NBER WORKING PAPER SERIES

A NATION OF IMMIGRANTS: ASSIMILATION AND ECONOMIC OUTCOMES IN THE AGE OF MASS MIGRATION

Ran Abramitzky
Leah Platt Boustan
Katherine Eriksson

Working Paper 18011
http://www.nber.org/papers/w18011

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2012

We are grateful for the access to Census manuscripts provided by Ancestry.com. We acknowledge financial support from the National Science Foundation (No. SES-0720901), the California Center for Population Research and UCLA's Center for Economic History. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration
Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson
NBER Working Paper No. 18011
April 2012
JEL No. F22,J61,N31

ABSTRACT

During the Age of Mass Migration, the US maintained open borders and absorbed 30 million European immigrants. Using cross-sectional data, prior work on this era finds that immigrants held lower-paid occupations than natives upon first arrival but experienced rapid convergence. In newly-assembled panel data following immigrants over time, the initial immigrant earnings penalty disappears almost entirely, and immigrants experience occupational upgrading at the same rate as natives. Cross-sectional patterns are driven by declines over time in arrival cohort quality and the departure of negatively-selected return migrants. We show that these findings vary substantially across sending countries and explore potential mechanisms.

Ran Abramitzky
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
and NBER
ranabr@stanford.edu

Leah Platt Boustan
Department of Economics
8283 Bunche Hall
UCLA
Los Angeles, CA 90095-1477
and NBER
lboustan@econ.ucla.edu

Katherine Eriksson
Department of Economics
8283 Bunche Hall
UCLA
Los Angeles, CA 90095-1477
kath722@ucla.edu

I. Introduction

This paper examines the performance of European migrants in the US labor market in the early twentieth century, both upon first arrival and over their first few decades in the US. We focus on the Age of Mass Migration (1850-1913) for two reasons. First, this era was one of the largest migration episodes in modern history. European countries lost up to a third of their population through emigration. The US absorbed 30 million immigrants in this era, and by 1910 22 percent of the US labor force was foreign born (compared with 17 percent today). Migration in this period was thus large enough to affect the labor supply and economic development on both sides of the Atlantic. Second, the US maintained an open border policy for European migrants in this period and had yet to develop a comprehensive welfare state, allowing the study of immigrants' labor market performance in the absence of immigrant selection policies or government support.

We use newly-constructed historical panel data to address two sets of questions. First, how did European immigrants perform in the American labor market upon first arrival? Did immigrants' (and their children's) performance converge to that of US natives? Second, were migrants who returned to their home countries drawn from the high end or the low end of the skill distribution, relative to the migrant pool? Understanding the selection of temporary versus permanent migrants is important in this context because over 25 percent of migrants returned to Europe (Gould, 1980; Bandiera, Rasul and Viarengo, 2010). Moreover, the direction of selection of return migrants is conceptually ambiguous. Return migrants could be negatively selected if, for example, migrants who were not successful in the US returned home. Many migrants in this era employed a deliberate strategy of temporary migration to the New World (Piore, 1980; Wyman, 1996). These temporary migrants could be negatively selected if they work in low-paid

occupations before returning home or could be positively selected if more productive migrants reached their "target savings" goal faster (Galor and Stark, 1991).

There is an extensive literature in economic history on immigrants' labor market assimilation in the early twentieth century. Yet addressing such fundamental questions about this era of mass migration has remained a challenge because of a lack of historical panel data. Inferring immigrants' assimilation from cross-sectional data introduces well-known biases that arise because one cannot follow the same immigrants over time. First, comparing migrants who just arrived with those who arrived years earlier in a single cross-section does not allow the researcher to distinguish differences in the quality of immigrant cohorts by arrival year from the assimilation of individual immigrant cohorts over time (Borjas, 1985). For example, if immigrants who arrived in 1900 were more skilled than their counterparts who arrived in 1910, any apparent relationship between earnings and years spent in the US may instead be due to differences in skills across cohorts rather than to the assimilation of any particular migrant cohort. Second, even in repeated cross-sections, when cohorts of immigrants are compared across Censuses, inferences on migrant assimilation can be biased by the process of return migration (Duleep and Dowhan 2002, Lubotsky 2007). For example, if less-skilled migrants were more likely to return to their countries of origin, then apparent increases in skills across years may be driven solely by compositional changes, and not in fact indicate assimilation of migrants.

To address these challenges, we construct a large panel dataset of 24,000 native born Americans and immigrants from 16 sending countries by matching men by name, age and place of birth between the 1900, 1910 and 1920 US Censuses. Assembling such panel data is possible because US Census policy makes complete individual records (including names) publicly

available after 72 years. In particular, we match immigrants and US natives from the 1900 Census manuscripts to the 1910 and 1920 Census manuscripts using the genealogy website Ancestry.com. This new panel dataset allows us to study how the same migrants (those who remained in the US long term) performed in the US over time without encountering the problems inherent in using cross-sectional data to measure assimilation. Moreover, by contrasting the assimilation patterns in the repeated cross-section and panel data, we can infer the nature of selection of return migrants relative to migrants who remained in the US long term. In particular, differences in the assimilation profile in the repeated cross-sections and the panel are due to selective attrition, which likely reflects selective return migration.[1] Indeed, when information on return migrants is observed, our indirect approach of inferring selection of return migrants is shown to be consistent with a direct approach of comparing return migrants with migrants who remained in the US.

Consistent with the existing literature for this time period, we find that in each cross-section it appears as if immigrants initially held lower-paid occupations than did natives but converged upon natives over time (Blau, 1980; Hatton, 1997; Minns, 2000). Controlling for year of arrival in repeated cross-sections halves the initial earnings gap between immigrants and natives. In the panel data, immigrants' initial penalty disappears almost entirely (completely disappearing in some specifications). We conclude that the apparent convergence in a single cross-section is driven by a decline in the quality of immigrant cohorts over time and the departure of negatively-selected return migrants. We focus on occupation-based earnings because individuals' actual earnings were not recorded in population Censuses before the mid-twentieth century. We thus match each individual's recorded occupation to the median earnings

[1] Selective attrition could also be driven by selective mortality or selective name changes. We discuss these possibilities in section III.D.

for that occupation in 1950. The unavoidable reliance on occupation-based earnings suggests an important limitation: our measure only captures convergence between occupations and is silent about within-occupation income convergence.

Moreover, our analysis suggests the importance of accounting for heterogeneity, as these patterns vary substantially across sending countries. Immigrants to the US from five sending countries, including the English-speaking countries of England, Scotland and Wales, held significantly higher-paid occupations than US natives upon first arrival, while immigrants from other sending countries started out in equivalent or lower-paid occupations. Similarly, we find that the nature of selection of return migrants varies substantially by country, with negative selection of migrants who returned to England, Italy, Norway, Russia, and Switzerland, and positive selection of migrants who returned to Finland. Direct evidence on return migrants to Norway using the 1910 Norwegian Census confirms the finding of negatively selected return migration for this case. We explore potential mechanisms underlying these results and find suggestive evidence that immigrants hailing from countries with higher real wages, lower shares of the labor force in agriculture, and more similar cultures and religions held higher-paying occupations in the US upon arrival.

Occupational convergence between immigrants and natives may take more than one generation. In the final section, we study how the children of immigrants who came during this era performed in the US labor and marriage markets. On one hand, these second generation migrants spoke English better than their parents did and, having grown up in the US, they might have been more exposed to US norms and culture. On the other hand, occupational differences could persist over generations if, for example, second generation migrants grew up in migrant enclaves, inherited skills from their parents, or used their parents' networks to find jobs. We find

that the evidence is consistent with persistence across generations: when migrants from a certain sending country outperformed US natives, so did second generation migrants, and vice versa. Furthermore, despite substantial variation across countries in the probability of marrying within one's ethnic group, there is considerable persistence in the likelihood of entering an endogamous marriage across generations.

The remainder of the paper proceeds as follows. Section 2 discusses the historical context and related literature. Section 3 describes the data construction and the matching procedures. Section 4 presents our empirical strategy and main results. We estimate the occupation-based earnings penalty (or premium) as well as the earnings distribution for the typical immigrant and then show country-by-country results. Section 5 contains direct evidence on the selection of return migrants from the 1910 Norwegian Census. In Section 6, we consider possible mechanisms for the cross-country variation in immigrant performance. Section 7 analyzes the performance of second generation immigrants relative to their parents, and section 8 concludes.

II. Historical context and related literature

A. Historical context

The US absorbed 30 million migrants during the Age of Mass Migration (1850-1913). By 1910, 22 percent of the US labor force was foreign-born. The foreign-born share of the labor force was even larger outside of the South (29.8 percent), especially in urban areas (38.3 percent).[2] Initially, migrants hailed from countries in northern and western Europe. By 1880, migrant sending countries had shifted toward the poorer regions of southern and eastern Europe (Hatton and Williamson, 1998). Not only were these new immigrants culturally, linguistically and religiously distinct from previous waves, but they were also more likely to be low skilled.

[2] Authors' calculations using the 1910 Integrated Public Use Microdata Series (IPUMS).

For example, in 1900, only 51.2 of Italian immigrants could read and write, compared with 92.7 percent of the German born.[3]

Many native-born residents expressed concerns about the concentrated poverty in immigrant neighborhoods and the low levels of education among immigrant children. Newcomers often lived in overcrowded city tenement buildings with poor ventilation and sanitation (Muller, 1993). Children from immigrant families were more likely than children of the native born to leave school at young ages in order to work in textile factories and other manufacturing industries (Moehling, 1999). Progressive reformers believed that immigrants' behaviors could be changed and championed a series of private initiatives and public legislation, including child labor laws and compulsory schooling requirements, to aid immigrant communities (Lleras-Muney, 2002; Carter, 2008; Lleras-Muney and Shertzer, 2011). Nativist politicians and commentators instead believed that new arrivals would never be able to assimilate into American society (Higham, 1988; Jacobson, 1999).

