 and 1900-1930 (Base 1900), the sample is limited to non-southern white men aged 18-40 in the base year. For 1900-1930 (Base 1930), the sample is limited to non-southern white men aged 44-74 in the base year excluding unlinked immigrants arriving after 1900; in this case, links to men not aged 18-40 in 1900 are not counted as links.

Table 2: Summary statistics

	1850		1880		1900		1930	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable	Native	Immigrant	Native	Immigrant	Native	Immigrant	Native	Immigrant
Age	27.427	28.409	57.251	58.136	25.843	28.086	55.237	57.710
	(6.450)	(6.179)	(6.679)	(6.536)	(7.147)	(6.765)	(8.291)	(7.520)
log(Occ. Wealth)	7.866	7.231	8.132	7.731	7.304	7.298	7.775	7.643
	(0.922)	(0.961)	(0.868)	(1.019)	(1.036)	(0.961)	(1.024)	(1.029)
log(PH Occ Score)	6.358	6.261	6.418	6.367	6.211	6.290	6.419	6.398
	(0.291)	(0.313)	(0.283)	(0.307)	(0.436)	(0.357)	(0.360)	(0.343)
Unskilled	0.152	0.381	0.108	0.236	0.389	0.337	0.231	0.275
	(0.359)	(0.486)	(0.311)	(0.425)	(0.488)	(0.473)	(0.422)	(0.446)
Farmer	0.418	0.141	0.509	0.325	0.182	0.118	0.236	0.176
	(0.493)	(0.348)	(0.500)	(0.468)	(0.386)	(0.323)	(0.425)	(0.380)
Craft	0.190	0.216	0.140	0.157	0.137	0.200	0.189	0.215
	(0.392)	(0.412)	(0.347)	(0.363)	(0.343)	(0.400)	(0.392)	(0.411)
Operative	0.128	0.181	0.078	0.145	0.117	0.205	0.090	0.135
	(0.334)	(0.385)	(0.269)	(0.352)	(0.322)	(0.404)	(0.287)	(0.341)
White Collar	0.113	0.081	0.165	0.137	0.175	0.140	0.253	0.200
	(0.317)	(0.272)	(0.371)	(0.344)	(0.380)	(0.347)	(0.435)	(0.400)
Average Occ Rank	0.515	0.383	0.593	0.524	0.432	0.453	0.529	0.497
	(0.233)	(0.267)	(0.201)	(0.239)	(0.281)	(0.253)	(0.254)	(0.256)
Occ. Wealth Rank	0.510	0.335	0.641	0.538	0.427	0.437	0.552	0.515
	(0.275)	(0.261)	(0.243)	(0.274)	(0.299)	(0.264)	(0.295)	(0.291)
PH Score Rank	0.519	0.432	0.545	0.510	0.436	0.469	0.506	0.479
	(0.277)	(0.326)	(0.238)	(0.268)	(0.307)	(0.283)	(0.285)	(0.281)
Literacy	0.961	0.917			0.979	0.904	0.984	0.933
	(0.193)	(0.275)			(0.145)	(0.294)	(0.125)	(0.251)
Speaks English					0.989	0.897	0.997	0.980
					(0.107)	(0.304)	(0.057)	(0.140)
Urban	0.197	0.494	0.199	0.453	0.355	0.617	0.513	0.677
	(0.398)	(0.500)	(0.399)	(0.498)	(0.479)	(0.486)	(0.500)	(0.468)
Midwest	0.379	0.316	0.428	0.418	0.575	0.452	0.521	0.416
	(0.485)	(0.465)	(0.495)	(0.493)	(0.494)	(0.498)	(0.500)	(0.493)
Northeast	0.580	0.663	0.496	0.482	0.340	0.461	0.334	0.446
	(0.494)	(0.473)	(0.500)	(0.500)	(0.474)	(0.499)	(0.472)	(0.497)
South			0.032	0.055				
			(0.176)	(0.227)				
West	0.041	0.021	0.045	0.045	0.085	0.086	0.145	0.137
	(0.199)	(0.144)	(0.206)	(0.207)	(0.279)	(0.281)	(0.352)	(0.344)
Old Source						0.422		
						(0.494)		
Years in US						13.300		
						(8.738)		
Observations	206,557	30,538	206,583	30,537	551,776	98,959	566,599	101,462

