Can Immigrants Transfer their Skills? Evidence from International Census Records

Adrian Haws

November 29, 2025

Click for updated version

Immigrants tend to earn less, on average, than native-born workers in high-income countries. A possible explanation for this earnings gap is that immigrants do not have the necessary skills to compete in the host-country labor market. Alternatively, they may be unable to transfer their skills across countries. Although qualitative evidence supports the hypothesis of imperfect skill transferability, the literature on immigrant earnings assimilation lacks a comparison between immigrants and natives with similar earnings potential. I construct an international panel dataset of Norwegian and Canadian immigrants during the Age of Mass Migration (1850–1913), making it possible to compare immigrants and U.S. natives possessing similar occupational skills. I document four facts. First, the earnings gap persists when accounting for pre-migration skills. Second, there is a large skill-adjusted earnings gap for French Canadians but no gap for English Canadians, suggesting that limited English proficiency inhibits skill transfer. Third, immigrants continue to earn less than similarly-skilled natives for up to 30 years after arrival. Fourth, children of Norwegian and French Canadian immigrants earn less than children of similarly-skilled U.S. natives.

JEL codes: J62, J15, N21

This research is supported by a joint grant from the Upjohn Institute for Employment Research and the Russell Sage Foundation, as well as the Center for the Study of Inequality and the Cornell Population Center at Cornell University. Author may be contacted by email at aah226@cornell.edu.

"These days, the person behind the wheel of any taxi cab or the counter of any hamburger stand may well be a foreigner who once did something much more elevated for a living."
—Constable (1997), *Washington Post*

1 Introduction

Economists have long debated the extent to which immigrants earn less than the native-born population and improve their relative earnings over time.[1] Existing studies compare the earnings of immigrants and natives with similar observed characteristics, but they lack information on immigrants' occupational skills. Without this information, it is impossible to know whether the earnings gap is driven by differences in skills or by imperfect skill transferability across countries.

Skills may be difficult to transfer across countries because of language differences, sparse social networks, unfamiliarity with business practices, or formal barriers such as occupational licenses (Chiswick, 1978). The imperfect skill transferability hypothesis is grounded in anecdotes of taxi-driving doctors or engineers who earn far less than native-born workers with similar training and qualifications. If immigrants are unable to fully transfer their skills, their full potential earnings and contributions to the economy remain unrealized.

Previous studies have found that immigrants receive lower compensation for their education and work experience than the native-born populations in several high-income countries.[2] However, educational attainment is only a weak proxy for earnings potential because earnings and skills have a high degree of variance within the same education level (Ingram and Neumann, 2006). Additionally, lower returns to immigrants' education could be explained by variation in educational quality across countries (Mattoo et al., 2008).

In this paper, I obtain information on immigrants' occupational skills by linking census records between the U.S. and two of the largest sending countries during the Age of Mass Migration: Canada and Norway.[3] The linked samples include recently-arrived immigrants and

[1] Chiswick (1978) used cross-sectional data from the 1970 U.S. Census to show that immigrants with more experience in the U.S. rapidly closed an earnings gap with natives. Later contributions estimated immigrant earnings assimilation while accounting for cohort characteristics (Borjas, 1985), selective return migration (Lindstrom and Massey, 1994; Lubotsky, 2007; Abramitzky et al., 2014; Rho and Sanders, 2021), labor market context (LaLonde and Topel, 1992; Collins and Zimran, 2023), and race (Villarreal and Tamborini, 2018)

[2] Friedberg (2000) relied on education reported in the Israel census. A similar approach has been done for Canada (Ferrer and Riddell, 2008), the U.S. (Mattoo et al., 2008), and Germany (Basilio et al., 2017)

[3] Recent studies of the Age of Mass Migration (1850–1913) have provided important insights because of the

native-born men in the 1900 or 1910 U.S. Census who are working-age when observed in the U.S. and their country of birth.[4] I combine genealogical records, emigrant lists, and a highly customized machine learning approach to create probabilistic record links that are highly accurate and broadly representative of the study population. I then create a continuous measure of immigrants' counterfactual earnings potential by assigning them to the percentile rank of the occupational earnings distribution that they would inhabit in the U.S. if they continued to hold their last observed occupation in the birth country.[5]

The main estimating equation compares post-immigration earnings ranks between immigrants and natives with similar earnings potential. Instead of estimating a linear rank-rank function, which is common in the mobility literature, I use weights to balance the immigrant and native skill distributions. The estimand is then the weighted average conditional difference, which I estimate separately with respect to either the immigrant or the native distribution (Li et al., 2018). This approach allows for nonparametric variation in the earnings gap across the skill distribution, and is robust to regions of low overlap in the distribution.

I find that immigrants achieve a substantially lower rank in the U.S. earnings distribution than natives with similar earnings potential. Within the first ten years since arrival, both Norwegian and French Canadian immigrants hold occupations ranked 10 percentiles lower, on average, than comparable U.S. natives. These skill-adjusted gaps in earnings ranks are 80% and 65% as large as the unadjusted gaps, respectively, suggesting that only 20-35% of the earnings disadvantage for immigrants can be attributed to differences in skills.

By contrast, Canadian immigrants with English or other ethnic origins do not have lower earnings than natives, with or without adjusting for pre-migration skills. French and English Canadians' divergent outcomes suggest an outsized role for language proficiency, as fewer than one-fourth of French Canadians in the same birth cohorts spoke English proficiently.

I also use links from the Census Tree Project (Buckles et al., 2025) to follow immigrants over their lives in the U.S. By the third census, up to 30 years after arrival, Norwegian immigrants

availability of administrative data, high levels of immigration, and a less complex policy landscape (Abramitzky and Boustan, 2017)

[4]Lower linkage rates and selection into the labor force result in extremely small samples for women.

[5]See Section 3 for a discussion of occupational differences across countries.

continue to achieve an earnings rank about 5 percentiles lower than similarly-skilled natives. I do not observe a statistically significant gap for French Canadians by their second U.S. census.

Finally, I estimates differences in the intergenerational mobility of immigrants and U.S. natives, accounting for parents' pre-migration skills. The sons of Norwegian and French Canadian immigrants achieved a lower average occupational earnings rank than sons of similarly-skilled U.S. natives, suggesting that the gaps in skill transferability persists into the second generation. Additionally, the gap in immigrants' intergenerational mobility is larger than estimates based on fathers' occupations after arrival in the U.S. Thus, the tendency for immigrants' children to be more upwardly mobile than children of the native-born could be due to the fact that their parents are under-placed in the earnings distribution (Abramitzky et al., 2021).

The paper proceeds as follows. Section 2 reviews the existing literature on immigrant skill transferability. Section 3 discusses migrants' motivations during the Age of Mass Migration and the comparative economic development and educational resources across Norway, Canada, and the United States during the study period. Section 4 describes the international panel data that includes immigrants' occupations in the birth country and the U.S. This section also details the occupation-based measures of earnings mobility. Section 5 presents the empirical strategy for measuring imperfect skill transferability. Section 6 presents estimates of the skill-adjusted gap in immigrants' and natives' occupational attainment shortly after arrival, throughout their working lives, and into the second generation. Section 7 concludes.

2 Theory and evidence on immigrant skill transferability

Chiswick (1978) first developed an informal theory of imperfect skill transferability across countries. The core assumption is that some determinants of labor productivity are specific to a particular country. For example, even if a store clerk's task requirements are identical across two countries, an experienced worker will be unable to perform the job if they are not proficient in the destination-country language. Occupational skills may also be difficult to transfer because of sparse social networks, less knowledge about business practices, and lack of recognized certifications. Imperfect skill transferability would result in lower earnings than natives

with the same skills. Immigrants may be able to recover their disadvantage by learning the language, expanding their networks, gaining experience with business practices, and working to gain professional certifications. If a gap in earnings persists, it could be driven by a persistent knowledge deficiency or by discrimination.

Other studies have formally modeled Chiswick's argument. Duleep and Regets (1999) propose that imperfect skill transferability spurs newly-arrived immigrants to invest in human capital at higher rates. Imai et al. (2019) model the impact of language proficiency on the transferability of skills, where immigrants with cognitive skills may transition to manual-intensive jobs if they lack proficiency in the dominant language.

Despite prevalent research and policy debates on immigrant economic assimilation, there is little direct evidence to support or contradict these theoretical hypotheses. Most of the evidence on immigrant skill transferability is derived from surveys that ask immigrants about their previous and current occupations. Using these surveys, Chiswick et al. (2005) and Akresh (2006) show that immigrants to Australia and the U.S. decline in occupational prestige relative to their origin countries. Imai et al. (2019) uses similar data to show that immigrants to Canada often switch from cognitive-intensive to manual-intensive occupations.

Yet there are several challenges to making inferences from retrospective surveys of new immigrants. First, these studies include samples of fewer than 3,000 immigrants dispersed across many source countries, hindering attempts at observing systematic differences across origins. Second, high rates of non-response introduce concerns of selection bias. Third, the surveys do not track immigrants for more than a few years, so their long-run outcomes are unknown. Fourth, and most importantly, the lack of a comparison group makes it impossible to know how much "downgrading" is typical among the native population over the same period.

Ferrie (1997) takes an alternative approach by linking a sample of passenger records between 1840 and 1860 to U.S. Census records to observe transitions across broad occupational categories. This study also suffers from small samples and the lack of a comparison group, as well as severe measurement error in reported occupations. However, I follow the approach of linking publicly-available population records to observe immigrants' occupations before and after immigration, with two key improvements. First, I use occupations reported at the time of residence in the

birth-country census. Second, I compare the occupational mobility of immigrants to the mobility of native-born workers linked to the U.S. census over a similar period.

3 Norway and Canada during the Age of Mass Migration

In the year 1900, the United States had both the largest economy and the highest GDP per capita in the world. Canada had the eighth highest GDP per capita among Western European and North American countries (between the income of Germany and Austria) and Norway was ranked at 13 (between Italy and Spain) (Bolt and van Zanden, 2025). It is reasonable to wonder whether experienced immigrants from less wealthy countries were equipped to fill similar occupations in the destination country. If there were large differences in production technologies across countries, occupational skill requirements might be difficult to compare. Yet during this period, Norway and Canada were both highly industrialized and would have followed similar production practices as the U.S.

Canada and Norway also devoted public resources to education and had strong traditions of skilled craftsmanship. Craftsmen in Europe and North America received up to ten years of training as apprentices and journeymen (Lane, 2005), and some immigrants were recognized for their specialized skills. For example, marble mosaic and terrazzo workers from Alpine foothills were known as the "aristocracy" of Italian Americans because of their prized skills (Grossutti, 2021). Additionally, communities with Danish American dairy farmers adopted more modernized practices (Boberg-Fazlić and Sharp, 2024). Norway and Canada likewise specialized in different industries, such as maritime trade and fishing in Norway, and timber in Canada. Separate from the question of training and experience, immigrants may be exceptionally talented and motivated in ways that are difficult to observe (Chiswick, 1999; Borjas et al., 2019).

During the Age of Mass Migration, Norway was the only country with a higher emigration rate than Ireland. Although early waves of Irish emigrants were refugees from famine, Norwegian emigrants were primarily motivated by economic opportunities. Emigration to America steadily increased after the Homestead Act in 1862 opened western lands to rapid settlement. The peak years of emigration coincided with periods of stagnation in Norway and growth in America. As

a case in point, the highest number of Norwegian immigrants arrived during the 1882 migration season, in the midst of a "long depression" and deflation in Norway and during rapid growth before the U.S. was hit by a recession. Prospective migrants were inspired by "America letters" sent from family and friends who had left, and by opportunities for employment and wealth in contrast to population pressures in Norway.

Norway began large-scale manufacturing in the 1840s, and experienced a second wave of industrialization in the 1860s and 1870s. Norway was a major force in global trade, with the world's fourth largest merchant fleet in 1875 (Grytten, 2008). Because of Norway's dominance in shipping and generations of accumulated knowledge in seamanship, it is likely that Norwegian sailors brought specialized knowledge to the U.S.

Up to 70% of Norway's population was literate by 1870 (Grytten and Basberg) because of investment in local educational resources and a law requiring training in reading. By 1889, seven years of education was mandatory. Twenty one years later, in 1910, only 29 U.S. states had the same standard (Goldin and Katz, 2011).

Canadian immigrants were the third-largest group in the 1900 U.S. Census. Immigrants encountered an economy, geography, and cultural environment in the U.S. that was similar in many ways. Many Canadian immigrants originated from French-speaking communities, including experienced sawyers and loggers.[6]

4 Data

4.1 Linking international census records

Census records show the occupations and other characteristics of recently-arrived immigrants in the U.S., but provides no information on occupations prior to entry. To obtain this information, I construct a new dataset linking immigrants from the U.S. census to the last census when they lived in their country of birth. I use a highly customized machine learning method with information from genealogical records and emigrant lists, resulting in two linked samples: a set of Norwegian immigrant men linked from the 1900 Norway Census to the 1910 U.S. Census, and

[6]Appendix Figure B2 illustrates that only about one-quarter of French Canadians were proficient in English in the 1911 Canada Census.

a set of Canadian immigrant men linked from the 1881 Canada Census to the 1900 U.S. Census (Ruggles et al., 2025a; University of Tromsø, 2008; Dillon, 2008; Ruggles et al., 2025b, 2024). I focus on men for this analysis because a combination of lower linkage rates and selection into the labor force would result in very small samples for women.

An important aspect of this study is to compare the rate of immigrants' occupational mobility to the rate of natives' mobility. Thus I also linked U.S.-born men in the same age range to previous census years that coincide with those of immigrants: 1880 is linked to 1900, and 1900 is linked to 1910. For these links, I use the Census Tree Project (Buckles et al., 2025), which provides the largest set of historical U.S. census links at a high level of accuracy. I also follow the practice of related papers by restricting both the immigrant and native sample to white men living outside the U.S. South (Abramitzky et al., 2014; Collins and Zimran, 2023).