Concerns about immigrant assimilation prompted Congress to convene a special commission in 1907 to study the social and economic conditions of the immigrant population. The resulting 41-volume report, which was published in 1911, concluded that immigration, particularly from southern and eastern Europe, was a threat to the economic and social fabric of the country. Members of the commission particularly singled out the trend of temporary and return migration as an impediment to assimilation. Two authors of the report, Jeremiah Jenks and W. Jett Lauck, later summarized this view, writing:

> "if an immigrant intends to remain permanently in the US and become an American citizen, he naturally begins at once… to fit himself for the conditions of his new life…If, on the other hand, he intends his sojourn in this country to be

[3] Over 70 percent of German immigrants were literate as early as 1850.

short… the acquisition of the English language will be of little consequence… The chief aim of a person with this intention is to put money in his purse… not for investment here but for investment in his home country"

(quoted in Wyman, 1996, p. 99-100).

The Immigration Commission report provided fuel for legislators seeking to restrict immigrant entry (Benton-Cohen, 2010). In 1917, Congress succeeded in passing a literacy test (after three prior attempts failed), which required potential immigrants to demonstrate the ability to read and write in any language (Goldin, 1994). In 1924, Congress further restricted immigrant entry by setting a strict quota of 150,000 arrivals per year, with more slots allocated to northern and western European countries.

B. Related literature: Immigrant assimilation in the early 20th century

Immigration to the United States picked up again after passage of the 1965 Immigration and Nationality Act, which not only increased the number of visas granted, but also shifted the emphasis for admission from country-specific quotas to preferences based on immigrant skills and family reunification with US citizens. Within a few years of this historic legislation, a literature emerged in economic history re-assessing immigrant performance in the labor market of the early twentieth century.[4] The earliest studies in this area (re-)analyzed the aggregate wage data published by the Immigration Commission and find that immigrants caught up with the native-born after 10 to 20 years in the US (Higgs, 1971; McGoldrick and Tannen, 1977; Blau, 1980).

[4] In a related body of work, Ferrie (1997, 1999) measures immigrant assimilation in the Antebellum period. Lieberson (1980) and Alba and Nee (2003) are two core references in the sociological literature on immigrant assimilation.

A second generation of scholarship examined individual-level wage data from surveys conducted by State Labor Bureaus (Hannon, 1982; Eichengreen and Gemery, 1986; Hanes, 1996). The first analyses of these sources found substantially lower rates of earnings growth for immigrant workers; in some cases, immigrants appear to have experienced no wage convergence with native workers at all. Although differences between these sources present something of an empirical puzzle, Hatton (1997) argues that this discrepancy is due to specification choice. He re-analyzes the state data with two simple modifications and finds that immigrants who arrived at age 25 fully erased the wage gap with natives within 13 years in the US.[5]

A more recent work on immigrant assimilation incorporates data from the federal Census of Population. Unlike the State Labor Bureau surveys, which are confined to specific industries in particular locations (Michigan, Iowa and California), the Census offers complete industrial and geographic coverage. However, in lieu of individual-level wage data, the Census only contains information on occupation. Relying on the 1900 and 1910 Census cross-sections, Minns (2000) finds partial convergence between immigrants and natives outside of the agricultural sector.[6] Immigrants eliminate 30 to 40 percent of their (between-occupation) earnings deficit relative to natives after 15 years in the US.

Overall, the existing literature suggests that immigrant workers experienced substantial occupational and earnings convergence with the native-born in the early twentieth century. In three different datasets – the Immigration Commission reports, state- and industry-level surveys, and the 1900 and 1910 Censuses – immigrants appear to eliminate between 40 and 100 percent

[5] In particular, Hatton (1997) allows for differences in the return to experience for younger and older workers and separates immigrants who arrived as children from those who arrived as adults. The convergence figure reported in the text is based on Hatton (1997, Table 4, columns 1 and 3). Because Hatton estimates different returns to experience parameters for immigrants and the native born, the size of the initial wage gap varies by age. For this calculation, we consider an immigrant who arrives at age 25, at which point the implied wage gap with natives is 0.275, a gap which is erased after the immigrant spends 13 years in the US.

[6] Consistent with our results, Minns finds that the full immigrant population actually earn as much as (or more than) natives. The immigrant deficit explored in his paper is present only outside of the agricultural sector.

of the earnings gap with natives after 15 years in the US. However, all these analyses compare earnings in a single cross-section, a method that suffers from two potentially important sources of bias: selective return migration, and changes in immigrant cohort quality over time.[7] The next section reviews these concerns in the context of the literature on contemporary immigrant flows.

C. Two sources of bias in cross-sectional studies of immigrant assimilation

Workers commonly experience wage growth with time spent in the labor market due to on-the-job training, learning-by-doing, or promotion to supervisory roles. Immigrants may also accumulate country-specific skills with time spent in the US, for example, by learning English and acquiring specific information about the US labor market. Immigrants may start below natives and experience convergence relative to natives if their earnings grow faster with each year of labor market experience due to the accumulation of US-specific skills with time spent in the US. The extent of immigrant-native convergence is estimated using a standard age-earnings profile. We illustrate one such (stylized) profile in equation 1:

$$\ln(earnings_i) = \alpha + f\left(experience_i, B\right) + \gamma I\left(ForeignBorn_i\right) + g\left(YearsInTheUS_i, \Delta\right) + \varepsilon_i \qquad (1)$$

where i indexes individuals. The coefficients in the vector B indicate how labor market experience translates into earnings for the typical worker. γ measures the additional earnings penalty (or premium) that immigrants face upon first arrival in the US. The coefficients in the vector Δ specify whether immigrants are able to subsequently erase some of this penalty with time spent in the US.

[7] We note that Minns (2000) acknowledges the potential bias from changes in the quality of immigrant arrival cohorts.

The methodological debate in the literature centers around the source of identifying variation for the "years in the US" parameters. An early paper by Chiswick (1978) relied on data from a single cross-section of the US Census. He found that, in 1970, the foreign-born experienced faster wage growth than the native-born and overtook natives within 15 years of arrival. However, changes in the quality of arrival cohorts over time can lead to biased estimates of immigrant wage growth in a single cross-section (Borjas, 1985). In our context, immigrants who arrived in the year 1900 were more likely to hail from the southern and eastern European countries, such as Italy and Poland, while migrants who arrived in 1880 were drawn from northern and western Europe. This shift in sending countries, as well as variation in the quality of migrants within a sending country over time, can generate a spurious positive relationship between earnings and time in the US. This concern can be addressed by pooling data from multiple cross-sections and following arrival cohorts over time.[8] In equation 1, this corresponds to replacing the single indicator variable for being foreign born with a vector of dummy variables for year-of-arrival cohorts. Borjas (1985) concludes that, in 1980, half of the apparent convergence in a single cross-section is driven by changes in cohort quality over time.

A second source of bias, which is present even in repeated cross-sections, is selective return migration. The composition of the immigration population can change over Census periods as some migrants return to their home countries (Jasso and Rosenzweig, 1988). Return migration rates were very high in the early twentieth century with estimates ranging from 25 to 75 percent (Gould, 1980; Bandiera, Rasul and Viarengo, 2010). These return migrants may not have been randomly selected from the immigrant population; return migrants may have been mainly the less successful migrants, who left after a trial period in the US, or alternatively they

[8] Hatton (1997) partially addressed the shift in sending countries by separately analyzing assimilation profiles by country of origin for three sending countries (Britain, Ireland and Germany).

may have been primarily the successful migrants, who had saved enough to return home. If return migrants are primarily negatively selected, the immigrant population will lose its lowest-earning members over time, thus causing average income to increase over time and mimicking the pattern of immigrant assimilation.

The problem of selective return migration can be addressed by re-estimating equation 1 with a balanced panel of individuals. The panel identifies individuals who stay in the US over multiple periods; in this case, the "years in US" parameters are identified by following a group of immigrants as they spend an increasing number of years in the country. Lubotsky (2007) uses social security earnings records to build such a panel for the contemporary period; he finds that around 40 percent of the observed convergence between immigrants and natives in repeated cross-sectional data can be attributed to negatively-selected return migration.[9]

III. Data and matching

A. Matching men between the 1900, 1910 and 1920 US Censuses

This section describes the construction of our new panel dataset that follows native-born workers and immigrants through the US Censuses of 1900, 1910 and 1920. We restrict our attention to men between the ages of 18 and 35 in 1900, an age range in which men are both old enough to be employed in 1900 and young enough to still be in the workforce in 1920. We further limit the immigrant portion of the sample to men who arrived in the US between 1880 and 1900. For comparability with the foreign born, 95 percent of whom live outside of the South,

[9] For other panel data analyses, see Borjas (1989), Hu (2000), Edin, Lalonde and Aslund (2000), Duleep and Dowhan (2002), Constant and Massey (2003), Eckstein and Weiss (2004) and Kim (2011). Lubotsky's conclusion is consistent with descriptive evidence from Zakharenko (2008) documenting that return migrants leaving the United States are negatively selected from the immigrant population.

we exclude native-born men residing in a southern state and all black natives regardless of place of residence.[10]

We identify a sample of men in the base year (1900) from two Census sources. For large sending countries (listed in Table 1, panel A), we rely on the 1900 5 percent Integrated Public Use Microdata Series (IPUMS) (Ruggles, 2010) to find immigrants from large sending countries and to randomly select a sample of 10,000 native-born men. To ensure a sufficient sample size for smaller sending countries (Table 1, panel B), we instead compile the full population in the relevant age range in 1900 from the genealogy website Ancestry.com. Altogether, we identify immigrants from 16 sending countries.[11]

We search for viable matches for these men in 1910 and 1920 using the iterative matching strategy developed by Ferrie (1996) and employed more recently by Abramitzky, Boustan and Eriksson (forthcoming) and Ferrie and Long (2011). Figure 1 illustrates our matching procedure by showing one observation in our dataset. The Census manuscript of 1900 reveals that Alexander James was born in Wales in 1871 and moved to the US in 1893. In the US, Alexander worked as a coal miner in 1900. Based on his name, age, and country of birth, we find Alexander James in the 1910 Census. He was still working as a miner. When we find Alexander again in 1920, he had become a foreman, i.e. he had moved up the occupational ladder.