Notes: For 1850 and 1880, the table includes white males linked from 1850 to 1880 who were ages 18-40 in 1850, who had an occupation in both years, and who did not live in the South in 1850. For 1900 and 1930, the table includes white males linked from 1900 to 1930 who were ages 18-40 in 1900, who had an occupation in both years, and who did not live in the South in either year. Observation numbers are the minimum with data for all variables except literacy, old source, and years in US. Different observation numbers for the same cohort across years are due to occupations that could not be scored. Observations weighted to correct for selection into linkage. Standard deviations in parentheses.

Table 3: Occupational category differences, immigrants vs. natives

	1850–1880			1900–1930		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	1850	1880	Diff.	1900	1930	Diff.
Unskilled	0.238[a]	0.128[a]	−0.111[a]	0.005	0.046[a]	0.041[a]
	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
Farmer	−0.278[a]	−0.187[a]	0.091[a]	−0.107[a]	−0.084[a]	0.024[a]
	(0.003)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)
Craft	0.023[a]	0.018[a]	−0.006[c]	0.053[a]	0.031[a]	−0.022[a]
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Operative	0.053[a]	0.068[a]	0.015[a]	0.092[a]	0.054[a]	−0.038[a]
	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
White Collar	−0.037[a]	−0.026[a]	0.011[a]	−0.043[a]	−0.048[a]	−0.005[c]
	(0.005)	(0.005)	(0.003)	(0.002)	(0.002)	(0.003)

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1
Notes: Robust standard errors in parentheses. Sample limited to individuals with occupations in both years. All specifications include a quartic in age and are weighted by inverse linkage probability, as described in text. Columns (1), (2), (4), and (5) present estimates of β_t from equation (1) for each year with the listed variable on the left-hand side. Columns (3) and (6) present estimates of $\beta_t - \beta_{t-30}$.

Table 4: Dissimilarity indices between natives and immigrants

	(1)	(2)	(3)	(4)
	1850	1880	1900	1930
Dissimilarity	0.3148	0.2133	0.1500	0.1315
	(0.0037)	(0.0039)	(0.0027)	(0.0027)
Observations	237,203	237,203	668,061	668,061

Notes: Dissimilarity indices between natives' and immigrants' occupational distributions in each year, controlling for a quartic in age as described in equation (2). Robust delta method standard errors in parentheses. Observations weighted to correct for selection into linkage. Sample limited to individuals with occupations in both years.

Table 5: Occupational income differences, immigrants vs. natives

	1850–1880			1900–1930		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	1850	1880	Diff.	1900	1930	Diff.
Average Occ. Rank	-0.138^a	-0.069^a	0.069^a	-0.015^a	-0.035^a	-0.020^a
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
log(Occ. Wealth)	-0.661^a	-0.404^a	0.257^a	-0.151^a	-0.164^a	-0.013
	(0.010)	(0.013)	(0.009)	(0.006)	(0.006)	(0.008)
log(PH Occ. Score)	-0.107^a	-0.049^a	0.059^a	0.027^a	-0.021^a	-0.048^a
	(0.004)	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)
Occ. Wealth Rank	-0.183^a	-0.105^a	0.078^a	-0.032^a	-0.046^a	-0.014^a
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
PH Score Rank	-0.094^a	-0.034^a	0.060^a	0.002	-0.024^a	-0.025^a
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1
Notes: Robust standard errors in parentheses. Sample limited to individuals with occupations in both years. All specifications include a quartic in age and are weighted by inverse linkage probability, as described in text. Columns (1), (2), (4), and (5) present estimates of β_t from equation (1) for each year with the listed variable on the left-hand side. Columns (3) and (6) present estimates of $\beta_t - \beta_{t-30}$.