To link immigrants, I first select a set of Norway-born men in the 1910 U.S. Census who immigrated in 1901 or later and a set of Canada-born men in the 1900 U.S. Census who immigrated in 1891 or later.[7] I additionally restrict these samples to men who were between the ages of 18 and 65 in both census years.

Previous efforts to link international census records have faced low match rates and potential accuracy concerns. Abramitzky et al. (2012) linked 5% of a sample of immigrants from the 1900 U.S. Census to the 1865 Norway Census using a series of rules for obtaining a unique match and discarding non-unique matches.[8] Connor (2019) modifies this approach to improve accuracy, but links only 3% of a sample of immigrants in the 1910 US Census to the 1901 Ireland Census.[9] Yet Ó Gráda et al. (2023) find that the algorithm developed by Abramitzky et al. (2012) results in up to one-half false positives when linking Irish immigrants across U.S. census records.

Why is it so difficult to link immigrants? Without unique identifiers such as Social Security Numbers, the researcher relies on information such as names, birth years, and birth places, as well as contextual information such as residence places and family members. The primary

[7]The 1900 Norway Census was conducted in December 1900, so immigrants who arrived in 1900 should not be tabulated in Norway.

[8]Abramitzky et al. (2012) report 2,613 links to the 1865 Norway Census, of the total 49,101 Norwegian-born males ages 38 to 50 in the 1900 US Census who report immigrating between 1866 and 1900.

[9]Connor (2019) reports 1,953 links to the 1901 Ireland Census, of the total 70,889 Irish-born males ages 15 to 34 in the 1910 US Census who report immigrating between 1902 and 1910.

challenges are common in other settings but to a larger extent for immigrants:

1. Non-unique matches; some of this results from name Americanization ("Jan," "Johan," and "Jonas" lose dimension as "John"), but many Norwegian immigrants also have common surnames such as "Petersen" and "Olsen".
2. Unstable characteristics; immigrants often choose different names in the new destination (Biavaschi et al., 2017) and English-speaking enumerators misspell unfamiliar names (Ghosh et al., 2024). Also, historical census records often round ages to the nearest five years. An especially acute challenge for Norwegians is that they often adopted the traditional name of their farm as a surname after immigrating.
3. Lack of contextual information; 100% of the sample changed countries of residence, and many immigrants came to the U.S. without family members.

I use several strategies to address these challenges. First, I use a machine learning algorithm to flexibly learn about a wide range of features among a set of verified ground truth links. These verified links were created through a collaborative Family Tree on the FamilySearch.org, where users have access to vital records and their own personal information about their ancestors. The links are highly accurate and have been used to establish links across U.S. census records as part of the Census Tree Project (Price et al., 2021; Buckles et al., 2025). Second, I use information from the Family Tree links and historical context to extract crucial information from potential links, including 10 variations of last name comparisons based on patronyms (names ending in "-sen"), surnames, and farm names. I also include features of coresident family members. Third, I incorporate information from a subset of Norway census records that I have linked to lists of emigrants to provide a signal that a match exists. Fourth, I use existing links across the 1900 and 1910 Norway Census from the Norwegian Historical Population Register (Holden et al., 2025) as indicators that a match to the U.S. is not likely to exist. The links for Canadian immigrants are only from the Family Tree, but I will expand these using my machine learning approach. The linking process is described in more detail in Appendix A.

Among 26,958 immigrants, I link 3,026 for a match rate of 11.2%.[10] Roughly half (1,411) were established independently of the Family Tree links. There is a tradeoff between match rate

[10]This is slightly more than the 2,889 reported in Table 1 because occupations must be non-missing.

and accuracy; I selected links conservatively, with an out-of-sample accuracy of 93.4%.

Because these linkages are not representative of the population of interest, I reweight the linked individuals by the inverse of their probability of being linked. I estimate the following probit equation separately for immigrants and natives:

$$\begin{aligned} P(Linked_{ijk} = 1) = \Phi\big(&\alpha_j + \gamma_k + \theta_1 Married_i + \theta_2 Age46to65_i \\ &+ \theta_3 Married_i * Age46to65_i + \theta_4 ImmRelative_i \\ &+ \theta_5 g(NameLastUniq_i) + \theta_6 g(NameFirstUniq_i)\big) \end{aligned} \tag{1}$$

for individual i in census region j (Midwest, West, Northeast, South) with occupational category k (white collar, manual, farmer, craftsman or operative). The variable $ImmRelative_i$ indicates where a relative who recently immigrated from Norway is present in the household. The uniqueness of first and last names is measured using the normalized log rarity, and $g(.)$ is a quadratic function.[11] These variables were chosen because of their association with linkage, but the equation is somewhat sparse because of the small sample size. Appendix Figure A5 shows that the propensity scores have common support across the linked and unlinked populations.

The weighting procedure is only valid if the propensity score function is correctly specified. Although this condition is impossible to prove, Figure 1 shows reassuring results from a validation exercise. Panel (a) shows, unsurprisingly, that the weighted sample is more representative in terms of the variables used in weighting. Panel (b) then takes a set of variables *not* used in weighting, and shows that the weighted sample better represents these as well. Together, these results provide evidence that the propensity score function is associated with linkage and does not suffer from severe omitted variables bias.

[Figure 1 here]

In addition to the main linked sample, I construct two types of samples with further linkages. To study immigrants' long-run outcomes, I link them over up to 30 years after arrival. For

[11] The normalized log rarity is bounded by 0 and 1, increasing in uniqueness, and less skewed than other measures such as inverse frequency. This measure if computed as $-ln(namecount/N)/ln(N)$ where $namecount$ is the number of observations with the name and N is the total number of observations.

Norwegian immigrants, this involves linking from the 1910 U.S. Census to the 1920 and 1930 census years. For Canadian immigrants, I only link the 1900 U.S. Census to the 1910 Census because the sample is older. I also link the sons of immigrants to their adult census records 30 years after childhood, among those who were 16 or younger in the starting census year. For each of these linked samples, I construct new weights using the propensity of being linked to all relevant census years.

4.2 Occupation and earnings data

Skill transferability is difficult to measure. A previous approach classifies all occupations into broad categories such as "manual-intensive" or "cognitive-intensive" (Imai et al., 2019) or as "white-collar", "skilled", "semi-skilled", "unskilled", or "farmer" (Ferrie, 1997). With this approach, an immigrant may lack skill transferability if they transition from a "semi-skilled" occupation in the origin country to an "unskilled" occupation in the destination country. Although this approach is straightforward, it risks missing mobility within categories. Additionally, it is not clear how large of a gap exists in earnings or training between "white-collar" and "skilled", or even that every "white-collar" job is preferable.

I construct a continuous measure of relative mobility in occupational skills that can be calculated for international migrants. This measure allows me to ask whether an immigrant's current occupation pays better or worse than the occupation of a native-born worker with the same earnings ability. I build this measure in three steps. First, I convert detailed occupational titles, both inside and outside the U.S., into the amount that a U.S. worker could expect to earn in that occupation. Second, I observe the distribution in occupational earnings for men in the U.S., separately at every age. Third, I assign each worker, both inside and outside the U.S., to the percentile rank of their age-specific distribution that they *would earn* if they held the same occupation within the U.S. If an immigrant switches occupations but earns the same as they would earn with their previous occupation, there is no change in the income rank measure.[12]

Individual-level occupations are recorded in each of the censuses in my linked dataset (an example is shown in Appendix Figure B1.) Occupational titles have been standardized across

[12]I refer to income and earnings interchangeably, but income always refers to wage income.

countries and over time using the Historical International Standard Classification of Occupations (Leeuwen et al., 2002). I convert each occupation into the typical amount a worker could expect to earn in the U.S. annually, using estimates of mean income by occupation based on four sources from around the year 1900 (Preston and Haines, 2014; Collins and Wanamaker, 2022).[13] I then compute the percentile ranks in the occupational income distribution within sex, age, and census year in the U.S., as rank-based measures are simple to interpret and not sensitive to means. An immigrant can simultaneously move downward in relative mobility while earning more real income after migration.

Although the rank-based occupational income measure is useful, it has weaknesses. First, it does not capture income differences within occupations, which may understate immigrant-native earnings differences if immigrants earn lower wages for the same work. Second, it is a noisy measure of occupational skills because it is measured once and up to 10 or 20 years prior to migration. However, this is likely to result in attenuation bias.

4.3 Sample characteristics

Table 1 shows the characteristics of the linked sample with weights for improved representativeness. Compared to U.S. natives in the 1910 sample, Norwegian immigrants are more likely to have an occupational background as manual workers, craft workers, or farmers. However, these occupations in Norway only translate to a four-point lower occupational income rank than native-born men. Canadian immigrants with a French ethnic origin also have a lower rank than U.S. natives, by about six points, and are much more likely to be manual workers. However, English-origin Canadian immigrants have similar occupational backgrounds and nearly identical income ranks on average. Immigrant-native gaps in occupational income ranks are much more pronounced after arrival in the U.S. Within the first ten years after arrival, Norwegians have a 13-point lower income rank and French Canadians have a 16-point lower income rank. English Canadians only have one point lower average occupational income ranks.

[Table 1 here]

[13] Economic historians have developed several ways to impute earnings from occupations because the U.S. census does not directly report income until 1940. The advantage of the Preston-Haines imputation is its proximity to the time period under study, as the more common alternatives are based on 1940 or 1950.

Figure 2 shows the prevalence of upward and downward occupational mobility for immigrants and native workers, with respect to their position in their birth country. Although natives most commonly had very little change in their occupational income rank from 1900 to 1910 (shown in panel (a)), there is a wide distribution of upward and downward mobility. The distribution in rank differences is much less concentrated among Norwegian immigrants, however. There is somewhat of a peak in the prevalence of upward mobility of about 25 percentile ranks, but downward mobility is much more common.

Unsurprisingly, the mobility distribution between 1880/1881 and 1900 is less concentrated around zero for both natives and immigrants (shown in panel (b) of Figure 2). For English Canadians, there is more upward and downward mobility than U.S. natives, but it is not clear which direction prevails. However, French Canadians are clearly downwardly mobile on average, with the most common difference being about -20 percentile ranks.

[Figure 2 here]

5 Empirical strategy

5.1 Reweighting the occupational earnings distributions

The goal of my analysis is to understand whether immigrants work in lower-paying occupations than natives with similar occupational experience. A possible way to make this comparison would be to use a linear control for the income rank of individuals' occupations in their country of birth, and to interact the linear term with an indicator for immigrant status. This would be similar to the rank-rank regressions used in studies of intergenerational mobility (e.g. Chetty et al. (2014)). However, this approach requires a rigid functional form assumption, and the estimates may be biased if there is low overlap in the distributions. I overcome both issues by constructing weights to balance the occupational earnings distributions across immigrant status.

The random variable $Z_i \in \{1, 0\}$ indicates whether person i is an immigrant. The functions $f_1(x)$ and $f_0(x)$ define the population density of the birth-country income ranks for $Z_i = 1$ and $Z_i = 0$ respectively. Balancing weights, are then given by a function $w(x)$ such that the weighted population densities are equivalent: $f_1(x)w_1(x) = f_0(x)w_0(x)$ (Li et al. 2018).

Balancing weights are given by:

$$\begin{cases} w_1(x) = \dfrac{h(x)}{e(x)}, & Z_i = 1 \\ w_0(x) = \dfrac{h(x)}{1 - e(x)}, & Z_i = 0 \end{cases}$$

where $e(x) = Pr(Z_i = 1 | Rank_i = x)$ is the propensity score function and $h(x)$ is a custom function depending on the target population. I estimate the function $e(x)$ using a cubic spline with seven knots, where the sample includes U.S. natives and either Norwegian, French Canadian, or English Canadian immigrants. A critical point is that the propensity score function is estimated on a sample with linkage weights applied (Ridgeway et al., 2015). Appendix Figure B3 plots the propensity scores by group and demonstrates that there is common support.

The choice of the $h(x)$ function depends on the target population for the estimand (Li et al., 2018). For a descriptive estimand akin to the "average treatment effect for the treated population," the h function is $e(x)$ and the weights are $w_1(x) = 1$ and $w_0(x) = e(x)/(1 - e(x))$. For an estimand akin to the "average treatment effect for the control population," the h function is $1 - e(x)$ and the weights are $w_1(x) = (1 - e(x))/e(x)$ and $w_0(x) = 1$.

Figure 3 demonstrates the original and weighted income distributions (with linkage weights already applied). Panel (a) shows the distributions with Norwegian immigrants as the target population. The weighted density $f_0(x)w_0(x)$ for natives closely tracks the original density $f_1(x)$ for immigrants, with some gaps in the lower part of the distribution. Panel (b) shows the distribution with U.S. natives as the target population, again showing that the weighted density $f_1(x)w_1(x)$ mostly tracks the original density $f_0(x)$. Reweighted distributions for French Canadians and English Canadians are shown in Appendix Figures B4 and B5.

[Figure 3 here]

5.2 Estimating equation

The main estimating equation is an ordinary least squares function with the following form.

$$IncomeRank_i = \alpha + \beta Immigrant_i + \epsilon_i \tag{2}$$

The variable $IncomeRank_i$ is the income percentile rank of the occupation held by person i within their age- and year-specific income distribution. In the baseline model, β gives the average difference in the income percentile ranks between a particular group of recent immigrants and U.S. natives. The estimation sample includes working-age men who have been linked to a prior census year in their birth country, either the U.S., Norway, or Canada. To account for selection into linkage, I reweight all observations using the weights v_i. I use heteroskedasticity-robust standard errors.