More formally, our matching procedure proceeds as follows:

[10] We also tried including native-born men living in the South into the sample. Because men who live in the South held lower-paid occupations, the immigrant earnings premium increases by around $1,000 in both the repeated cross-section and the panel. Yet the extent of convergence in both samples and the comparison between immigrants in the cross-section and panel (relative to natives) is preserved.

[11] We include men from all European sending countries with at least 3,000 migrants living in the US in 1900, with the exception of Poland, Czechoslovakia and the Netherlands, which made the cut but were nevertheless excluded. Individuals born in Polish or Czech territory were allowed to report these locations as their place of birth on the 1900 Census. This option was removed from the 1910 Census and then restored in 1920 after both countries gained their independence in 1918. Migrants from the Netherlands reported varied birthplaces on the Census, rendering it difficult to find these individuals in Ancestry.com and follow them over time.

(1) We begin by standardizing the first and last names of men in our 1900 samples to address orthographic differences between phonetically equivalent names using the NYSIIS algorithm (see Atack and Bateman, 1992). We restrict our attention to men in 1900 who are unique by first and last name, birth year, and place of birth (either state or country) in our sample. We do so because, for non-unique cases, it is impossible to determine which of the records should be linked to potential matches in 1910 and 1920. Table 1 presents information about the number of potential matches by country.

(2) We identify potential matches in 1910 and 1920 by searching for all men in our 1900 sample in the 1910 and 1920 Census manuscripts available from Ancestry.com. The Ancestry.com search algorithm is expansive and returns many potential matches for each case, which we cull using the iterative match procedure described in the next step.[12]

(3) We match observations forward from 1900 to either the full population (for small countries) or to the set of potential matches (for large countries) in 1910 and 1920 using an iterative procedure. We start by looking for a match by first name, last name, place of birth (either state or country) and exact birth year. There are three possibilities: (a) if we find a *unique* match, we stop and consider the observation "matched"; (b) if we find multiple matches for the same birth year, the observation is thrown out; (c) if we do not find a match at this first step, we try matching within a one-year band (older and younger) and then with a two-year band around the reported

[12] The Ancestry.com search engine aims to maximize potential 'hits' under the assumption that individual users can identify their relatives from a longer list by hand. To this end, it uses many approaches to convert names into their phonetic equivalents and applies a very lax matching rule. For small sending countries, we instead match the complete 1900 population to the complete 1910 and 1920 populations obtained from Ancestry.com.

birth year; we only accept unique matches. If none of these attempts produces a match, the observation is discarded as unmatched.

(4) After matching each sample in 1900 separately to 1910 and 1920, we create our final dataset by restricting to men who were located both in 1910 and 1920.

The second and third columns in Table 1 present match rates and final sample sizes for each sending country and for native born men. Our matching procedure generates a final sample of 22,070 immigrants and 1,891 natives. We can successfully match 19 percent of all native-born men forward from 1900 to both 1910 and 1920. For the foreign born, the average match rate across countries is lower (10 percent), which is expected given that a sizeable number of migrants return to Europe between 1900 and 1920. These double match rates are similar to those in Ferrie (1996) and Abramitzky, Boustan and Eriksson (forthcoming).

B. Occupation and earnings data

We observe labor market outcomes for our matched sample in 1900, 1910 and 1920. Because these Censuses do not contain individual information about wages or income, we assign individuals the median income in their reported occupation.[13] Table 2 reports the ten most common occupations for our sample of matched natives and foreign born workers. Although the top ten occupations are similar for both groups, migrants to the US were less likely to be farmers (18.1 versus 24.8 percent) and more likely to be mine operatives (3.3 versus 1.4 percent). The native born were more likely to be salesmen and clerks, two occupations with high returns to

[13] For observations taken from the 1900 IPUMS (the native born and immigrants from large sending countries), we use the occupation recorded in the digitized micro data. For the remaining countries in 1900 and for all countries in 1910 and 1920, we collect the occupation string by hand from the historical manuscripts on Ancestry.com. We then standardize occupation titles to match those identified in the 1900 IPUMS.

fluency in English. Other common occupations in both groups include managers, operatives, and general laborers.[14]

Our primary source of income data is the "occupational score" variable constructed by IPUMS. This score assigns to an occupation the median income of all individuals in that job category in 1950. For ease of interpretation, we convert this measure into 2010 dollars. Using this measure, our dataset contains individuals representing around 125 occupational categories. Our unavoidable reliance on median earnings by occupation prevents us from measuring the full convergence between immigrants and natives. In particular, we are able to capture convergence due to advancement up the occupational ladder (between-occupation convergence), but we cannot measure potential convergence between immigrants and natives in the same occupation.[15] A further concern with the IPUMS 'occupation score' variable is its reliance on occupation-based earnings in 1950. The decades of the 1940s and 1950s were a period of wage compression (Goldin and Margo, 1992). If immigrants were clustered in low-paying occupations, the occupation score variable may understate both their initial earnings penalty and the convergence implied by moving up the occupational ladder. We address this concern by using occupation-based earnings from the 1900 Cost of Living survey as an alternative dependent variable.

[14] Men who were not employed at the time of the survey reported their last-held occupation. 1910 was the only census in our time period to ask about unemployment. In that year, native-born men of native parentage (age 18-60) had an unemployment rate of 4.4 percent, while 5.7 percent of foreign born were unemployed. This differential unemployment likely contributed to the true earnings gap between immigrants and natives.

[15] We use the 1970 IPUMS to assess the share of total wage convergence between immigrants and natives that takes place between versus within occupational categories. The 1970 Census is the first to record both wage data and year of immigration or years spent in the US for the foreign born. We exclude immigrants who arrived after 1965, the year of major immigration policy change. Immigrants experience 9 log points of total wage convergence relative to natives after spending 30 years in the US and 3 log points of convergence when using an occupation-based measure of earnings. We conclude that our method is likely to capture around one third of total wage convergence between immigrants and natives. We note that this exercise is subject to all the problems we mentioned previously of inferring convergence from a cross sectional data.

C. Comparing matched samples with the full population

Our matched sample may not be fully representative of the immigrant and native born populations from which they are drawn. In particular, men with uncommon names are more likely to be successfully linked between Censuses, and the commonness of one's name could potentially be correlated with socio-economic status. We assess this possibility by comparing men in the cross-sectional and panel samples in 1920. By definition, men in both the panel and repeated cross-sections must have survived and remained in the US until 1920. Thus, by 1920, up to sampling error, any difference between the panel and the repeated cross-sections is due to an imperfect matching procedure.

Table 3 compares the mean occupation score of men in our cross-section and panel samples in 1920. We consider natives and the foreign born separately and re-weight the matched sample to reflect the distribution of country of origins in the 1920 population.[16] Immigrants in the matched sample slightly out-earn their native counterparts by 1920 ($23,500 vs. $23,200). Among natives, the difference in the mean occupation score in the matched sample and the population in 1920 is small ($53) and statistically indistinguishable from zero. In contrast, immigrants in the matched sample have a $369 advantage over immigrants in the representative sample. Country-by-country comparisons reveal that this gap is generated by five sending countries: Belgium, France, Ireland, Italy and Norway. Results are robust to dropping these five countries from the analysis. Overall, we have little concern that the matching procedure generates a systematic bias.

[16] We need to re-weight the matched sample because our universe of potential matches is drawn from 5 percent samples for large countries and from 100 percent samples for smaller countries. We weight according to the 1920 cross-section to reflect the fact that migrants in the panel sample remain in the US until 1920.

D. Other sources of selective attrition

We infer selection of return migrants relative to migrants who settled in the US long term indirectly, by comparing occupational upgrading patterns in the repeated cross-section versus the panel data. Any difference between the panel and the repeated cross-sections is due to selective attrition from the cross-section sample, which is arguably mostly due to selective return migration.[17] However, any form of selective attrition from the repeated cross-sections (such as selective mortality) could drive differences between the panel and the repeated cross-sections.

Selective mortality is not likely a concern. Mortality in 1900 for this age group (ages 15-45) was fairly low and fairly uniform across sending countries. The Irish were slightly more likely to die (8 per 1000) and the Russian were slightly less likely to die (3 per 1000), but mortality among people from other nationality and US natives were all around 5-6 per 1000 (figures by Marriam, 1903, based on 1900 Census). To further evaluate the role of selective mortality, we note that while selective mortality is a potential concern for both native- and foreign-born men, selective return migration is not an issue for the native born, as few native-born men emigrated away from the US. Therefore, one way to test for the presence of selective mortality in our sample is to compare the occupation-based earning patterns of native-born men in the repeated cross-section versus the panel data. We find that the occupation-based earnings of natives are similar in the repeated cross-sections and the panel in all years, suggesting that selective mortality is non-issue for them.[18] We note that this test for selective mortality relies on the assumption that native- and foreign-born men were subject to the same mortality process.