Table 6: Human Capital Differences, Immigrants vs. Natives

	1850–1880			1900–1930		
	(1)	(2)	(3)	(4)	(5)	(6)
Variable	1850	1880	Diff.	1900	1930	Diff.
Literacy	-0.044^a			-0.072^a	-0.052^a	0.020^a
	(0.002)			(0.002)	(0.002)	(0.002)
Numeracy	-0.083^a	-0.102^a	-0.019^a	-0.017^a	-0.025^a	-0.008^a
	(0.003)	(0.003)	(0.005)	(0.002)	(0.002)	(0.003)
Speaks English				-0.091^a	-0.016^a	0.075^a
				(0.002)	(0.001)	(0.002)

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1
Notes: Robust standard errors in parentheses. Sample limited to individuals with occupations in both years. All specifications include a quartic in age and are weighted by inverse linkage probability, as described in text. Columns (1), (2), (4), and (5) present estimates of β_t from equation (1) for each year with the listed variable on the left-hand side. Columns (3) and (6) present estimates of $\beta_t - \beta_{t-30}$.

Table 7: Conditional changes in rank

Variables	(1) 1850–1880	(2) 1900–1930	(3) 1850–1880	(4) 1900–1930	(5) 1850–1880	(6) 1900–1930	(7) 1900–1930	(8) 1900–1930
Immigrant	0.069[a]	−0.018[a]	−0.040[a]	−0.029[a]	−0.044[a]	0.082[a]	−0.017[a]	−0.034[a]
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)
Initial Avg. Occ. Rank			−0.781[a]	−0.712[a]				−0.713[a]
			(0.007)	(0.004)				(0.004)
Old Source							−0.003	0.012[a]
							(0.003)	(0.002)
Observations	237,016	667,717	237,016	667,717	237,011	659,693	667,717	667,717
R-squared	0.028	0.100	0.470	0.427	0.068	0.077	0.100	0.427
Weights					1900	1850		

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1
Notes: Dependent variable is change in average occupational rank. Robust standard errors in parentheses. All specifications include a quartic in age and are weighted by inverse linkage probability. Excluded group in all specifications are natives. Weights indicate that the data are reweighed to match the occupational distribution of the year listed in the last row of the table, for immigrants and natives separately. Sample limited to individuals with occupations in both years.

Table 8: Correlates of change in rank

Variables	(1) 1850–1880	(2) 1900–1930	(3) 1850–1880	(4) 1900–1930	(5) 1850–1880	(6) 1900–1930	(7) 1850–1880	(8) 1900–1930
Immigrant	0.078[a]	0.018[a]	−0.049[a]	−0.038[a]	0.071[a]	0.017[a]	−0.055[a]	−0.041[a]
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Initial Avg. Occ. Rank			−0.789[a]	−0.741[a]			−0.794[a]	−0.745[a]
			(0.003)	(0.003)			(0.003)	(0.003)
Literate in Initial Year	−0.056[a]	−0.015[b]	0.052[a]	0.059[a]	−0.055[a]	−0.015[a]	0.051[a]	0.059[a]
	(0.005)	(0.006)	(0.004)	(0.005)	(0.004)	(0.006)	(0.004)	(0.004)
Urban in Initial Year	−0.022[a]	−0.083[a]	0.023[a]	0.024[a]	−0.020[a]	−0.076[a]	0.027[a]	0.034[a]
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Moved to Urban					0.013[a]	0.025[a]	0.028[a]	0.039[a]
					(0.004)	(0.004)	(0.003)	(0.003)
Moved County					−0.007[a]	−0.014[a]	−0.030[a]	−0.025[a]
					(0.002)	(0.002)	(0.001)	(0.002)
Observations	211,847	667,717	211,847	667,717	211,626	667,712	211,626	667,712
R-squared	0.068	0.151	0.473	0.448	0.095	0.171	0.496	0.466
Initial County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Final County FE	No	No	No	No	Yes	Yes	Yes	Yes

Significance levels: [a] p<0.01, [b] p<0.05, [c] p<0.1
Notes: Dependent variable is change in average occupational rank. Robust standard errors in parentheses. All specifications include a quartic in age and are weighted by inverse linkage probability. Excluded group in all specifications are natives. Sample limited to individuals with occupations in both years.