I then account for differences in pre-migration skills by applying balancing weights, with combined linkage and balancing weights $u_i = v_i * w_i$. The interpretation of the estimated β varies based on the target of the balancing weights w_i. The most straightforward interpretation uses immigrants as the target population, providing the expectation of the immigrant-native gap in earnings rank across the skill distribution of immigrants. I also provide estimates using natives as the target population.

6 Skill-adjusted immigrant earnings differences

Three questions are at hand. First, do recently-arrived immigrants earn less than native-born workers with similar earnings ability? Second, if so, does this disadvantage persist with more experience in the U.S.? Third, is there a difference in the intergenerational mobility of immigrants and natives with similar ability? I will also discuss implications of varying results across source countries and ethnic groups.

6.1 Short-run outcomes

I first estimate the immigrant-native gap in income ranks without accounting for pre-migration skills. Column 1 of Table 2 shows these results. Norwegian immigrants with fewer than 10 years in the U.S. work in occupations with an average of 13.1 lower percentile ranks in the income distribution. The gap for French Canadian immigrants is even larger, at 16 lower percentile ranks. In sharp contrast, English Canadians only have 1.1 lower percentile ranks on average, and this is not statistically distinguishable from zero.

[Table 2 here]

Column 2 of Table 2 reports estimates with the U.S. native distribution in occupational income ranks reweighted to represent the immigrant distribution. These estimates account for differences in the earnings ability of natives and immigrants, based on their occupations in the same pre-migration census year, and provide a descriptive estimate analogous to a "treatment on the treated" effect. Any gap for English Canadians disappears completely when accounting for pre-migration skills. However, most of the gaps for Norwegians and for French Canadians remain. Both groups have occupational income that is 10.4 percentile ranks lower than natives when accounting for skills. Alternative estimates with data weighted to represent the native distribution (shown in column 3) or using fixed effects in the ventile rank of the income distributions (shown in column 4) show similarly large gaps.

A clear result from Table 2 is that the experience for French Canadian immigrants is more similar to that of Norwegians than to their countrymen. Language is the critical difference between French and English Canadians, who otherwise share geography, economic development, proximity, and public institutions.[14] This finding concurs with studies showing positive effects of host-country language proficiency on earnings (Dustmann and Van Soest, 2002; Shields and Price, 2002; Bleakley and Chin, 2004; Heller and Mumma, 2023) and the possible role of language in skill transferability (Chiswick and Miller, 2002).

6.2 Long-run outcomes

Some of the causes of imperfect skill transferability could wane over time if immigrants are able to learn the language, build networks, and gain the other information and credentials necessary for success. However, some of these barriers may be difficult to overcome. I track the long-run outcomes of immigrants by using the Census Tree Project to link them to later census years in the U.S. The full panel for immigrants and natives observed in the 1910 U.S. Census covers 1900–1930 and is restricted to ages 28–45 in 1910. The full panel for immigrants and natives observed in the 1900 U.S. Census covers 1880–1910 and is restricted to ages 38–55 in 1900. Importantly,

[14] As shown in Appendix Figure B2, most Canadians with French ethnicity did not have conversational English skills in this era.

estimates based on the first census after immigration may vary from Table 2 because of selection into return migration and because the panel sample omits men who were older.

Figure 4 reports estimates of the skill-adjusted differences in occupational income ranks between immigrants and natives. I use the regression specification from Equation 2 with the pre-migration income distribution for natives re-weighted to the immigrant distribution. I also apply linkage weights and use robust standard errors, and plot the 95% confidence intervals.

Similar to the previous estimates, the first census after immigration shows large and negative gaps between immigrants and natives with similar skills. The picture is somewhat unclear for French Canadians by the second census because the confidence intervals are very large. If the gap indeed does disappear, it would suggest that French Canadians are able to eventually apply their skills by overcoming language deficiencies. For Norwegian immigrants, a gap of approximately 5 percentile ranks persists in the third census, up to 29 years after immigration.

[Figure 4 here]

6.3 Second-generation outcomes

Theory on intergenerational mobility posits that children inherit earnings ability from their parents (Mogstad, 2017; Becker et al., 2018). If parents' true earnings ability is not reflected in their labor earnings, the association between parent and child outcomes is biased downward. Thus, the result that immigrants earn less than similarly-skilled natives has direct implications for previous estimates of immigrants' intergenerational mobility (Abramitzky et al., 2021).

I study intergenerational mobility by linking sons of immigrants and natives from the first census after arrival to their census in adulthood 30 years later using the Census Tree Project. The outcome variable for each father is the average of their sons' percentile rank in the income distribution, measured after 30 years. I estimate the following equation.

$$MeanIncomeRankSons_i = \alpha + \beta Immigrant_i + \epsilon_i \tag{3}$$

Table 3 reports two sets of intergenerational mobility estimates: conditional on father's income rank in the U.S., and conditional on father's income rank in the birth country. I will only

discuss estimates with immigrants as the target population for reweighting, but estimates are also reported using weights for the native distribution.

When accounting for occupations in the birth country, the intergenerational mobility of immigrants from non-English-speaking backgrounds is lower than natives, with the opposite result for English Canadians. Column 3 of Table 3 reports that sons of Norwegian immigrants are less mobile than sons of natives (4.4 percentiles lower). The estimate is similarly large and negative for sons of French Canadian immigrants, although statistically insignificant. Yet sons of English Canadians exceed their father's birth-country income rank by 2.2 percentiles more than sons of natives ($p < .1$).

Table 3 also shows that imperfect skill transferability causes post-immigration estimates of intergenerational mobility to be biased downward. When accounting for fathers' pre-migration occupations, the estimate for Norwegians is 2.2 percentiles lower ($p < .1$) than when accounting for post-migration occupations (as shown in column 1). The same should be true for French Canadians, but the current sample of 127 linked fathers and sons makes this difficult to measure. Fitting with the lack of within-skill earnings gaps, sons of English Canadians have no statistically significant difference based on when their fathers are observed.

[Table 3 here]

7 Conclusion

This study is the first to use international linked census records to study the transferability of immigrants' skills. My findings confirm theoretical hypotheses that skills are not perfectly transferable across countries. Indeed, I find that most of the gap in immigrants' and natives' earnings in my study population is not driven by a skills gap, and that it persists when comparing workers with similar earnings potential. An important insight is that even immigrants from Canada, with all of its similarities with the United States, may be unable to transfer their skills if they have limited English language proficiency. Yet English Canadians are able to transfer their skills, as predicted by theory.[15] I also show that estimates of the difference in

[15] Chiswick et al. (2005) argue that "[i]mmigrants from countries very similar to the destination — for example, an English-speaking Canadian moving to the United States — may experience little or no downward mobility on

intergenerational mobility between immigrants and natives are biased downward if they do not account for immigrants' skill downgrading after arrival.

Many countries have policies that favor high-skilled immigrants, which is justified by evidence that immigrants boost productivity (Moser et al., 2014; Peri et al., 2015). But my findings suggest that skilled immigrants may not reach their full productive potential. Improving occupation skill matching for immigrants may also reduce competition with natives for low-wage jobs, which is a salient policy concern (Ottaviano and Peri, 2012; Dustmann et al., 2013). Thus, skill transferability has direct relevance for immigration policy goals.

An existing literature suggests factors that could improve the transferability of skills. Several studies show a strong relationship between language learning and wage growth (Dustmann and Van Soest, 2002; Shields and Price, 2002; Berman et al., 2003). There is also evidence that earnings rise with increased social connections with the native-born population (Engdahl et al., 2024).[16] Additionally, immigrants earn a large wage premium from obtaining an occupational license, although language ability is a strong predictor of licensing (Cassidy and Dacass, 2021). A growing literature also studies the role of obtaining legal documentation.[17] Strengthening the contributions of skilled immigrants has the potential to benefit both immigrants and the countries in which they live.

migration."

[16]The need for broad social connections could be related to the negative effects of residential ethnic enclaves (Xie and Gough, 2011; Eriksson, 2020; Abramitzky et al., 2024). However, social connections with natives and with immigrants are not zero-sum, and some evidence shows positive effects of ethnic enclaves (Edin et al., 2003).

[17]The legal standing of immigrants in the U.S. has changed many times through policies such as the Immigration Reform and Control Act of 1986 and the Deferred Action for Childhood Arrival program starting in 2012. Documented immigrants earn more than those who are undocumented (Rivera-Batiz, 1999; Borjas and Cassidy, 2019), and there is also an earnings benefit to gaining naturalized citizenship (Bratsberg et al., 2002).

References

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

Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson. Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, 102(5):1832–1856, August 2012.

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

Ran Abramitzky, Leah Boustan, Elisa Jacome, and Santiago Perez. Intergenerational Mobility of Immigrants in the United States over Two Centuries. *American Economic Review*, 111(2): 580–608, February 2021.

Ran Abramitzky, Leah Boustan, and Dylan Shane Connor. Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office. *The Journal of Economic History*, 84(2):352–394, June 2024.

Ilana Redstone Akresh. Occupational mobility among legal immigrants to the united states. *International Migration Review*, 40:854–884, 12 2006.

Leilanie Basilio, Thomas K. Bauer, and Anica Kramer. Transferability of Human Capital and Immigrant Assimilation: An Analysis for Germany. *LABOUR*, 31(3):245–264, 2017.

Gary S. Becker, Scott Duke Kominers, Kevin M. Murphy, and Jörg L. Spenkuch. A Theory of Intergenerational Mobility. *Journal of Political Economy*, 126(S1):S7–S25, October 2018.

Eli Berman, Kevin Lang, and Erez Siniver. Language-skill complementarity: returns to immigrant language acquisition. *Labour Economics*, 10(3):265–290, June 2003.

Costanza Biavaschi, Corrado Giulietti, and Zahra Siddique. The Economic Payoff of Name Americanization. *Journal of Labor Economics*, 35(4):1089–1116, October 2017.

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

Nina Boberg-Fazlić and Paul Sharp. Immigrant Communities and Knowledge Spillovers: Danish Americans and the Development of the Dairy Industry in the United States. *American Economic Journal: Macroeconomics*, 16(1):102–146, January 2024.

Jutta Bolt and Jan Luiten van Zanden. Maddison-style estimates of the evolution of the world economy: A new 2023 update. *Journal of Economic Surveys*, 39(2):631–671, 2025.

George J Borjas. Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of Labor Economics*, 3:463–489, 1985.

George J. Borjas and Hugh Cassidy. The wage penalty to undocumented immigration. *Labour Economics*, 61:101757, December 2019.

George J Borjas, Ilpo Kauppinen, and Panu Poutvaara. Self-selection of Emigrants: Theory and Evidence on Stochastic Dominance in Observable and Unobservable Characteristics. *The Economic Journal*, 129(617):143–171, January 2019.

Bernt Bratsberg, Jr. Ragan, James F., and Zafar M. Nasir. The Effect of Naturalization on Wage Growth: A Panel Study of Young Male Immigrants. *Journal of Labor Economics*, 20 (3):568–597, July 2002.

Kasey Buckles, Adrian Haws, Joseph Price, and Haley E. B. Wilbert. Breakthroughs in historical record linking using genealogy data: The Census Tree project. *Explorations in Economic History*, 98:101717, October 2025.

Hugh Cassidy and Tennecia Dacass. Occupational Licensing and Immigrants. *The Journal of Law and Economics*, 64(1):1–28, February 2021.

Raj Chetty, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. Where Is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4):1553–1624, 2014.

Barry Chiswick. Are Immigrants Favorably Self-Selected? *American Economic Review*, 89(2): 181–185, May 1999.

Barry R Chiswick. The effect of americanization on the earnings of foreign-born men. *Journal of Political Economy*, 86:897–921, 1978.

Barry R. Chiswick and Paul W. Miller. Immigrant earnings: Language skills, linguistic concentrations and the business cycle. *Journal of Population Economics*, 15(1):31–57, January 2002.

Barry R. Chiswick, Yew Liang Lee, and Paul W. Miller. A longitudinal analysis of immigrant occupational mobility: A test of the immigrant assimilation hypothesis. *International Migration Review*, 39:332–353, 2005.

William J. Collins and Marianne H. Wanamaker. Data and code for: African American intergenerational economic mobility since 1880 – Collins and Wanamaker (2021), June 2022. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

William J. Collins and Ariell Zimran. Working Their Way Up? US Immigrants' Changing Labor Market Assimilation in the Age of Mass Migration. *American Economic Journal: Applied Economics*, 15(3):238–269, July 2023.

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

Pamela Constable. Facing the Harsh Reality of the American Dream. *The Washington Post*, June 29, 1997.

Lisa Dillon. 1881 canadian census project, north atlantic population project, and minnesota population center. national sample of the 1881 census of canada (version 2.0). montréal, qc: Département de démographie, université de montréal [distributor], 2008.

Harriet Orcutt Duleep and Mark C. Regets. Immigrants and Human-Capital Investment. *American Economic Review*, 89(2):186–191, May 1999.

C. Dustmann, T. Frattini, and I. P. Preston. The Effect of Immigration along the Distribution of Wages. *The Review of Economic Studies*, 80(1):145–173, January 2013.

Christian Dustmann and Arthur Van Soest. Language and the Earnings of Immigrants. *ILR Review*, 55(3):473–492, April 2002.

P.-A. Edin, P. Fredriksson, and O. Aslund. Ethnic Enclaves and the Economic Success of Immigrants–Evidence from a Natural Experiment. *The Quarterly Journal of Economics*, 118 (1):329–357, February 2003.

Mattias Engdahl, Sébastien Willis, and Olof Åslund. Professional networks and the labour market assimilation of immigrants. Technical Report 2024:9, IFAU - Institute for Evaluation of Labour Market and Education Policy, March 2024.