[17] During this period, some immigrants engaged in circular migration, migrating to the US and returning to Europe multiple times (Piore, 1980; Wyman, 1996). Circular migrants will enter the panel sample only if they happen to live in the US on the Census years; otherwise, they will be treated as temporary migrants.

[18] We regress occupation-based earnings score on a dummy for being in the panel sample for the native born. In 1900, for example, the coefficient on this dummy variable is -0.212 (s.e. = 0.294). After adjusting for age differences between the two samples, the difference falls further to -0.130 (s.e. = 0.288). This finding is consistent

Selective name changes by immigrants are also not a likely concern. First, most name changes occurred upon entry to the US and were processed by state or federal officials (for example, at Ellis Island). Any such change would have taken place before we first observe migrants in the 1900 Census and would thus affect neither the panel nor the cross-sectional data. Second, men who changed their name between Censuses are not likely to affect the results. To see this, note first that even though such men would never be included in the panel sample, they would stay in the repeated cross-sections before and after their name change. We can thus test whether migrants in the repeated cross-section are less likely to have a "foreign" name (an indication that they may have changed their name).[19] Indeed, we find that foreign-born men in the panel sample have slightly more "foreign" names than their foreign-born counterparts in the cross-section, which is consistent with the fact that men who change their name after arriving in the US do not enter the panel. Yet the small difference in the "foreignness" index is associated with only a $60 difference in occupation-based earnings (in 2010 dollars) and so is not quantitatively large enough to affect the results.

IV. Immigrant assimilation in panel data

A. Estimating equation

Our main analysis compares the occupational mobility of native-born and immigrant workers by estimating a modified version of equation 1:

$$Occupation\_score_{ijmt} = \gamma_{t-m} + \lambda_m + \eta_t + \alpha_j + \beta_1 Age_{it} + \beta_2 I\left[Age_{it} \geq 35\right] + \beta_3 Age_{it} I\left[Age_{it} \geq 35\right] + \varepsilon_{ijmt} \qquad (2)$$

with the presence of a minimal relationship between socio-economic status and health in the early twentieth century (Frank and Mustard, 1994; Hummer and Lariscy, 2011).

[19] The "foreignness" index is constructed by first calculating the probability of being foreign born conditional on having a given first name (and, separately, a given last name) in the 1900-20 IPUMS samples. The "foreignness" index is then the sum of the two probabilities; the index varies between zero and two. Foreign-born men in the cross-section (panel sample) have an index value of 1.13 (1.23).

where *i* denotes the individual, *j* denotes the country of origin, *m* is the year of arrival in the US, *t* is the (Census) year, and *t-m* is thus the number of years spent in the US.[20] *Occupation score* is a proxy for labor market earnings that varies between (but not within) occupations. The coefficients β_1 through β_3 relate years of labor market experience to the worker's position on the occupational ladder. Following Hatton (1997), we allow the slope of the experience profile to vary by age to account for steep returns to labor market experience for young workers in the early twentieth century.

The vector γ_{t-m} separates the foreign-born into five categories according to time spent in the US (0-5 years; 6-10 years; 11-20 years; 21-30 years; 30 or more years). Equation 2 includes a dummy variable for each time interval, with the native born constituting the omitted category.[21] The sign and magnitude of the coefficient on the first dummy variable (0-5 years) indicates whether immigrants received a premium upon arrival to the US, whereas the difference between this indicator and the remaining dummy variables reveal whether immigrants eventually catch up with or surpass the earnings of natives. Our main specification divides the foreign born into two year-of-arrival cohorts (pre-1890 arrivals versus those who arrived after 1890) to allow for differences in earnings capacity by arrival cohort (Section IV.C explores the sensitivity of the results to the choice of the number of arrival cohorts).

We begin by estimating equation 2 with data from the 1900, 1910 and 1920 IPUMS samples and include Census year fixed effects (results are similar when looking at each single cross-section separately, see online Appendix B). In this case, we omit the arrival cohort dummy (λ_m) to mimic the cross-sectional studies that do not take into account arrival cohort.

[20] In contrast to the existing literature, we include country fixed effects in all specifications. As a result, we do not rely on variation in typical sending countries across arrival years but instead compare immigrants from the same country of origin who arrive in different years.

[21] The rates of convergence for immigrants in the cross-section and the panel are similar if, instead, we estimate a quadratic in years spent in the US (see online Appendix A).

Next, we estimate the same regressions in the repeated cross-sections of 1900-10-20 using an arrival cohort dummy. We can infer from the coefficient on the arrival cohort dummy whether the cohort quality changed. Moreover, comparing between the cross-section and the repeated cross-section allows us to infer how much of the initial occupational penalty can be attributed to differences in the quality of arrival cohorts.

We next compare the repeated cross-section with the panel. The repeated cross-sectional data follows arrival cohorts, rather than individuals, across Censuses. Therefore, comparing the estimates in the repeated cross-sections and panel data allow us to infer whether and to what extent return migrants were positively or negatively selected from the immigrant population. In 1900, the cross-sectional data includes both temporary and permanent migrants. Over time, the temporary migrants return home, leaving only permanent migrants in the cross-section by 1920. In contrast, the panel is restricted to permanent migrants in all years. If we observe more (less) convergence in the cross-section than in the panel, we can infer that the temporary migrants are drawn from the lower (upper) end of the occupation-earnings distribution, thereby leading their departure to increase (decrease) the immigrant average.

In particular, we estimate a single regression that pools the 1900-1920 cross-sections with the matched panel sample. We allow the variables of interest, including the arrival cohort (λ_m) and years spent in the US (γ_{t-m}) fixed effects, to have separate coefficients in the cross-section and panel samples.[22] As before, we reweight observations in the panel sample by country of birth to be representative of the full population, both native- and foreign-born, in 1920.

[22] Note that, by pooling the two data sources, we constrain the year, country of origin, and age effects to be common across the two samples. Results are similar when we run equation 2 separately for the panel and the repeated cross section or when we restrict the arrival cohort effects to be the same in both samples (results are shown in online Appendix C).

We emphasize that observing the occupation assimilation of an individual immigrant requires a panel dataset that follows the individual over time. With the panel sample, we can estimate an assimilation profile for permanent migrants – defined here as migrants who remain in the US for at least 20 years - who demonstrate continued participation in the US labor market (specifically migrants who migrated between 1880 and 1900 and remained until 1920). The repeated cross-sections instead allow us to estimate an assimilation profile for complete migrant cohorts, consisting of both permanent migrants and temporary migrants who will later return to their home country. These patterns are of interest in themselves because they represent the experience of the average migrant in the US at a point in time.

B. Occupational convergence in cross-sectional and panel data

In this section, we estimate equation 2 with the full sample of immigrant and native-born workers. We show that: (1) In the cross-section, immigrants initially hold lower-paid occupations but converge upon natives over time. (2) Following arrival cohorts from 1900 to 1920 in the repeated cross-sections lowers the initial migrant disadvantage. (3) Permanent immigrants (as represented by the panel data) hold higher-paid occupations than natives upon first arrival and experience similar occupational upgrading over time. We conclude that the apparent immigrant disadvantage in a single cross-section is driven by the lower quality of later arrival cohorts (1900 versus 1880) and the negative selection of temporary migrants who eventually return to Europe. Specifically, a comparison between the single and repeated cross-sections suggests that the quality of immigrant cohorts declined over time, while comparing the panel and repeated cross-sectional data implies that return migrants were negatively selected.

Table 4 presents estimates of equation 2 for the cross-section, the repeated cross-sections and our newly-constructed panel sample. The coefficients on the "years in the US" dummy variables indicate the gap between immigrants of a given vintage and the native born. In the cross-section, new immigrants hold occupations that earn $1200 in 2010 dollars below natives of similar age and appear to completely make up this gap over time (column 1). Columns 2 and 3 report coefficients from a specification that pools the cross-section and panel. In the repeated cross-section, immigrants who arrived after 1890 had significantly lower occupation-based earnings than did earlier arrivals, receiving an arrival cohort penalty of $750. Thus, simply by controlling for arrival cohort in column 2, the occupation score gap between recently-arrived immigrants and natives shrinks to $300. In other words, *even within sending countries*, around three-quarters of the initial gap in the pooled cross-section is due to the lower occupational skills of immigrants who arrived after 1890. In the repeated cross-section, immigrants again appear to completely close this (smaller) occupation gap with natives after spending time in the US.

Coefficients for the panel data are reported in column 3. In this subsample, we find *no initial occupation score gap* between immigrants and natives. If anything, immigrants start out $300 ahead of natives, although this difference is not statistically significant. Comparing the two samples suggests that the initial earnings gap in the repeated cross-section is capturing the negative selection of immigrants who end up returning to Europe (temporary migrants).

The differences in the initial immigrant-native gaps and implied rates of convergence between the cross-section and panel samples are underscored in Figure 2. This figure graphs the coefficients on the five 'years in the US' dummy variables in the pooled cross-section, the repeated cross-sections and the panel dataset. In graphical form, it is even easier to see that, in the cross-section, immigrants face an occupation score gap relative to natives upon first arrival,

but are able to erase this gap over time. In contrast, immigrants in the pre-1890 arrival cohort arrived with a much smaller occupation score gap relative to natives. Finally, permanent immigrants in the panel data hold somewhat higher-paying occupations than do natives, even upon first arrival, and retain this slight advantage over time.

C. Alternative specifications and earnings measures

Table 5 reports results from a series of alternative specifications and measures of occupation-based earnings. For brevity, we show only the main coefficients on the "years in US" indicators for the repeated cross-section and panel samples. Online Appendix D contains graphs for each of these specifications in the format of Figure 2. The first section of Table 5 considers alternative specifications for equation 2. In Panel A, we omit the country-of-origin fixed effects, thereby estimating the assimilation profile using variation that occurs both within and between sending countries. In this case, the permanent immigrants fare somewhat better than natives, earning $600 more than a similar native even upon first arrival. This modification does not alter the comparison of permanent and temporary immigrants nor the degree of convergence experienced in each sample.

Panel B includes indicators for a series of finer arrival cohorts (arrived between 1886-1890; 1891-1895; 1896-1900; arrival before 1885 is the omitted category). These controls reduce the initial earnings gap between migrants and natives in the repeated cross-section from -$300 to +$45. In this case, all of the apparent convergence in the cross-section is due to changes in arrival cohort quality. Yet, permanent immigrants continue to earn $600 more than the average immigrant in the cross-section upon first arrival, revealing a similar degree of negative selection among return migrants. In Panel C, we interact the country-of-origin fixed effects with the initial

arrival cohort dummy (arrival after 1890). The premium earned by permanent immigrants relative to both temporary migrants and natives is preserved.

The next section of Table 5 introduces alternative dependent variables. Panel D uses the logarithm of our occupation-based earnings measure. In this case, immigrants in both the cross-section and the panel out-earn natives upon first arrival, by 5 percent and 9 percent respectively. Permanent migrants maintain their advantage relative to the total migrant pool. Differences between the logarithm and levels specifications are driven by the concentration of natives at the top end of the occupation-based earnings distribution; these lucrative occupations are more heavily weighted in the levels specification. The next sub-section discusses the earnings distribution of migrants and natives in more detail.

As we mentioned above, the income distribution was particularly compressed in 1950. In addition, by 1950 farmers earned below the median, whereas from 1900-20 farming was a relatively high-paid occupation. To examine the sensitivity of the results to farmers' earnings, Panel E arbitrarily raises farmers' income by 20 percent in the 1950 occupation-based earnings data. Natives were more likely than immigrants to be owner-occupier farmers; thus, when raising farmers' earnings, the immigrant earnings penalty relative to natives increases. However, the degree of convergence and the comparison between immigrants in the cross-section and panel are unchanged.

Panel F instead replaces the 1950-based earnings measure with mean earnings by occupation from the 1900 Cost of Living survey. Given the greater income inequality in 1900 and the concentration of immigrants in the lower half of the income distribution, the initial gap between immigrants and natives is substantially larger in this specification (between $2,700 and $3,200 in 2010 dollars). Immigrants in both the cross-section and matched panel samples

experience more convergence relative to natives with time spent in the US (around $1000). Nevertheless, as before there is a substantial gap between permanent immigrants and the total immigrant pool in the cross-section, suggesting the presence of negatively-selected return migration. We note that this measure has several disadvantages relative to our main measure based on the Censuses. First, the Cost of Living surveys were not nationally representative but instead focused on urban married households. Second, income in the surveys is missing for a number of occupations (including farmers, which we instead infer from the US Census of Agriculture).

The final section of Table 5 presents estimates of equation 2 that address aspects of the migration decision. Panel G excludes the 20 percent of the migrant sample who arrived in the US before the age of 10, an age at which most people did not work, even in this historical period; we also try cutoffs of age 12 or 14. Young immigrants may experience systematically different rates of assimilation due to heightened fluency in English or education in the US school system (Friedberg, 1993; Hatton, 1997; Bleakley and Chin, 2010). Yet we find similar results to the full sample when we exclude child immigrants.

Panels H and I introduce state fixed effects and interactions between state fixed effects and an indicator for living in an urban area. The state to which a migrant moves is a choice, and so including state fixed effects raises concerns of endogeneity. However, these specifications may shed light on the mechanism underlying the earnings difference between immigrants and natives. First, immigrants may achieve earnings parity with natives by moving to locations with an industry mix conducive to high-paid occupations (Borjas, 2001). Second, immigrants may earn the same nominal wage as natives but face lower real wages if they settle in more expensive states or urban areas. Adding location fixed effects results in a larger initial gap between

immigrants and natives (around $1500), suggesting that the earnings parity in the main specification is achieved largely through location choice. As before, we find a gap between permanent immigrants and the total migrant pool, suggesting negative selection of return migrants, as well as a similar amount of convergence between immigrants and natives.[23]

D. Earnings distribution of natives and immigrants

We find that, on average, immigrants earn less than natives upon first arrival in the cross-section. Table 6 reports percentiles of the earnings distribution for natives and for recent immigrants (those who arrived within the past 10 years) in both the cross-section and panel samples.

The earnings distribution reveals why immigrants in the cross-section face an earnings penalty relative to natives upon first arrival. Although immigrants earn more than natives at the low-end of the earnings distribution, natives hold higher-paid occupations than immigrants at the 75th, 90th and especially the 99th percentiles. The weight placed on the high-end of the earnings distribution in the levels specification explains why the immigrant occupation penalty in levels becomes an occupational premium in logs.

The better performance of long term immigrants relative to natives even upon first arrival is also apparent in the earnings distribution. Permanent immigrants out-earn natives at percentiles below the median. By extension, one can readily see the earnings advantage of permanent immigrants relative to the total migrant pool. Permanent immigrants have a smaller left-tail in the earnings distribution, earning more than the typical migrant at the 10th percentile of the occupation-based earnings distribution.

[23] In online Appendix D, we also graph the implied effects of years spent in the US from a specification that allows immigrants to have their own age-earnings profile. Results are qualitatively similar to those in the main specification.

E. Heterogeneity in convergence and selectivity of return migration by sending country

The typical permanent immigrant in the panel sample holds a slightly higher-paid occupation than the average native, even upon first arrival. However, this pattern masks substantial heterogeneity across sending countries. Figure 3 illustrates cross-country variation in the occupation-based earnings of immigrants relative to the native born. Six of the 16 countries in the current sample hold occupations that pay significantly less than those held by the native born upon first arrival. The size of this occupation-based earnings penalty varies from $1000 (Finland) to $4000 (Portugal) in 2010 dollars. In contrast, immigrants from three English-speaking countries (England, Scotland and Wales), a developed country in Western Europe (France) and one country from the new immigrant stock (Russia) arrived with statistically-significantly more occupation-based skill than the typical native-born worker. The remaining five countries exhibit little difference in earning power relative to natives (Austria, Germany, Ireland, Italy and Sweden).[24]

Figure 4 compares the degree of convergence relative to natives across the 16 sending countries in the panel sample. Convergence is defined as the difference between relative immigrant occupation scores after 30 years in the US and the relative immigration score after just 0-5 years in the country. On the whole, permanent immigrants from every country appear to experience occupation-based earnings growth at the same pace as the native born. Migrants from eight countries experience between $500 and $1000 of convergence relative to natives over this period, while migrants from seven countries actually experience up to $1000 of divergence relative to natives. We note that none of these patterns are statistically significant. Immigrants

[24] Consistent with Table 5, panel H, the average relative immigrant earnings by country declines when we include state-by-urban fixed effects. However, the order and statistical significance of the country-specific immigrant earnings penalties are primarily robust to include state-by-urban fixed effects. Exceptions are the earnings premium of Russian and Scottish migrants, which disappears with these added controls, and the neutral earnings of Irish immigrants, which becomes a $1,700 earnings penalty (in 2010 dollars).

from Finland are the only group that exhibits a statistically-significant amount of divergence, falling further behind natives by over $2000 from their year of first arrival.

Factors like superior levels of education, training or health, could explain why immigrants from the United Kingdom, France and Russia countries hold higher-earning occupations than native born workers upon first arrival in the US and preserve this advantage over time. In contrast, immigrants from Scandinavian countries, Belgium, Switzerland and Portugal display no such advantage and may in fact exhibit lower levels of factors such as education and training than the native born. These immigrants face an earnings penalty upon first arrival and are not able to acquire US-specific skills fast enough to close the earnings gap relative to natives with time spent in the US. Immigrants from Austria, Germany, Ireland and Italy appear similar in their occupation patterns to the native born. We explore suggestive explanations for these cross-country differences below.

The average immigrant in the cross-section and panel samples differ both because of declines in arrival cohort quality and negatively-selected return migration. The direction and magnitude of these two biases vary by country-of-origin. Figures 5 and 6 report evidence of heterogeneity in each factor in turn. We begin by estimating a version of equation 2 with four arrival cohorts (see Table 5, Panel B). Figure 5 reports differences by country between immigrants who arrived between 1880 and 1884 and those who arrived between 1895 and 1900. Countries like Russia and Italy whose immigration waves only began in large numbers in the early 1880s are among those with the largest decline in immigrant arrival cohorts over the period, perhaps because positively-selected "pioneer" migrants are replaced by the more typical migrant over time. However, old immigrant groups like the English and the Irish experience large declines in arrival cohort quality as well during this time.

Figure 6 explores heterogeneity in the implied selection of return migrants by sending country. In particular, we report the difference between immigrants' occupational upgrading relative to natives in the cross-section versus the panel sample by sending country; recall that a negative value indicates that return migrants are negatively selected. The figure reveals statistically-significant negative selection in the return migration flow back to five sending countries (England, Italy, Norway, Russia and Switzerland) and significant positive selection to one country (Finland). The return migrant flow to the remaining ten countries is neutral. Russia is a particularly interesting case. Figure 3 shows that Russian migrants performed well in the US upon first arrival and Figure 4 suggests that return migrants to Russia were particularly negatively selected. These patterns can be explained by the ethnic composition of the Russian migration. The Russian migrant flow is made up of two groups, Jews and non-Jews, who were primarily Poles and other non-ethnic Russians. The Jewish immigrants were both higher skilled and less likely to return to Russia than their non-Jewish counterparts (Perlmann, 1999). In fact, only 7.1 percent of Russian Jews returned to Europe compared with 87 percent of Russian non-Jews (Gould, 1980). Therefore, the return migrant flow is made up primarily of low-skilled non-Jewish Russians.