Katherine Eriksson. Ethnic enclaves and immigrant outcomes: Norwegian immigrants during the Age of Mass Migration. *European Review of Economic History*, 24(3):427–446, August 2020.

Ana Ferrer and W. Craig Riddell. Education, credentials, and immigrant earnings. *Canadian Journal of Economics/Revue canadienne d'économique*, 41(1):186–216, 2008. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2966.2008.00460.x.

Joseph P Ferrie. The entry into the u.s. labor market of antebellum european immigrants, 1840–1860, 1997.

Rachel M. Friedberg. You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics*, 18(2):221–251, April 2000.

Arkadev Ghosh, Sam Il Myoung Hwang, and Munir Squires. Links and Legibility: Making Sense of Historical U.S. Census Automated Linking Methods. *Journal of Business & Economic Statistics*, 42(2):579–590, April 2024.

Claudia Goldin and Lawrence F. Katz. Mass Secondary Schooling and the State: The Role of State Compulsion in the High School Movement. In Dora L. Costa and Naomi R. Lamoreaux, editors, *Understanding Long-Run Economic Growth: Geography, Institutions, and the Knowledge Economy*. University of Chicago Press, October 2011. ISBN 978-0-226-11634-1.

Javier P. Grossutti. From Guild Artisans to Entrepreneurs: The Long Path of Italian Marble Mosaic and Terrazzo Craftsmen (16th c. Venice – 20th c. New York City). *International Labor and Working-Class History*, 100:60–86, October 2021.

Daniel Valen Grytten and Bjørn L Basberg. Historical Index of Human Development on Scandinavian Countries 1820-2020.

Ola Grytten. The economic history of norway. *EH.Net Encyclopedia*, edited by Robert Whaples, March 2008. URL https://eh.net/encyclopedia/the-economic-history-of-norway/.

Blake H. Heller and Kirsten Slungaard Mumma. Immigrant integration in the united states: The role of adult english language training. *American Economic Journal: Economic Policy*, 15:407–437, 2023.

Lars Holden, Svetlana Boudko, and Bjørn Fjellvoll. Historisk befolkningsregister som et autoritetsregister for personer og verktøy for lokalhistorisk forskning. Working paper, 2025.

Susumu Imai, Derek Stacey, and Casey Warman. From engineer to taxi driver? language proficiency and the occupational skills of immigrants. *The Canadian Journal of Economics*, 52:914–953, 2019.

Beth F. Ingram and George R. Neumann. The returns to skill. *Labour Economics*, 13:35–59, 2 2006.

Robert J. LaLonde and Robert H. Topel. The Assimilation of Immigrants in the U. S. Labor Market. In *Immigration and the Work Force: Economic Consequences for the United States and Source Areas*, pages 67–92. January 1992.

Joan Lane. *Apprenticeship In England, 1600-1914*. Routledge, London, November 2005. ISBN 978-0-203-99262-3.

Marco H. D. van Leeuwen, Ineke Maas, and Andrew Miles. *HISCO: Historical International Standard Classification of Occupations*. Leuven University Press, 2002. ISBN 978-90-5867-196-7. Google-Books-ID: EMPtAAAAIAAJ.

Fan Li, Kari Lock Morgan, and Alan M. Zaslavsky. Balancing Covariates via Propensity Score Weighting. *Journal of the American Statistical Association*, 113(521):390–400, January 2018.

D. P. Lindstrom and D. S. Massey. Selective Emigration, Cohort Quality, and Models of Immigrant Assimilation. *Social Science Research*, 23(4):315–349, December 1994.

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

Aaditya Mattoo, Ileana Cristina Neagu, and Çağlar Özden. Brain waste? Educated immigrants in the US labor market. *Journal of Development Economics*, 87(2):255–269, October 2008.

Magne Mogstad. The Human Capital Approach to Intergenerational Mobility. *Journal of Political Economy*, 125(6):1862–1868, December 2017.

Petra Moser, Alessandra Voena, and Fabian Waldinger. German Jewish Émigrés and US Invention. *American Economic Review*, 104(10):3222–3255, October 2014.

Gianmarco I. P. Ottaviano and Giovanni Peri. Rethinking the Effect of Immigration on Wages. *Journal of the European Economic Association*, 10(1):152–197, February 2012.

Giovanni Peri, Kevin Shih, and Chad Sparber. STEM Workers, H-1B Visas, and Productivity in US Cities. *Journal of Labor Economics*, 33(S1):S225–S255, July 2015.

Samuel H. Preston and Michael R. Haines. *Fatal Years: Child Mortality in Late Nineteenth-Century America*. Princeton University Press, July 2014. ISBN 978-1-4008-6189-7.

Joseph Price, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley. Combining family history and machine learning to link historical records: The Census Tree data set. *Explorations in Economic History*, 80:101391, April 2021.

Deborah Rho and Seth Sanders. Immigrant Earnings Assimilation in the United States: A Panel Analysis. *Journal of Labor Economics*, 39(1):37–78, January 2021.

Greg Ridgeway, Stephanie Ann Kovalchik, Beth Ann Griffin, and Mohammed U. Kabeto. Propensity Score Analysis with Survey Weighted Data. *Journal of causal inference*, 3(2): 237–249, September 2015.

Francisco L. Rivera-Batiz. Undocumented workers in the labor market: An analysis of the earnings of legal and illegal Mexican immigrants in the United States. *Journal of Population Economics*, 12(1):91–116, February 1999.

Steven Ruggles, Matt A. Nelson, Matthew Sobek, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Evan Roberts, and J. Robert Warren. IPUMS ancestry full count data: Version 4.0 [dataset]. Minneapolis, MN: IPUMS, 2024.

Steven Ruggles, Lara Cleveland, Rodrigo Lovaton, Sula Sarkar, Matthew Sobek, Derek Burk, Dan Ehrlich, Jane Lee, and Nate Merrill. Integrated public use microdata series, international: Version 7.6 [dataset]. minneapolis, mn: Ipums, 2025a.

Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rodgers, Jonathan Schroeder, and Kari C.W. Williams. IPUMS USA: Version 16.0 [dataset]. Minneapolis, MN: IPUMS, 2025b.

Michael A. Shields and Stephen Wheatley Price. The English language fluency and occupational success of ethnic minority immigrant men living in English metropolitan areas. *Journal of Population Economics*, 15(1):137–160, January 2002.

University of Tromsø. The digital archive (the national archive), norwegian historical data centre (university of tromsø) and the minnesota population center. national sample of the 1900 census of norway, version 2.0. tromsø, norway: University of tromsø, 2008.

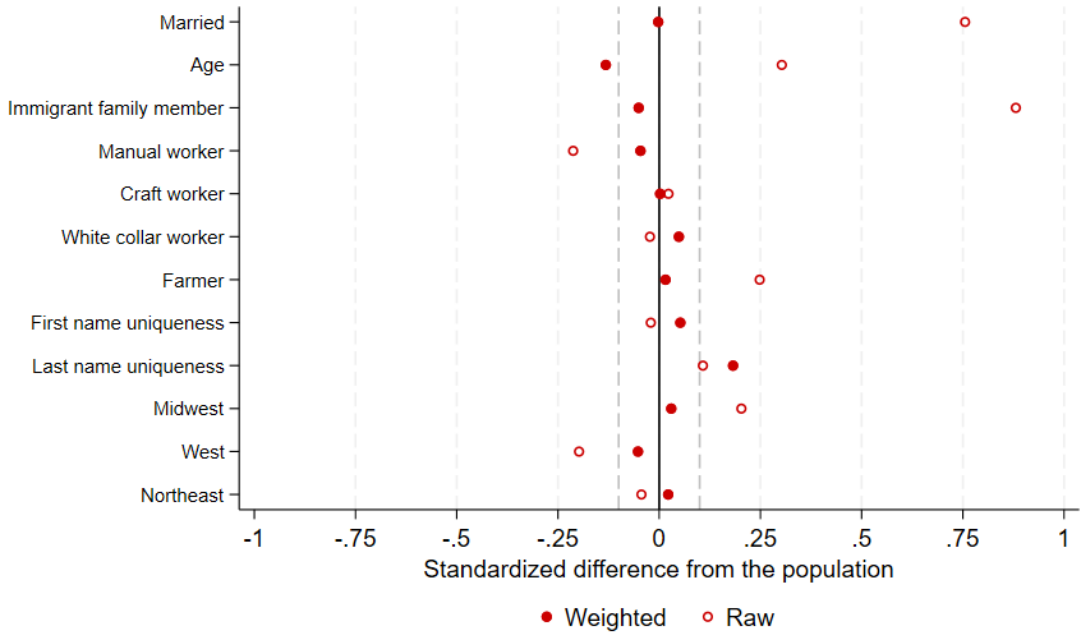
Andrés Villarreal and Christopher R. Tamborini. Immigrants' Economic Assimilation: Evidence from Longitudinal Earnings Records. *American Sociological Review*, 83(4):686–715, August 2018.

Yu Xie and Margaret Gough. Ethnic Enclaves and the Earnings of Immigrants. *Demography*, 48(4):1293–1315, August 2011.

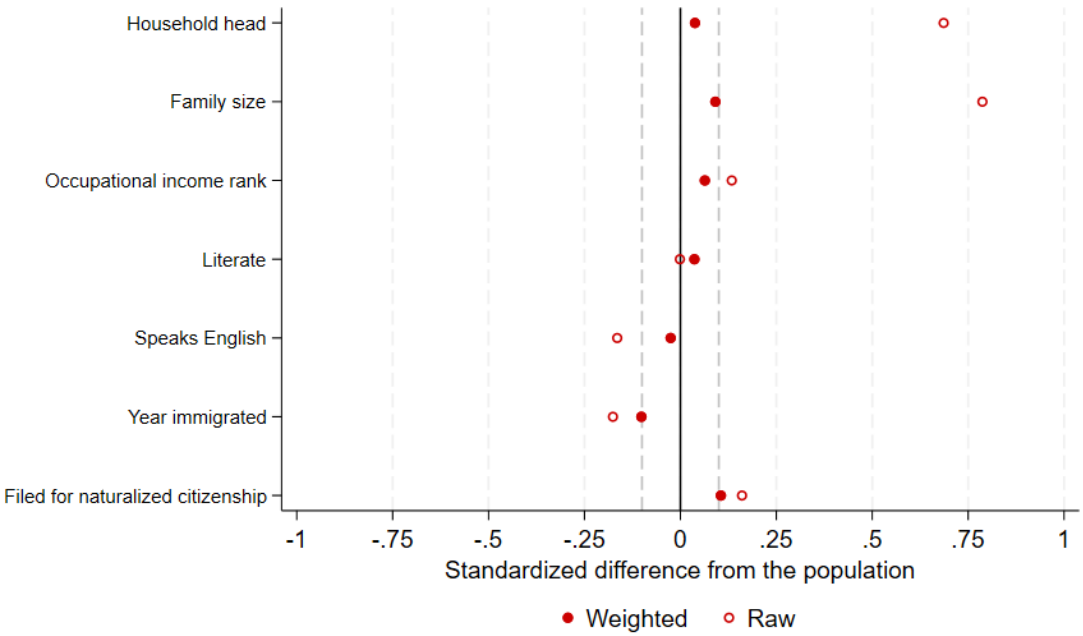
Cormac Ó Gráda, Tyler Anbinder, Dylan Connor, and Simone A. Wegge. The problem of false positives in automated census linking: Nineteenth-century New York's Irish immigrants as a case study. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 56 (4):240–259, October 2023.

Figure 1: Representativeness of international census links to Norway

(a) Weighting variables



(b) Validation variables



Notes: Standardized differences between the subset of immigrants in the 1910 U.S. Census who are linked to the 1900 Norway Census compared to the full population of recently-arrived Norwegian immigrants satisfying the sample restrictions. "Weighting variables" were used to fit a probit model of the probability of linkage to the Norway census, which generated the linkage weights. "Validation variables" were not used in the weighting process.