The height of the bars in Figure 6 do not distinguish between strong negative (positive) selection of a few return migrants, on the one hand, and slight negative (positive) selection of many return migrants on the other. To address this issue, we used country-specific return migration rates reported in either in Gould (1980) or in Bandiera, Rasul and Viarengo (2011) to normalize the coefficients. While the two sources report different average return migration rates (26.3 percent or 59.9 percent respectively), our results are similar when using either source as a benchmark for adjustment. In particular, we multiply each coefficient by the ratio of the average

migration rate across countries to the country's own migration rate. Thus, for a country with average return migration we simply drew a bar identical to that in Figure 6, and for a country with an above- (below-) average return migration rate we drew a shorter (longer) bar than in Figure 6.[25] Because there is little cross-country variation in the rates of return migration, the resulting picture is nearly identical to the pattern reported in Figure 6 (not shown). The one difference worth noting is that the negative selection of return migration to Russia looks even more severe given the low return migration rates found in Bandiera, et al.[26]

V. Direct evidence on Norwegian return migration

Thus far, we have inferred the selection of return migration to Europe indirectly, by comparing cross-section and panel data. This section directly examines the selection of men who returned from the US to Norway, and compares the direct and indirect evidence on selection of return migrants in the case of Norway. Return migration was sufficiently high that the 1910 Norwegian Census added a supplement for individuals who had spent some time in the US. Return migrants were asked to report the date on which they left for US and the date on which they returned, as well as the occupation they held in the US. We use these data to compare the occupational distribution of Norwegian migrants who stayed in the US with those who returned to Norway.

In the 1910 Norwegian Census, occupations are coded according to the Historical International Standard Classification of Occupations (HISCO). For comparison, we convert these

[25] In particular, we used the observed selection term in Figure 6 and the known country-specific migration rates reported in Gould (1980) to back out what the actual selection term by country must have been. We then multiply these terms by the average return migration rate (0.263) in the sample as a whole. For the case of Russia, we use the average return rate of 23.1 percent, a weighted average of the Jewish and non-Jewish return migration rates (7.1 percent and 87 percent, respectively).

[26] This difference is consistent with a Roy selection model that predicts that the smaller the selected group, the more different it will be relative to the remaining population.

values into US Census occupation codes and then into 1950 income. We focus on men between the ages of 18 and 55 in 1910 who migrated to the US between 1880 and 1900 and who returned to Norway between 1900 and 1910 (if they returned). We observe the occupations held in the US by return migrants in the year before their return (sometime between 1900 and 1910), and the occupations of Norwegian migrants who stayed in the US in the 1910 US Census.

We pool 957 migrants in the US and 3,100 return migrants in Norway and regress occupation-based earnings on a dummy for having returned to Norway and a polynomial in age. The coefficient on being a return migrant is -$1659 (s.e. = 225) in 2010 dollars. In other words, return migrants held lower-paid occupations than migrants who remained in the US. This magnitude is remarkably similar to our inference on the extent of negative selection among return migrants generated by comparing cross-section and panel (-$1757, Figure 6).

VI. Explaining cross-country variation in immigrant performance

Figures 3 and 4 document substantial variation in the performance of immigrants from different sending countries in the US labor market. This section explores the relationship between economic and cultural characteristics of source countries and the initial earnings penalty (or premium) that immigrants from these countries face in the US. We emphasize that, because of the small sample size (16 countries) and lack of exogenous variation, these relationships are merely suggestive. Nonetheless, it is interesting to document the source country characteristics that predict success in the New World.

In particular, we regress the earnings penalty (or premium) of recently-arrived immigrants relative to natives on a set of economic characteristics for the sending country in

1880 and on measures of the linguistic, cultural and religious difference between the source country and the US. Results are reported in Table 7.

We begin in column three by regressing the earnings penalty on each sending country characteristic one-by-one. Immigrants from countries with a higher share of the labor force working in agriculture or a lower real wage hold lower-paid occupations relative to natives when they arrive in the US. Residents of these poorer, more agricultural countries may develop fewer skills in their native countries; alternatively, immigrants from these countries could be negatively selected from the sending population. In contrast, immigrants from countries that share a language, cultural background or religious affiliation with residents of the US are more successful in their new destination. Hailing from a country with a similar culture could help immigrants assimilate in the US or may prevent them from facing discrimination in the labor market. Population pressure and health conditions in the source country, as measured by the rates of natural increase and of infant mortality, have no relationship with subsequent immigrant outcomes.

With only 16 country-level observations, we have limited degrees of freedom. Yet in columns 4 and 5, we supplement each of the estimating equation with either the strongest economic or the strongest cultural characteristic. That is, we regress the immigrant earnings penalty on each country characteristic and either the share of the labor force in agriculture (column 4) or the cultural and religious distance between residents of the country and the US (column 5). We find that cultural and religious distances are the most robust predictors of labor market performance upon first arrival.

VII. Second generation migrants in the US labor and marriage markets

A. Labor market performance

Occupational convergence between immigrants and natives may take more than one generation. On the one hand, second generation migrants were educated in the US and, therefore, were likely fluent in English and may have been exposed to US norms and culture. On the other hand, occupational differences could persist over generations if, for example, second generation migrants grew up in migrant enclaves or inherited occupational skills from their parents.

We compare the occupation-based earnings of US-born men whose parents were born abroad to US-born men whose parents were born in the US (hereafter referred to as US natives, even though second generation immigrants are also born in the US). Because Census records are not publicly available, we are unable to construct panels for this era. Instead, we use the 1% IPUMS samples of the US Census from 1900-1950 to compare the children of first generation immigrants from various sending countries to their parents' generation and to US natives.[27] In particular, we define two samples of non-Southern males aged 20 to 60. For first generation immigrants, we use the Censuses of 1900 to 1920 to compare foreign-born men with US natives. Second generation immigrants, defined as men with two parents from the same country of origin, are drawn from the Censuses of 1920 to 1950 and compared with US natives in those years.

We estimate the following age-earnings profile separately for each group and for each country of origin: immigrants (first generation), US natives in the same Censuses and ages as the immigrants, sons of immigrants (second generation), US natives in the same Censuses and ages as the second generation sample:

[27] Note that for second generation migrants and US natives, assimilation patterns will not be biased by return migration. In addition, the first generation migrants in this analysis must have stayed in the US for at least twenty years and thus are unlikely to include men who will later return to Europe.

$$Y_{it} = \begin{Bmatrix} \alpha_t + \beta_a Age_{it} + \beta_{a2} Age_{it}^{\ 2} + \beta_{a3} Age_{it}^{\ 3} + \beta_{a4} Age_{it}^{\ 4}{}_i + \\ Migrant_i * \sum_k \gamma_k YearsIUS_{itk} + \varepsilon_{it} \end{Bmatrix} \quad (3)$$

As before, our outcome variable is occupation-based earnings converted to 2010 dollars. In Figure 7, we illustrate the results from equation 3 for a person who is 25 years old in either 1910 (first generation versus natives) or in 1930 (second generation versus natives). We assume the first generation migrant moved to the US in 1890.[28]

Figure 7 suggests strong evidence of persistence across generations. If the first generation immigrants out-performed natives (England, Scotland, Wales, France, Italy, Austria, Germany, Russia), so did the second generation and vice versa (Norway, Portugal). A notable exception is Finland, in which first generation migrants held lower-paid occupations but second generation migrants held higher-paid occupations. Consistent with Borjas (1994), there is evidence of convergence between natives and immigrants across the immigrant generations, although this convergence is slow for most countries.

B. Marriage market outcomes

Our paper has primarily focused on an economic measure of assimilation, occupational convergence with natives. Another sign of assimilation is the rate of inter-marriage between immigrants and either US natives or members of other immigrant groups. In contrast, groups that have not assimilated into US society may be more likely to enter endogamous marriages within their own community. Endogamy could reflect preferences or constraints. Immigrants or natives

[28] Results are robust to alternative specifications of the age effects.

(or both) may prefer to marry members of their own group, or migrants may simply be more likely to interact with each other at school or in their neighborhoods.

We use the IPUMS 1% sample of the 1920 and 1930 Census to construct the proportion of marriages that are endogamous by country of origin. We calculate the share of married immigrants whose spouse is a first or second generation immigrant from the same country of origin in the 1920 sample for the first generation and the 1930 sample for the second generation.

Figure 8 reports these endogamy rates for the 16 countries in our sample. First generation immigrants exhibit a strong tendency toward endogamy that weakens by the second generation. The mean probability of endogamous marriage falls substantially from 61 percent for the first generation to 32 percent for the second generation (endogamy rates were slightly higher for women). At the same time, there is sizeable variation in the endogamy rate across countries of origin. For example, 89 percent of first generation immigrants from Italy are married to another Italian, compared with only 28 percent of first generation immigrants from Scotland. Furthermore, there is strong persistence in the probability of in-group marriage across generations; the correlation between the first and second generation endogamy rate is 0.90. Some portion of this persistence could be explained by the relative sizes of these country-of-origin groups in the US.

It is interesting to ask whether first generation performance in the US labor market is correlated with assimilation into US society for the second generation. This correlation could arise if parental resources facilitate assimilation or if labor market performance is determined, in part, by cultural similarity (as in Table 7), which in turn enables social assimilation. Figure 9 suggests that immigrant groups that held well-paid occupations relative to natives upon first arrival are somewhat more likely to marry outside their nationality in the next generation. In

particular, we graph the initial earnings penalty (or premium) for first generation immigrants against the share of men in the second generation in an endogamous marriage by country of origin. We find a mild negative relationship whereby immigrant groups that held well-paid occupations relative to natives in the first generation are also more likely to eschew in-group marriage in the next generation. This relationship is weakened by two outliers, Denmark and Switzerland, countries that start out with sizeable earnings gaps relative to natives in the first generation but experience assimilation via the marriage market in the next generation.

VIII. Conclusion

We construct a new panel dataset of native- and foreign-born men in the US labor market during the Age of Mass Migration at the turn of the twentieth century, an era in which US borders were open to European migrants. This period is not only of interest in itself as one of the largest migration waves in modern history, but it is also informative about how migrants assimilated in a world without migration restrictions. Most of the previous research relying on a single cross-section of data has found that recent migrant arrivals to the US held considerably lower-paid occupations than migrants who had arrived many years earlier. This led researchers to conclude that immigrants in this period experienced a substantial degree of assimilation in the US labor market; migrants started with lower-paid occupations than US natives but caught up after they spent some time in the US.

When repeated cross-sections are used, which allow us to follow cohorts of migrants over time, we show that the initial earnings penalty of the typical migrant and his degree of convergence with natives were much smaller. This suggests that the patterns observed in the

single cross-sections reflect a substantial decline over time in the quality of migrant cohorts, rather than a large earning penalty upon arrival and strong degree of convergence to natives.

In our panel dataset, we are able to follow the same migrants (not just the same cohorts) over time. The panel data thus allow us to examine the assimilation of those migrants who remained in the US long term, rather than returning to Europe. We find that immigrants who remained in the US held higher-paid occupations even upon first arrival to the US, and their age-earnings profile was rather similar to US natives. The difference between the repeated cross-section and the panel data is driven by the change in composition of the repeated cross-section with years in the US: return migrants drop out over time. Thus the larger degree of convergence in the repeated cross-sections relative to the panel reflects negative selection of return migrants.

Our paper further shows that it is important to account for differences in migration patterns across sending countries. While permanent migrants from some countries performed better than US natives upon first arrival, migrants from other countries performed worse than natives. Moreover, these differences in performance across countries are shown to be persistent across generations. We further examine the assimilation in the marriage markets of first and second generation migrants from this age of mass migration and find high rates of endogamous marriage and persistence in endogamy across generations. We find evidence that cultural distances between sending country and the US are correlated with migrants' performance in the US labor markets upon arrival.

References

R. Abramitzky, L. Boustan and K. Eriksson, Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, forthcoming.

R. Alba and V. Nee, *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. (Cambridge, MA: Harvard University Press, 2003).

A. Alesina, A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg, Fractionalization. *Journal of Economic Growth* **8** (2003), pp. 155-194.

J. Atack and F. Bateman, 'Matchmaker, Matchmaker, Make Me a Match': A General Personal Computer-Based Matching Program for Historical Research *Historical Methods* **25** (1992), pp. 53-65.

O. Bandiera, I. Rasul and M. Viarengo, The Making of Modern America: Migratory Flows in the Age of Mass Migration. Manuscript, 2010.

K. Benton-Cohen, The Rude Birth of Immigration Reform. *Wilson Quarterly* 2010.

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

H. Bleakley and A. Chin, Age at Arrival, English Proficiency, and Social Assimilation Among U.S. Immigrants. *American Economic Journal: Applied Economics* **2** (2010), pp. 165-192.

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

G. Borjas, Immigrant and Emigrant Earnings: A Longitudinal Study. *Economic Inquiry* **27** (1989), pp. 21–37.

G. Borjas, Long-Run Convergence of Ethnic Skill Differentials: The Children and Grandchildren of the Great Migration. *Industrial and Labor Relations Review* **47** (1994), pp. 553-573.

G. Borjas, Does Immigration Grease the Wheels of the Labor Market? *Brookings Papers on Economic Activity* (2001), pp. 69-133.

L. Carter, A Hard Day's Night: Evening Schools and Child Labor in the United States, 1870-1910. Manuscript, 2008.

B. Chiswick, The Effect of Americanization on the Earnings of Foreign-born Men. *Journal of Political Economy* **86** (1978), pp. 897–921.

A. Constant and D. S. Massey, Self-selection, Earnings, and Out-Migration: A Longitudinal Study of Immigrants to Germany. *Journal of Population Economics* **16** (2003), pp. 631-653.

H. Duleep and D. Dowhan, Insights from Longitudinal Data on the Earnings Growth of U.S. Foreign-Born Men. *Demography* **39** (2002), pp. 485–506.

Z. Eckstein and Y. Weiss, On the Wage Growth of Immigrants: Israel, 1990-2000. *Journal of the European Economic Association* **2** (2004), pp. 665-695.

P-A. Edin, R. LaLonde and O. Aslund, Emigration of Immigrants and Measures of Immigrant Assimilation: Evidence from Sweden. *Swedish Economic Policy Review* **7** (2000), pp. 163-204.

B. Eichengreen and H. Gemery, The Earnings of Skilled and Unskilled Immigrants at the End of the Nineteenth Century. *Journal of Economic History* **46** (1986), pp. 441–454.

J. Fearon, Ethnic and Cultural Diversity by Country. *Journal of Economic Growth* **8** (2003), pp. 195-222.

J. Ferrie, A New Sample of Males Linked from the Public Use Micro Sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census Manuscript Schedules. *Historical Methods* **29** (1996), pp. 141-156.

----------, The Entry into the U.S. Labor Market of Antebellum European Immigrants, 1840–1860. *Explorations in Economic History* **34** (1997), pp. 295–330.

----------, *Yankees Now: Immigrants in the Antebellum United States, 1840-1860*. (New York: Oxford University Press, 1999).

J. Ferrie and J. Long, Long-Distance Migration and Intergenerational Mobility: British Emigration to the U.S. and Canada, 1851-1881. Manuscript (2011).

J. W. Frank and J. F. Mustard, The Determinants of Health from a Historical Perspective. *Daedalus* **123** (1994), pp. 1-19.

R. Fryer and S. Levitt, The Causes and Consequences of Distinctively Black Names. *Quarterly Journal of Economics* **119** (2004), pp. 767-805.

R. Friedberg, The Labor Market Assimilation of Immigrants in the United States: The Role of Age at Arrival. Manuscript (1993), Brown Univ.

O. Galor and O. Stark, Migrants' Savings, The Probability of Return Migration and Migrants' Performance. *International Economic Review* **31** (1990), pp. 463-467.

C. Goldin and R. A. Margo, The Great Compression: The Wage Structure in the United States at Mid- Century. *Quarterly Journal of Economics* **107** (1992), pp. 1-34.

C. Goldin, The Political Economy of Immigration Restriction in the United States, 1890 to 1921. In *The Regulated Economy: A Historical Approach to Political Economy,* eds. C. Goldin and G. D. Libecap. (Chicago: University of Chicago Press, 1994), pp. 223–57.

J. D. Gould, European Inter-Continental Emigration. The Road Home: Return Migration from the U.S.A. *Journal of European Economic History* **9** (1980), pp. 41-112.

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

J. Hannon, Ethnic Discrimination in a 19th-Century Mining District: Michigan Copper Mines, 1988. *Explorations in Economic History* **19** (1982), pp. 28–50.

T. Hatton, The Immigrant Assimilation Puzzle in Late Nineteenth-Century America. *Journal of Economic History* **57** (1997), pp. 34–62.

T. Hatton and J. Williamson, *The Age of Mass Migration: Causes and Economic Impact*. (Oxford: Oxford University Press, 1998).

R. Higgs, Race, Skills, and Earnings: American Immigrants in 1909. *Journal of Economic History* **32** (1971), pp. 420–428.

J. Higham, *Strangers in the Land: Patterns of American Nativism, 1860–1925*. 2nd ed. (New Brunswick, NJ: Rutgers University Press, 1988, Orig. pub. 1955.)

G. Hofstede, *Cultures Consequences: International Differences in Work-related Values*. (Beverly Hills, CA: Sage, 1980).

W-Y. Hu, Immigrant Earnings Assimilation: Estimates from Longitudinal Data. *A.E.R. Papers and Proceedings* **90** (2000), pp. 368–72.

R. A. Hummer and J. T. Lariscy, Educational Attainment and Adult Mortality. In *International Handbook of Adult Mortality. Vol. 2*, eds. R. Rogers and E. M. Crimmins. (New York: Springer), pp. 241-261.

M. F. Jacobson, *Whiteness of a Different Color: European Immigrants and the Alchemy of Race*. (Cambridge, MA: Harvard University Press, 1999).

G. Jasso and M. Rosenzweig, How Well Do US Immigrants Do? Vintage Effects, Emigration Selectivity, and Occupational Mobility. *Research in Population Economics* **6** (1988), pp. 229–253.

S. Kim, Economic Assimilation of Foreign-Born Workers in the United States: An Overlapping Rotating Panel Analysis, Manuscript (2011).
http://faculty.washington.edu/seikkim/seikkim_immorpm.pdf

A. Lleras-Muney, Were Compulsory Attendance and Child Labor Laws Effective? An Analysis from 1915 to 1939. *Journal of Law and Economics*, **45.2** (2002), pp. 401-435.

A. Lleras-Muney and A. Shertzer, The Americanization Movement, 1910-1930. Manuscript 2011.

S. Lieberson, *A Piece of the Pie: Blacks and White Immigrants Since 1880*. (Berkeley, CA: University of California Press, 1980).

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

W. Merriam, *Twelfth Census of the United States, Taken in the Year 1900: Statistical Atlas*, United States Census Office, Washington, 1903.

P. McGouldrick and M. Tannen, Did American Manufacturers Discriminate Against Immigrants Before 1914?. *Journal of Economic History* **37** (1977), pp. 723–746.