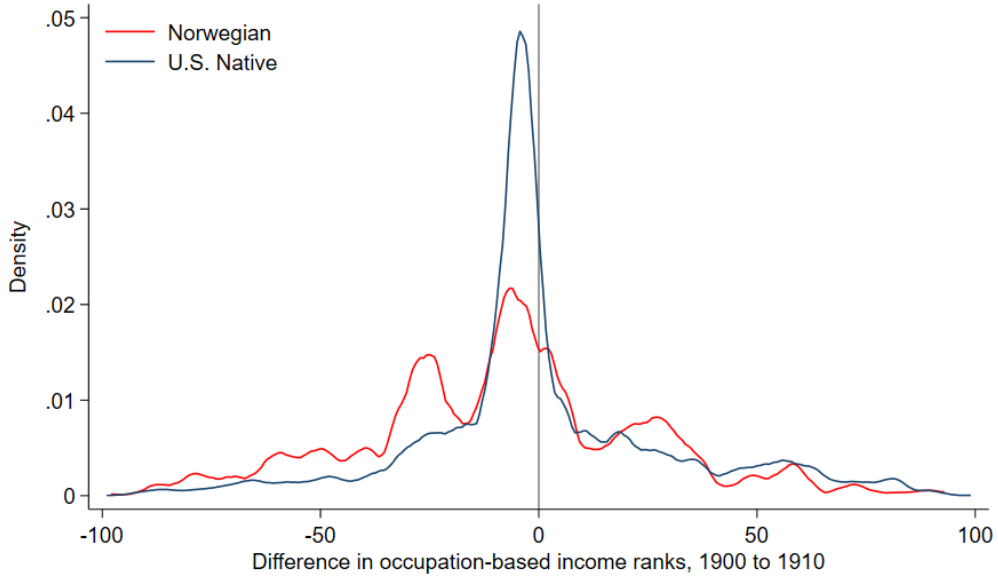
Table 1: Sample characteristics in the U.S. and birth countries

	1910 sample		1900 sample		
	Norwegian	U.S. Native	English Canadian	French Canadian	U.S. Native
Characteristics in U.S.					
Rank of occupational income	38.066	51.152	46.363	31.490	47.445
	[28.720]	[30.580]	[31.754]	[29.393]	[31.680]
Literate	0.965	0.984	0.960	0.553	0.964
	[0.183]	[0.126]	[0.195]	[0.498]	[0.187]
Age	34.013	41.203	47.335	46.606	47.256
	[6.946]	[9.740]	[7.697]	[6.333]	[7.638]
Midwest	0.596	0.456	0.273	0.126	0.494
	[0.491]	[0.498]	[0.446]	[0.333]	[0.500]
English	0.868	0.996	0.952	0.658	0.952
	[0.338]	[0.060]	[0.215]	[0.475]	[0.213]
Year immigrated	1904.108		1894.635	1894.822	
	[2.355]		[2.899]	[3.043]	
Filed for citizenship	0.596		0.411	0.254	
	[0.491]		[0.492]	[0.436]	
Characteristics in birth country					
Rank of occupational income	46.096	50.049	46.634	40.560	46.425
	[27.979]	[30.723]	[26.740]	[22.235]	[30.326]
Manual worker	0.344	0.213	0.276	0.453	0.186
	[0.475]	[0.409]	[0.447]	[0.499]	[0.389]
Craft worker	0.374	0.294	0.345	0.378	0.256
	[0.483]	[0.456]	[0.476]	[0.486]	[0.436]
White collar worker	0.090	0.267	0.143	0.044	0.203
	[0.286]	[0.442]	[0.351]	[0.206]	[0.403]
Farmer	0.142	0.133	0.099	0.028	0.214
	[0.350]	[0.340]	[0.299]	[0.166]	[0.410]
Additional linkages					
Linked to later census	0.315	0.333	0.477	0.441	0.522
	[0.465]	[0.471]	[0.500]	[0.497]	[0.500]
Son linked to later census	0.182	0.245	0.359	0.475	0.281
	[0.386]	[0.430]	[0.480]	[0.500]	[0.450]
Observations	2,889	5,024,398	1,755	267	2,300,974

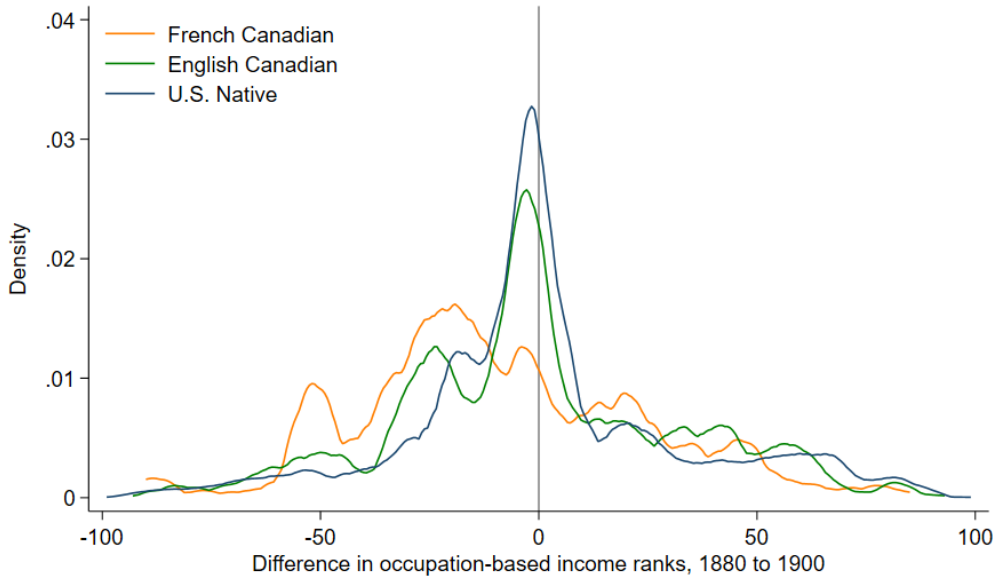
Notes: Samples include working-age immigrant men who arrived from Norway or Canada (separated by ethnic origin) within the last 10 years, as well as U.S.-born men. The table reports means and standard deviations of characteristics in the U.S. and birth country census for a linked sample that is reweighted for population representativeness. Individuals are assigned a percentile rank based on their place in the U.S. distribution of occupation-based income, by age and census year. For example, a 35-year-old carpenter in the 1900 Norway Census is assigned a rank of 64 because a typical carpenter in the U.S. earned more than 64% of 35-year-old men in 1900.

Figure 2: Distribution of rank-rank occupational income mobility between birth country and U.S. census

(a) Norwegian immigrants



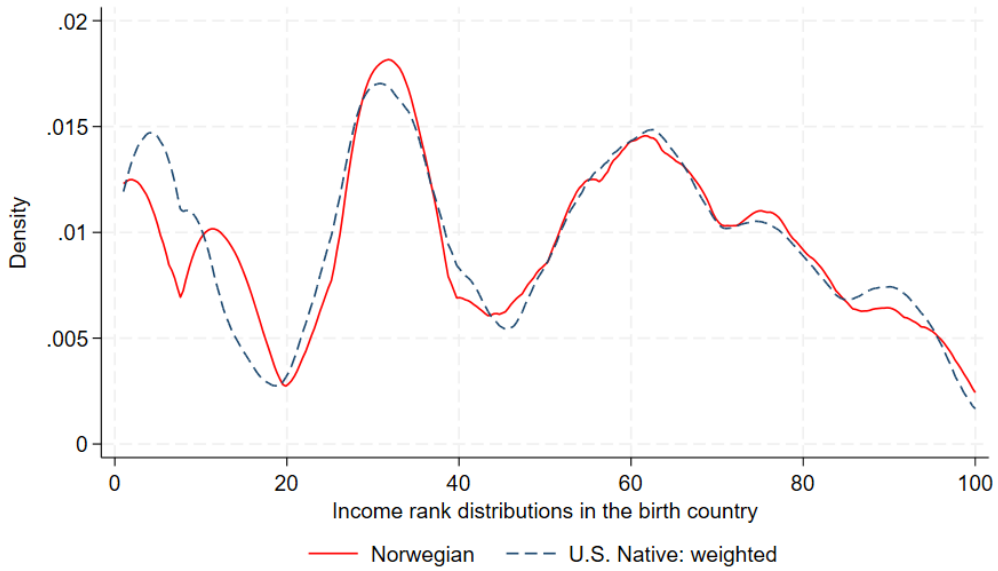
(b) Canadian immigrants



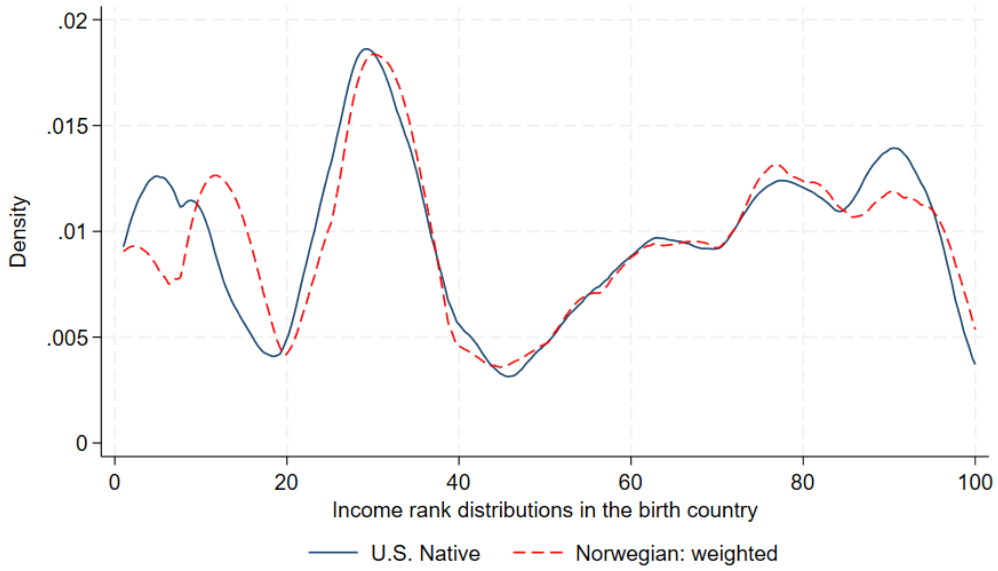
Notes: Differences in individual occupational income ranks between the earlier and later census year. Individuals are assigned a percentile rank based on their place in the U.S. distribution of occupation-based income, by age and census year. For example, a 35-year-old carpenter in the 1900 Norway Census is assigned a rank of 64 because a typical carpenter in the U.S. earned more than 64% of 35-year-old men in 1900.

Figure 3: Reweighted occupational income distributions in the birth country, Norway and U.S. Census in 1900

(a) Reweighted to the Norwegian distribution



(b) Reweighted to the U.S. native distribution



Notes: Weights generated from a probit model with splines, with linkage weights pre-applied to the data. Distributions of occupational income ranks among the U.S. population in 1900, by age.

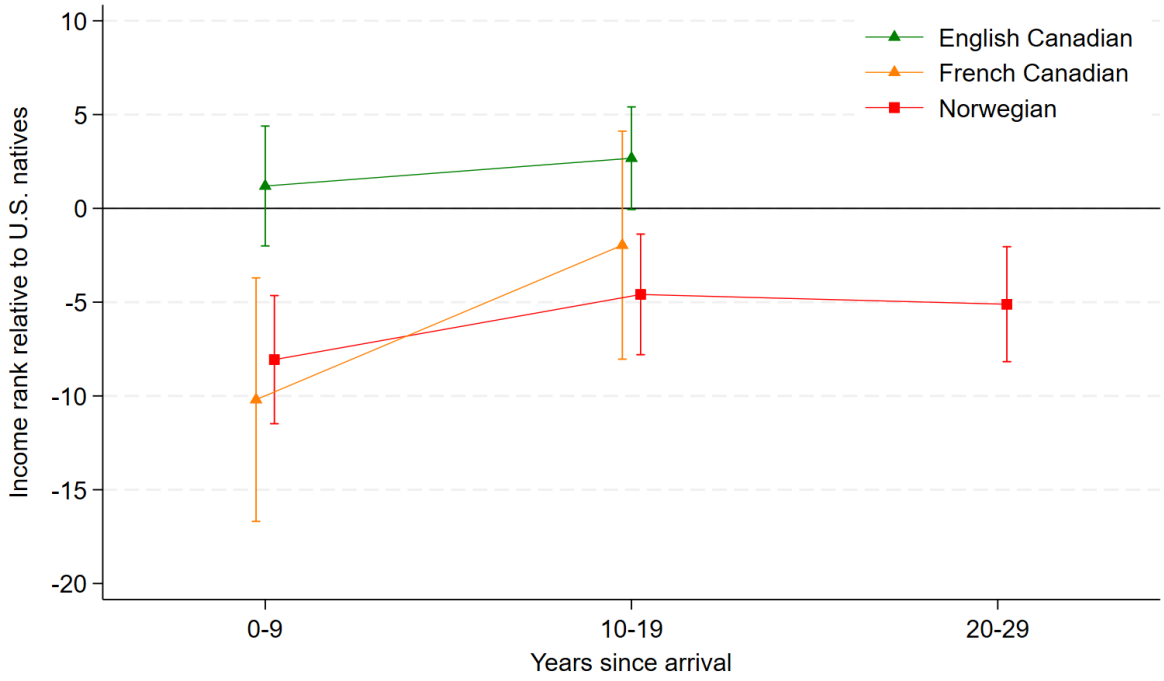
Table 2: Skill-adjusted occupational income gaps between immigrants and natives

	(1)	(2)	(3)	(4)
Panel A: Norwegian sample				
Immigrant	-13.086***	-10.446***	-10.937***	-10.655***
	(0.869)	(0.869)	(1.002)	(0.805)
Target of birth-country occ. rank weights		Norway	U.S.	
Birth-country occ. ventile rank FE				✓
Birth-country census linkage weights	✓	✓	✓	✓
Observations	4,686,991	4,686,991	4,686,991	4,686,991
Panel B: French Canadian sample				
Immigrant	-15.955***	-10.446***	-13.746***	-10.591***
	(2.132)	(2.132)	(2.811)	(2.112)
Target of birth-country occ. rank weights		Canada	U.S.	
Birth-country occ. ventile rank FE				✓
Birth-country census linkage weights	✓	✓	✓	✓
Observations	2,039,652	2,039,652	2,039,652	2,039,652
Panel C: English Canadian sample				
Immigrant	-1.082	-0.023	-0.848	-0.654
	(1.220)	(1.220)	(1.357)	(1.113)
Target of birth-country occ. rank weights		Canada	U.S.	
Birth-country occ. ventile rank FE				✓
Birth-country census linkage weights	✓	✓	✓	✓
Observations	2,040,973	2,040,973	2,040,973	2,040,973

$^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$

Notes: Comparing occupational income ranks between U.S. natives and immigrants within ten years of arrival. Samples are reweighted to balance on the birth-country distribution of occupational income ranks in either the U.S., Norway, or Canada, giving various types of treatment effects.

Figure 4: Trajectory of skill-adjusted occupational income gaps between immigrants and natives



Notes: OLS regression comparing occupational income of U.S. natives and immigrants up to 29 years after arrival. 95% confidence intervals are shown. Consistent panel sample observed in 1900 and 1910 for Canadian immigrants and 1910 through 1930 for Norwegian immigrants. Birth-country occupations are reweighted to the immigrant population, generating estimates of the average treatment effect for immigrants.

Table 3: Intergenerational mobility estimates, relative to father's occupational income rank in U.S. or the birth country

	(1)	(2)	(3)	(4)
Panel A: Norwegian sample				
Immigrant father	-2.282**	-2.142	-4.438***	-4.237***
	(1.114)	(1.303)	(1.114)	(1.251)
Target of U.S. occ. rank weights	Norwegian	U.S. Native		
Target of birth-country occ. rank weights			Norwegian	U.S. Native
Birth-country and U.S. census linkage weights	✓	✓	✓	✓
Difference from U.S.-weighted mobility estimate			-2.156*	-2.095
Observations	1,520,069	1,520,069	1,520,069	1,520,069
Panel B: French Canadian sample				
Immigrant father	-4.918*	-8.563***	-4.058	-6.549**
	(2.600)	(2.600)	(2.991)	(2.501)
Target of U.S. occ. rank weights	Canadian	U.S. Native		
Target of birth-country occ. rank weights			Canadian	U.S. Native
Birth-country and U.S. census linkage weights	✓	✓	✓	✓
Difference from U.S.-weighted mobility estimate			0.860	2.014
Observations	645,357	645,357	645,357	645,357
Panel C: English Canadian sample				
Immigrant father	1.280	2.202*	2.156*	1.992
	(1.163)	(1.159)	(1.163)	(1.291)
Target of U.S. occ. rank weights	Canadian	U.S. Native		
Target of birth-country occ. rank weights			Canadian	U.S. Native
Birth-country and U.S. census linkage weights	✓	✓	✓	✓
Difference from U.S.-weighted mobility estimate			0.876	-0.210
Observations	645,932	645,932	645,932	645,932

*$p < .1$, **$p < .05$, ***$p < .01$

Notes: Comparing occupational income ranks between sons of U.S. natives or immigrants, either in the U.S. within ten years of arrival in the birth country. Fathers are assigned a percentile rank based on their place in the U.S. distribution of occupation-based income, by age and census year. P-values for the differences between coefficients are computed using a standard normal estimator of the form $z = \frac{\beta_1 - \beta_2}{\sqrt{\beta_1^2 + \beta_2^2}}$.

Online Appendix for "Can Immigrants Transfer their Skills? Evidence from International Census Records"

Adrian Haws

A Record linking methodology **32**
A.1 Family Tree data . . . 32
A.2 Data pre-processing and candidate links . . . 33
A.3 Feature engineering . . . 34
A.4 Selecting unique links . . . 35
A.5 Representativeness . . . 37

B Additional tables and figures **41**
B.1 Norway census . . . 41
B.2 English-speaking French Canadians . . . 41
B.3 Balancing weights . . . 42

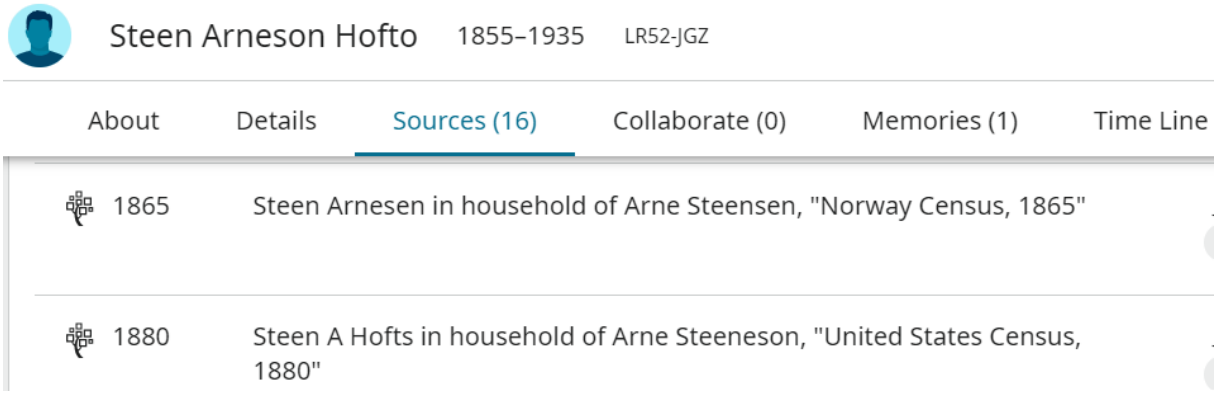
A Record linking methodology

Here I describe the methodology of my machine learning pipeline for linking Norwegian immigrants from the 1910 U.S. Census to the 1900 Norway Census. This approach will also be modified to link Canadian immigrants. I build on previous methods, but introduce several innovations. First, I develop methods to consider multiple possible surnames in the context where Norwegian immigrants often switch their last names upon arrival in the U.S. Second, I use information from two external datasets that provide information about emigrants and those who stayed in Norway. Third, I use a high-quality ground truth dataset of international census links to create a much more flexible algorithm than previous efforts at linking immigrants.

A.1 Family Tree data

Machine learning approaches have yielded substantial gains in match rates for historical record links, but these require high-quality training data. However, training data is costly to create, and in the case of Norwegian immigrants would be extremely difficult with external information. Consider the example shown in Figure A1. A public genealogy profile shows Steen Arneson Hofto, who was born in 1855. Census records are attached for 1865 and 1880, establishing a "true" link. Yet it is not at all clear why "Steen Arnesen" should be considered the same person as "Steen A. Hofts." Both of these records have been attached by Steen's descendants as part of a collaborative Family Tree on FamilySearch.org.

Figure A1: Example of a deceased person's profile on FamilySearch.org



The Family Tree provides many such instances of "true" links that have been created by volunteer genealogists who have access to vital records and collaborate on a shared tree. These links inform the machine learning algorithm of cases where a farm name (such as "Hoft") is adopted as a surname or how closely a first name should match across records. These links from the Family Tree also inform the pre-processing of census data.

A.2 Data pre-processing and candidate links

The Norway census data contain several fields that have some probability of becoming an immigrant's surname in the U.S. This is due to the unique historical context where many Norwegians did not have fixed surnames until after 1923. Instead, they used patronyms that changed with each generation (e.g. "Eriksen" for the children of Erik). To further distinguish between common names, individuals would sometimes append their farm name, as in the case of Steen Arnesen Hofts. Because neither the patronym or the farm name is a surname, immigrants could choose what to adopt as their surname. In some cases, they even adopted their father's patronym. Those who did have surnames in Norway lived in cities or were highly educated. Informed by this context, I separate and clean information on patronyms, names that could be surnames or place names, and the name of the current farm of residence.[18]

I then use a very flexible blocking strategy that attempts to limit the number of candidate links while retaining a high fraction of the true links. I require that candidate links have a difference in birth years of no more than one, and that they match on the NYSIIS phonetic encoding of the first name, after converting each first name to the most likely Americanized name (e.g. Jan, Johan, and Johannes are converted to John). To accomplish this, I create a name Americanization crosswalk from the Family Tree links. Finally, I require that the surname in the U.S. match on the first initial or at least one bigram with any of the patronym, surname, or farm name information from a Norway census record. To create the full set of training pairs, I label "false" pairs if a true link exists but does not match.

[18]Many Norwegians who lived on land with a historical farm name were not actually involved in farming.

A.3 Feature engineering

A key advantage of the machine learning process is that it support high dimensionality. I create a total of 36 features, many of which are a function of last names. For each comparison of last name types (patronym, surname, or farm name), I create interactions with key characteristics in the Norway census such as rural status. I also create features based on the similarity of information for the two family members that have the most in common across the censuses.

Some of the most important information comes from historical records that are external to the census. The first of these is Norwegian emigrant lists. Every person who left Norway during the period of my study should be included, but a large fraction do not include a birth year. Among those with quality information, I use a simple linking algorithm to identify individuals in the census before migration. A link to the emigrant lists is a strong indicator of having migrated, but these also include information on the year of immigration and the intended residence in the destination. The second external source is a set of links across the 1900 and 1910 Norway Census from the Norwegian Historical Population Register (Holden et al., 2025), which indicate that a person likely stayed in Norway.

Table A1 shows the number of features and their collective importance, among several categories. The most important features based on the focal record are last names and features based on external records. Features based on family members also have a high importance because they help establish uniqueness. Although given names and birth year appear to have low importance, that is because all of the candidate links were required to have similar given names and birth years as part of the blocking strategy.

Table A1: Feature importances in the XGBoost model

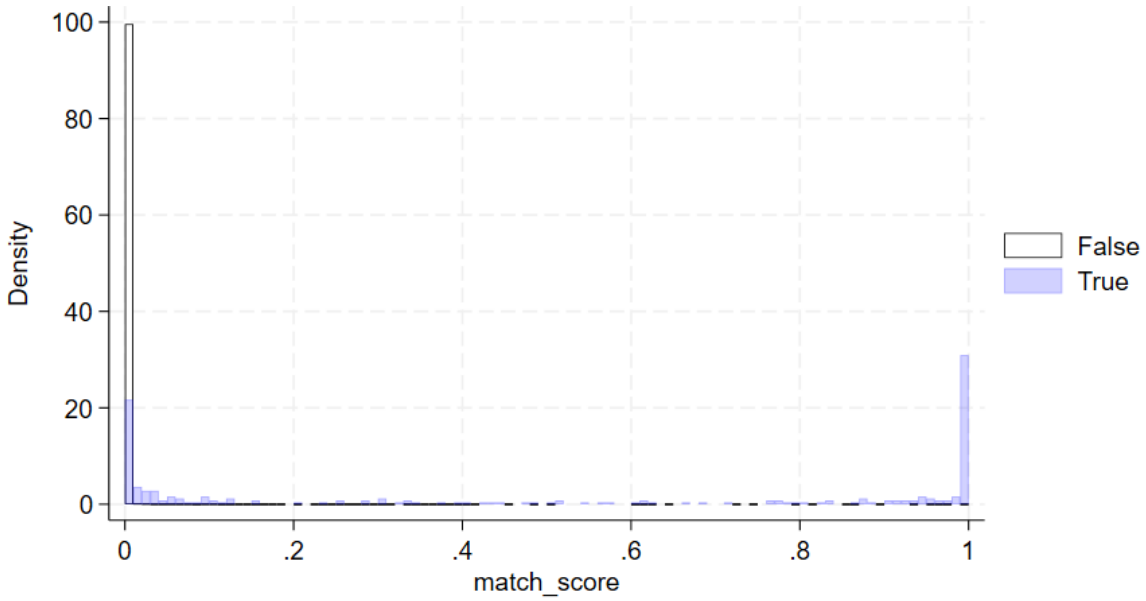
Category	Number of features	Importance
Self		
Last names	10	0.189
Linked to the 1910 Norway Census	1	0.120
Emigration records	3	0.120
Given names	3	0.042
Birth year	1	0.030
Family members		
Last names	10	0.152
Linked to the 1910 Norway Census	2	0.189
Emigration records	2	0.008
Given names	2	0.069
Birth year	2	0.082

Notes: The importance measure is calculated as the average increase in accuracy across nodes of the decision tree which use the feature. This is the "gain" measure of feature importance calculated by the XGBoost algorithm.

A.4 Selecting unique links

To generate predictions of whether each candidate link is true, I use the XGBoost algorithm because it performs well in historical record linking (Price et al., 2021). This results in continuous match scores between 0 and 1 for each candidate link. I then perform a validation exercise to show that the model successfully distinguishes between true and false links. Figure A2 takes a set of training links that were not used in training the model (the "test set") and shows the frequency of match scores among true and false links. Nearly all of the false pairs are given a near-zero score and the true pairs are heavily clustered near a score of 1.

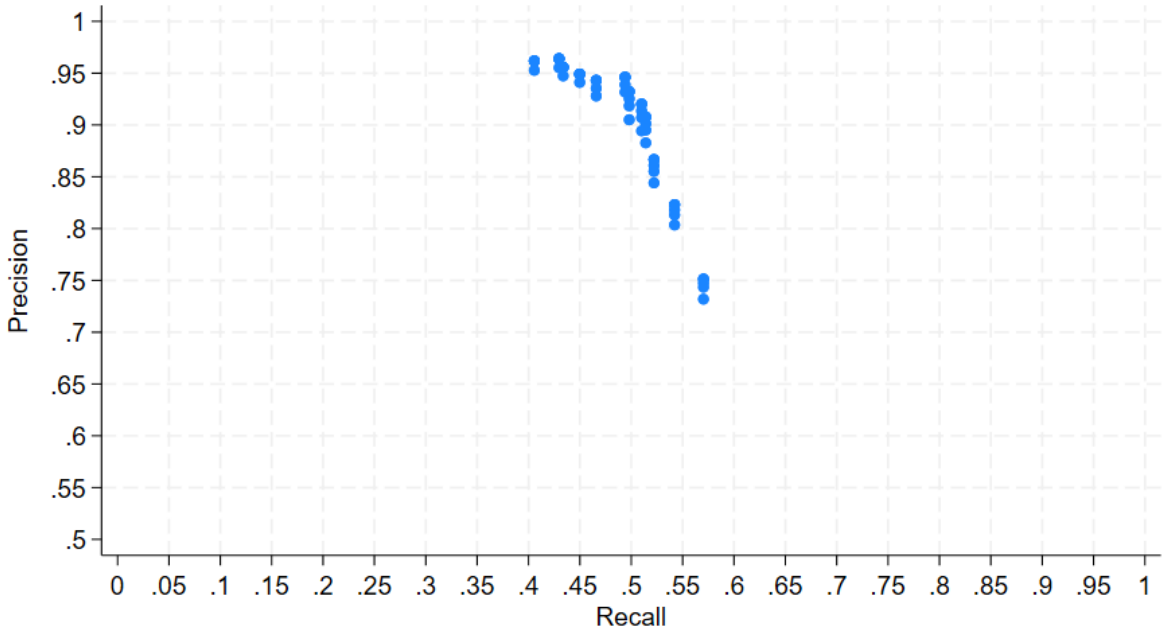
Figure A2: Distribution of match scores in the test set



Notes: Distribution of match scores in the test set from an XGBoost model with 36 features that was trained on 933 true pairs.

There is a trade-off in selecting the required match score. If all selected links must have a match score above .95, accuracy of the links would be very high but many true links would be missed. Figure A3 shows the tradeoff between precision (accuracy) and recall (match rate) among the test set, across many possible choices of cutoffs for match scores. Two parameters are varied to create the plot: the minimum match score and the minimum ratio between best and second-best match score. I then choose a set of values that results in a very high out-of-sample accuracy (93.4%) to ensure the dataset is high quality and the results are not driven by false links, and drop all non-unique links that remain. Combining the Family Tree and unique machine learning links produce a total of 3,026 links and a match rate of 11.2%.

Figure A3: Precision and recall of various match score cutoffs in the test set



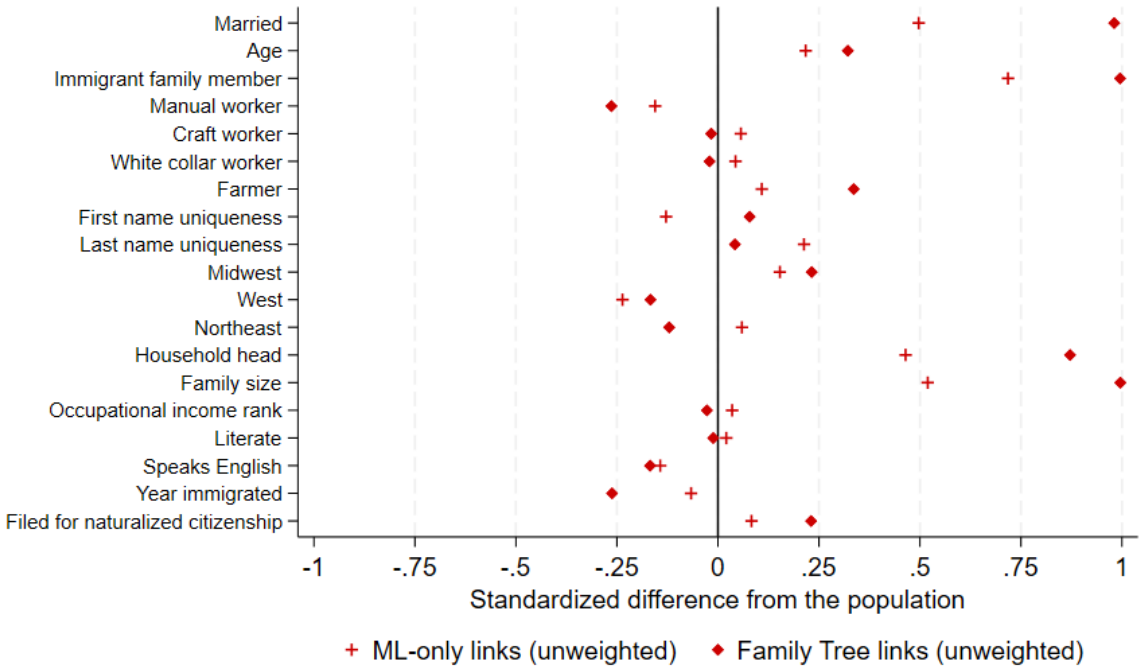
Notes: Values of precision and recall are plotted at many possible values of the minimum match score and minimum ratio between best and second-best match score, using the test set.

A.5 Representativeness

Although the Family Tree links are not representative of the full sample of recent Norwegian immigrants, they include support for different types of true links. Figure A4 shows that the machine learning links are much more representative than the Family Tree links for most variables. To account for the selection into linkage, I reweight each link (including the Family Tree and machine learning links) by the inverse probability of being linked. This process is described more in the main part of the paper. Here I show that there is common support in the probability of linkage (Figure A5 across the linked and unlinked samples.

Figure 1 showed that reweighting the Norway census links for representativeness also improves the representativeness across variables not included in the propensity score calculation. The results are not as promising for linked Canadian immigrants, as shown in Figure A6. However, the current paper does not include machine learning links for Canada, and the success of reweighting is likely to increase with the number of available links.

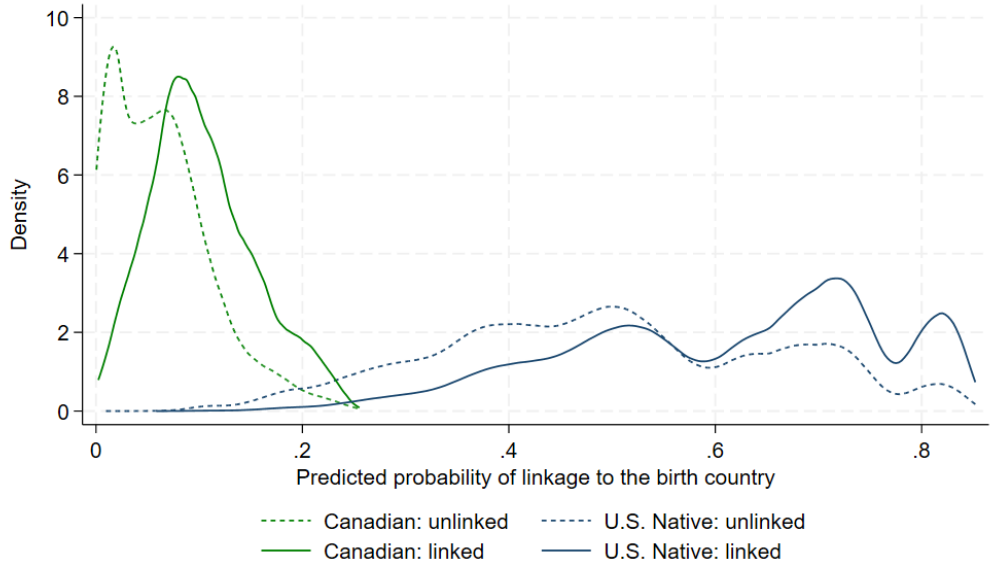
Figure A4: Representativeness of the Family Tree and machine learning links



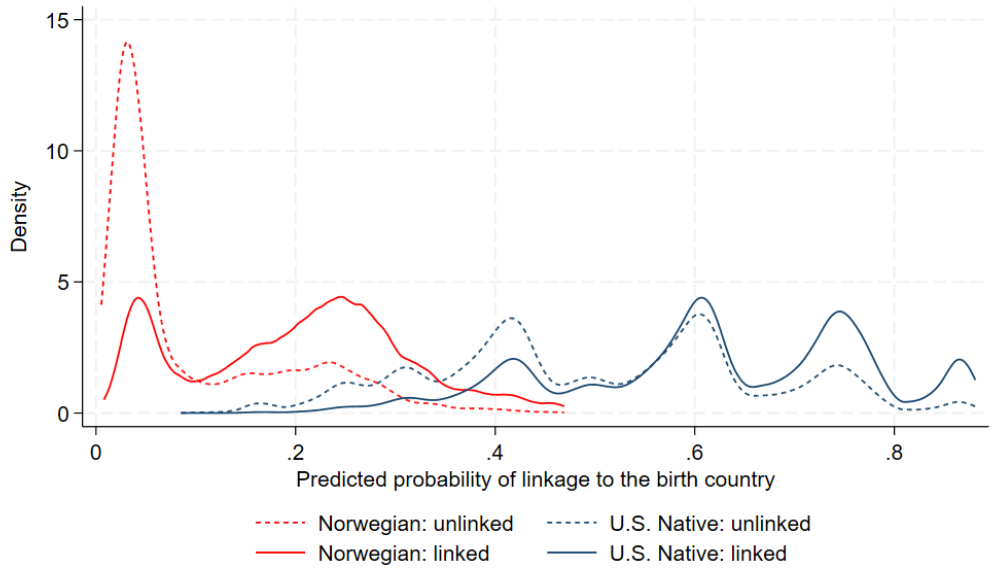
Notes: Standardized differences between the subset of immigrants in the 1910 U.S. Census who are linked to the 1900 Norway Census compared to the full population of recently-arrived Norwegian immigrants satisfying the sample restrictions. The Family Tree links were used to train a machine learning model.

Figure A5: Probability of linkage to an earlier census in the birth country

(a) Linked to the Norway or U.S. Census in 1900



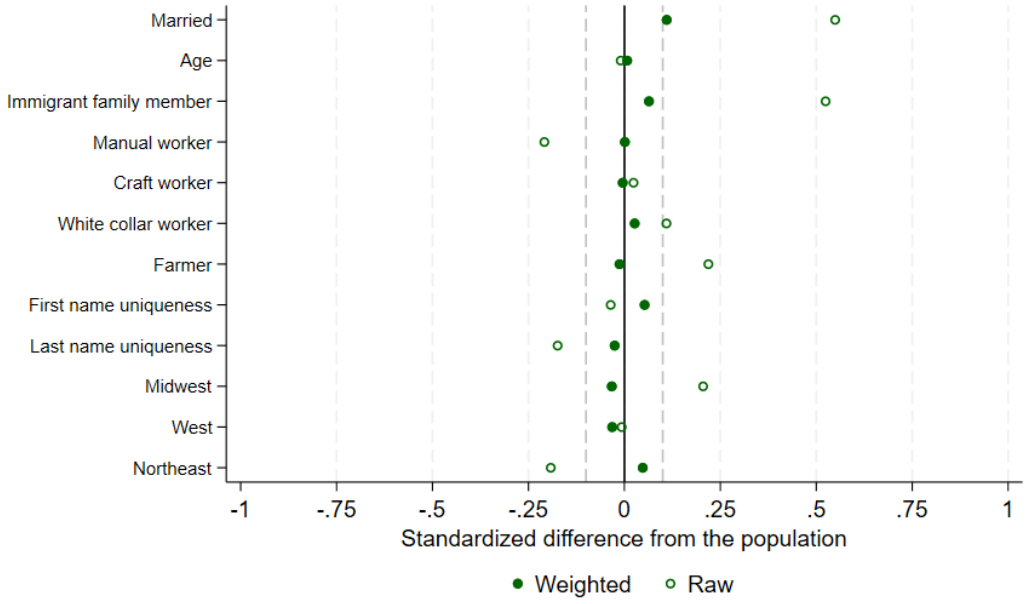
(b) Linked to the Canada or U.S. Census in 1880/1881



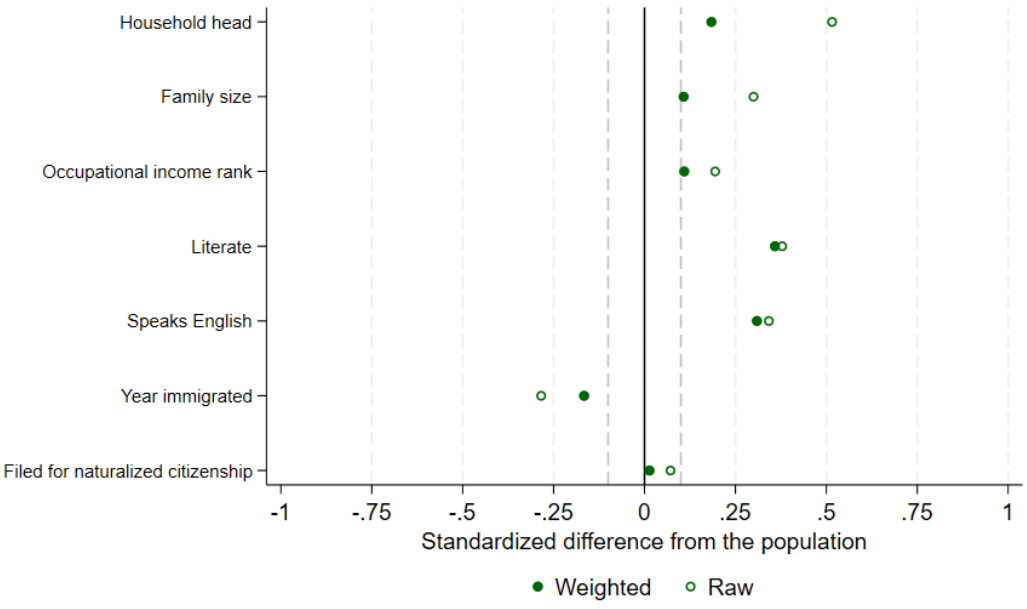
Notes: Probability of linkage from a probit model estimated separately for immigrants and U.S. natives.

Figure A6: Representativeness of international census links to Canada

(a) Weighting variables



(b) Validation variables

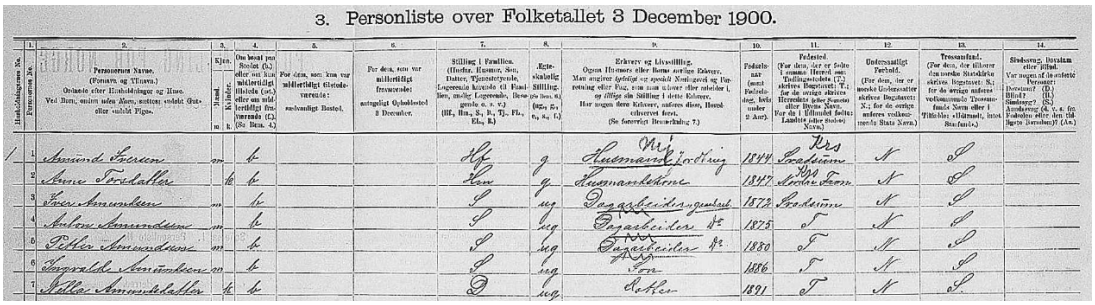


Notes: Standardized differences between the subset of immigrants in the 1900 U.S. Census who are linked to the 1881 Canada Census compared to the full population of recently-arrived Canadian immigrants satisfying the sample restrictions. "Weighting variables" were used to fit a probit model of the probability of linkage to the Canada census, which generated the linkage weights. "Validation variables" were not used in the weighting process.

B Additional tables and figures

B.1 Norway census

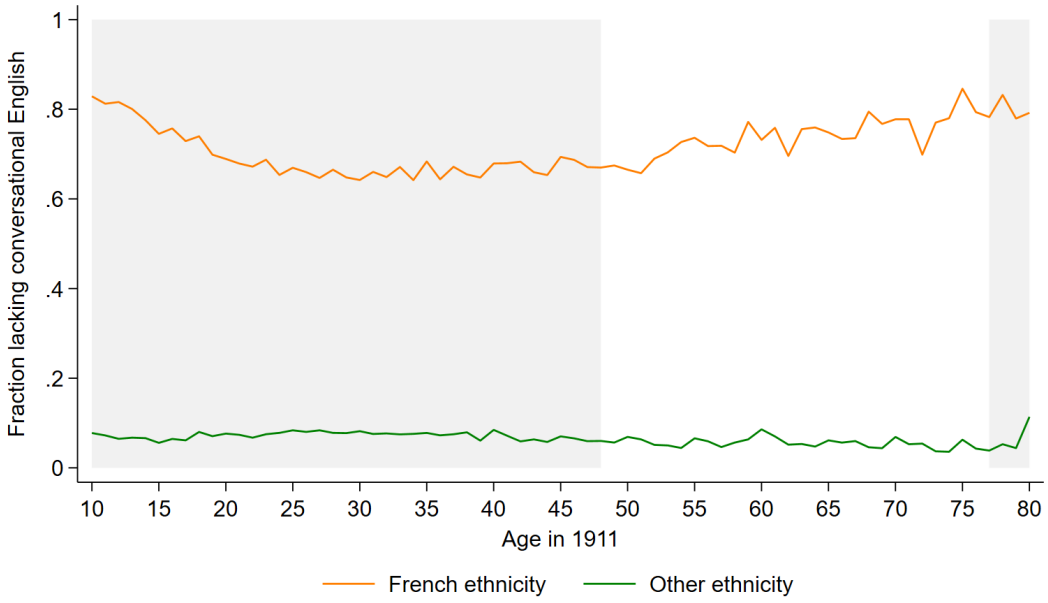
Figure B1: Image of the 1900 Norway Census



Notes: An image of a census sheet from the 1900 Norway Census, focusing on a nuclear family with two parents and five children. Column 9 includes occupations.

B.2 English-speaking French Canadians

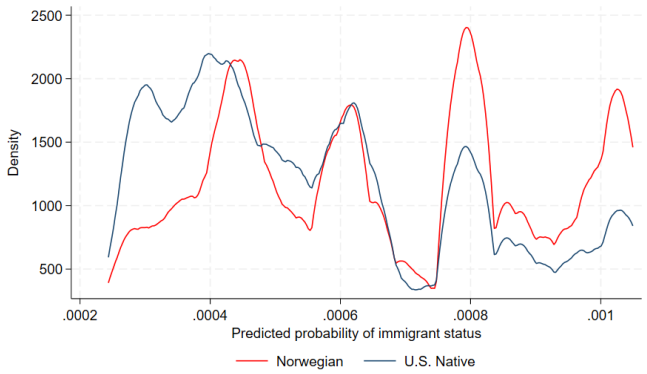
Figure B2: Fraction of French Canadians lacking conversational English, 1911 Canada Census



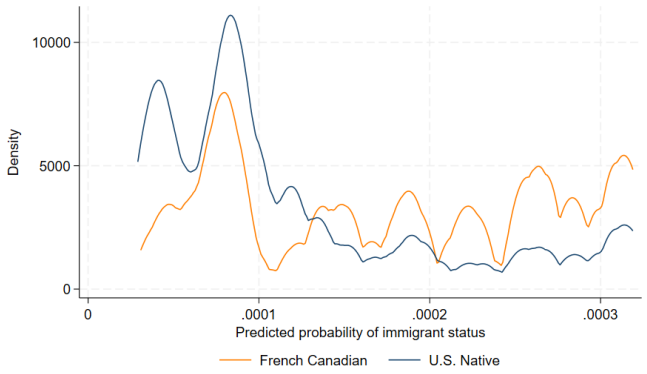
B.3 Balancing weights

Figure B3: Propensity score functions for balancing weights

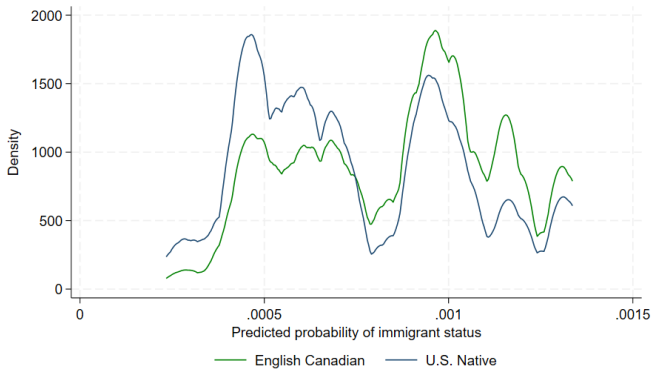
(a) Norwegian



(b) French Canadian



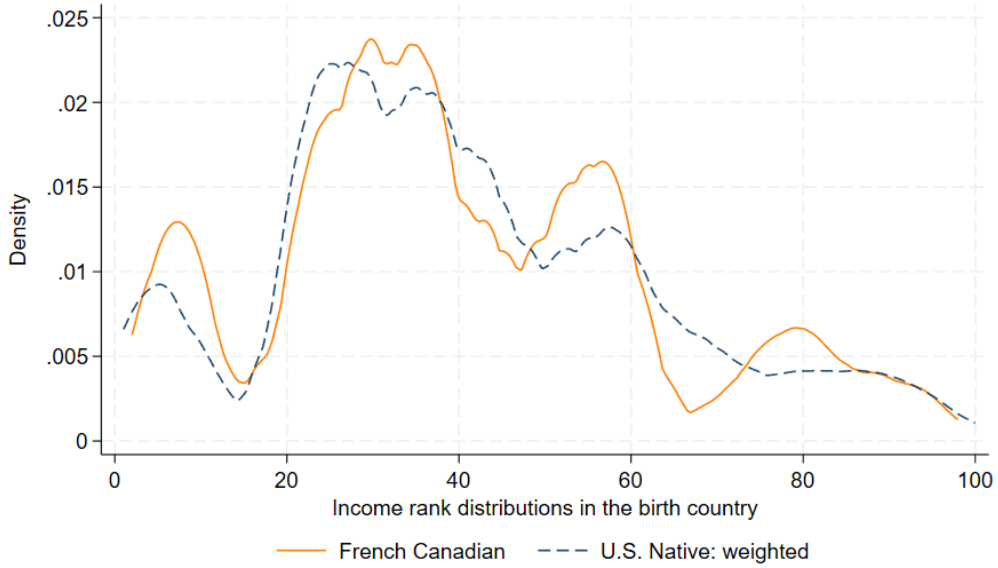
(c) English Canadian



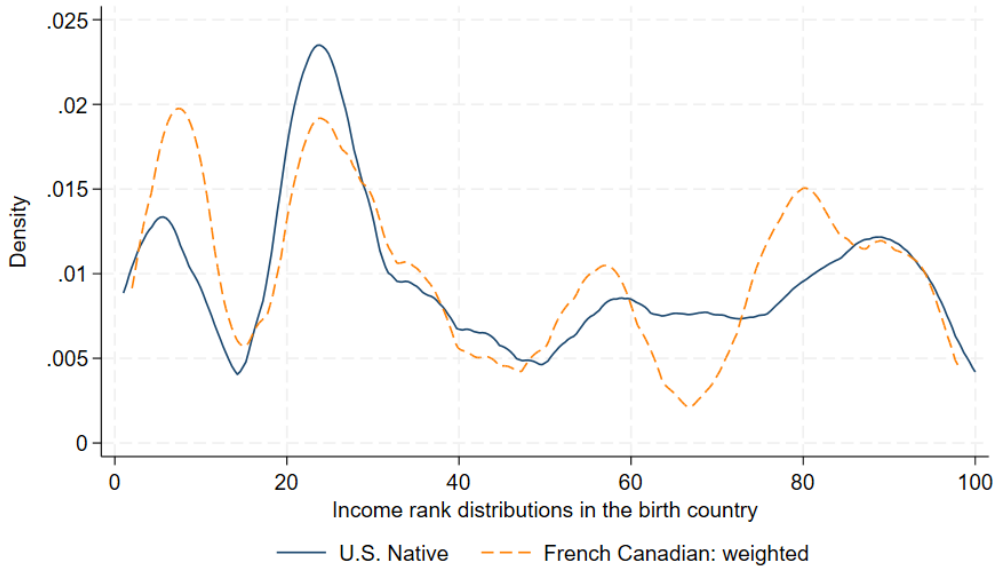
Notes: Probability of immigrant status from a probit model with splines for U.S. income ranks based on birth-country occupations, with linkage weights pre-applied to the data.

Figure B4: Reweighted occupational income distributions in the birth country, French Canadians and U.S. natives in 1880/1881

(a) Reweighted to the French Canadian distribution



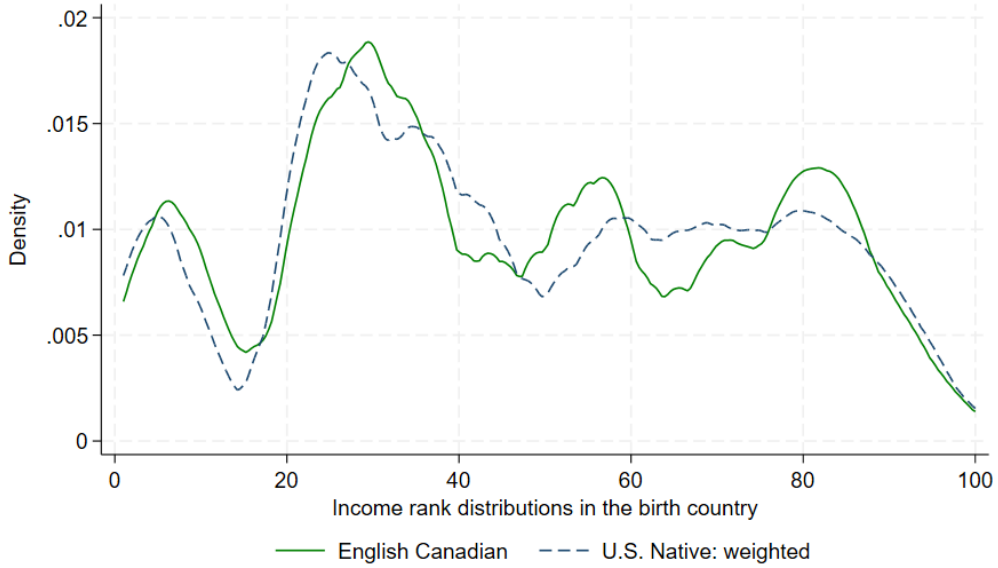
(b) Reweighted to the U.S. native distribution



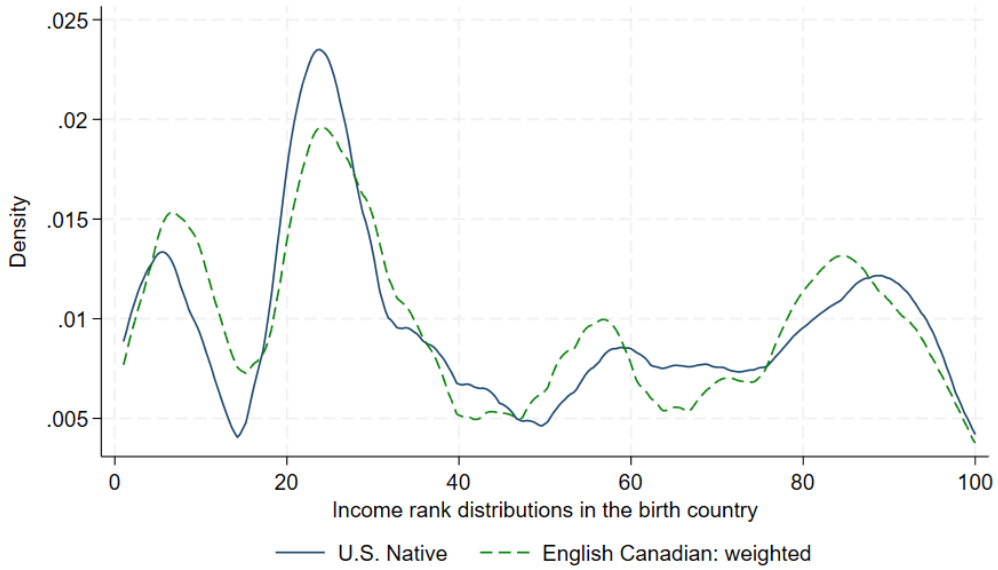
Notes: Weights generated from a probit model with splines, with linkage weights pre-applied to the data. Distributions of occupational income ranks among the U.S. population in 1880, by age.

Figure B5: Reweighted occupational income distributions in the birth country, English Canadians and U.S. natives in 1880/1881

(a) Reweighted to the English Canadian distribution



(b) Reweighted to the U.S. native distribution



Notes: Weights generated from a probit model with splines, with linkage weights pre-applied to the data. Distributions of occupational income ranks among the U.S. population in 1880, by age.