C. Minns, 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** (2000), pp. 326-350.

B. Mitchell, *International Historical Statistics 1750-2005: Europe*, *6th edition.* (New York: Palgrave Macmillan, 2007).

C. M. Moehling, State Child Labor Laws and the Decline of Child Labor. *Explorations in Economic History* **36** (1999), pp. 72-106.

T. Muller, *Immigrants and the American City*. (New York: New York University Press, 1993).

J. Perlmann, Selective Migration as a Basis for Upward Mobility? The Occupation of the Jewish Immigrants to the United States, ca. 1900, working paper 1999.

Piore, Michael J. *Birds of Passage: Migrant Labor and Industrial Societies*. (Cambridge University Press, 1980).

S. Ruggles, J. T. Alexander, K. Genadek, R. Goeken, M. B. Schroeder, and M. Sobek, *Integrated Public Use Microdata Series: Version 5.0* [Machine-readable database]. Minneapolis: University of Minnesota, 2010.

I. Sin, *Insights from Book Translations on the International Diffusion of Knowledge*, Stanford Dissertation, May 2011.

J. Williamson, The Evolution of Global Labor Markets since 1830: Background Evidence and Hypotheses. *Explorations in Economic History* **32** (1995), pp. 141-196.

M. Wyman. Round Trip to America: The Immigrants Return to Europe, 1880-1930. (Ithaca, NY: Cornell University Press, 1996).

R. Zakharenko, Return Migration: An Empirical Investigation. MPRA Paper No. 13755 (2009). http://mpra.ub.uni~muenchen.de/13755/

Table 1: Sample sizes and match rates by place of birth

Country	1900 # in universe	Number matched	Match rate, total	1900 # unique	Match rate, unique
A. 1900 source: IPUMS					
Austria	4,722	397	0.084	--	--
England	7,296	916	0.126		
France	11,615	728	0.063		
Germany	19,855	2,891	0.146		
Ireland	9,737	1,115	0.115		
Italy	6,649	1,076	0.162		
Norway	3,541	575	0.162		
Russia	5,641	771	0.136		
Sweden	6,164	633	0.102		
US natives	10,000	1,891	0.190	--	--
B. 1900 source: Ancestry.com					
Belgium	6,060	545	0.090	5,962	0.091
Denmark	34,594	1,980	0.058	17,425	0.114
Finland	23,843	828	0.035	22,197	0.037
Portugal	12,585	584	0.046	8,362	0.070
Scotland	53,091	4,349	0.082	15,529	0.280
Switzerland	22,276	3,311	0.149	20,588	0.161
Wales	17,767	1,342	0.076	9,876	0.135

Notes: The sample universe includes men between the ages of 18 and 35 in 1900. Immigrants must have arrived in the US between 1880 and 1900. We exclude all blacks and native born men living in the South. For large sending countries and the native born, we start with the 1900 IPUMS sample (Panel A). For smaller sending countries, we begin with the complete population in 1900. The text describes our matching procedure. The number of matched cases refers to men who match to both the 1910 and 1920 Censuses. We report the number of unique cases by first name, last name, age and country-of-birth and the match rate for this group in columns 4 and 5 for the smaller countries, for which we have a complete population.

Table 2: Common occupations for natives and foreign-born in matched samples, 1920

	Natives			Foreign-born		
	Occupation	Freq.	Percent	Occupation	Freq.	Percent
1.	Farmer	352	24.82	Farmer	3,301	18.09
2.	Manager	129	9.10	Manager	1,999	10.95
3.	Laborer	117	8.25	Laborer	1,791	9.81
4.	Salesman	75	5.28	Operative	1,102	6.04
5.	Operative	71	5.00	Foreman	603	3.30
6.	Clerical	45	3.17	Mine operative	596	3.27
7.	Carpenter	45	3.17	Machinist	578	3.17
8.	Machinist	45	3.17	Carpenter	529	2.90
9.	Farm laborer	39	2.75	Salesman	495	2.71
10.	Foreman	27	1.90	Clerical	326	1.79
Total		945	66.61		11,320	62.03

Notes: See notes to Table 1 for sample restrictions. Occupations based on 'OCC1950' IPUMS variable.

Table 3: Comparing matched panel sample with population, 1920
Occupation-based earnings in $2010 dollars

	Mean, Panel sample	Difference, Panel sample - population	
		Levels	Logs
Native born	$23,200	52.92	0.010
		(301.546)	(0.013)
Foreign born	$23,471	368.75	0.024
		(127.42)	(0.006)

Notes: Occupation-based earnings based on 'OCCSCORE' IPUMS variable, converted into 2010 dollars. Regressions in columns 2 and 3 pool the 1920 IPUMS cross-section with our matched sample and regress occupation-based earnings on a dummy variable for being in the matched sample. Standard errors are in parentheses.

Table 4: OLS estimates, Age-earnings profile for natives and foreign-born, 1900-1920, Occupation-based earnings in $2010 dollars

	(1) Cross-section	(2) Pooled cross-section and panel	
RHS variable		(a) Cross-section coefficients	(b) Panel coefficients
0-5 yrs in US	-1184.27	-314.14	324.16
	(223.14)	(185.97)	(225.66)
6-10 yrs US	-673.57	53.51	448.96
	(200.01)	(170.30)	(200.74)
11-20 yrs US	-378.28	126.81	295.88
	(171.53)	(131.57)	(143.65)
21-30 yrs US	-273.55	126.06	99.42
	(179.52)	(136.40)	(143.22)
30 yrs in US	-18.00	103.85	149.71
	(217.551)	(176.42)	(177.12)
Arrive 1891+	---	-742.61	-230.78
		(107.11)	(154.45)
Native born	---	---	-118.68
			(167.99)
N	205,458	262,248	

Notes: See Table 1 notes for sample restrictions. Columns report coefficients from estimation of equation 2. Column (1) pools three cross-sections (1900-20); the regression in column (2) adds the matched panel sample. The coefficients in sub-column (a) are interactions between the right-hand side variables listed and a dummy for being in the cross-section, while sub-column (b) reports interactions between the right-hand side variables and a dummy for being in the panel. The omitted category is native-born men in the cross-section. Coefficients on age, Census year dummies, and country-of-origin fixed effects not shown.

Table 5: Robustness for age-earnings profile in panel sample, 1900-1920

	A. Without country FE		B. 4 arrival cohorts		C. Country *x* cohort FE	
	RCS	Panel	RCS	Panel	RCS	Panel
0-5 years in US	-132.53	632.18	44.52	588.60	-6.806	640.95
	(169.89)	(214.37)	(219.02)	(262.28)	(247.55)	(277.98)
6-10 yrs in US	362.45	736.20	86.587	399.58	332.24	702.95
	(145.07)	(180.92)	(196.75)	(226.47)	(228.31)	(250.96)
11-20 yrs in US	474.73	569.06	226.58	347.85	436.95	586.86
	(92.613)	(110.09)	(155.93)	(167.73)	(203.68)	(211.28)
21-30 yrs in US	436.39	359.30	196.66	136.77	417.65	387.28
	(100.04)	(108.00)	(157.90)	(165.04)	(206.22)	(211.53)
30+ yrs in US	300.30	398.08	184.00	198.59	414.38	457.35
	(148.23)	(147.04)	(187.25)	(189.01)	(236.21)	(239.64)
N	262,248		262,248		262,248	
	D. ln(occupation score)		E. Raise farmer income		F. 1900 income	
	RCS	Panel	RCS	Panel	RCS	Panel
0-5 years in US	0.051	0.087	-670.49	26.20	-3241.10	-2717.00
	(0.010)	(0.011)	(182.15)	(221.63)	(148.88)	(186.45)
6-10 yrs in US	0.066	0.083	-340.30	120.39	-2694.03	-2033.67
	(0.008)	(0.009)	(166.45)	(196.15)	(142.32)	(164.90)
11-20 yrs in US	0.063	0.069	-248.11	-36.01	-2257.03	-1972.49
	(0.006)	(0.007)	(128.91)	(140.41)	(113.37)	(119.66)
21-30 yrs in US	0.053	0.060	-227.43	-206.59	-2055.00	-1945.88
	(0.006)	(0.006)	(133.38)	(139.12)	(115.48)	(119.79)
30+ yrs in US	0.044	0.057	-221.87	-129.70	-1828.41	-1755.42
	(0.008)	(0.008)	(170.33)	(169.90)	(137.24)	(138.07)
N	262,248		262,248		264,338	

--Continued on next page--

Table 5, continued

	G. Drop child migrants		H. State FE		I. State * urban FE	
	RCS	Panel	RCS	Panel	RCS	Panel
0-5 years in US	-419.28	212.13	-1654.14	-1138.32	-2218.65	-1430.21
	(191.28)	(232.69)	(198.71)	(301.55)	(197.41)	(301.41)
6-10 yrs in US	-95.30	259.10	-1285.28	-759.39	-2009.73	-1132.61
	(177.25)	(210.17)	(190.91)	(257.23)	(189.70)	(257.31)
11-20 yrs in US	-50.80	67.75	-1193.63	-960.18	-1860.04	-1032.53
	(144.09)	(160.15)	(157.29)	(201.16)	(156.93)	(199.08)
21-30 yrs in US	120.75	-81.71	-1073.68	-1139.49	-1659.34	-1111.17
	(148.59)	(161.72)	(163.78)	(209.51)	(163.46)	(210.92)
30+ yrs in US	118.75	62.60	-1008.40	-577.14	-1539.02	-478.06
	(201.37)	(209.79)	(196.08)	(229.82)	(193.29)	(230.63)
N	246,365		228,793		227,930	

Notes: See notes to Table 4 for sample restrictions. All regressions follow the specification in Table 4 with the exception of the modification listed in panel titles. In Panel B, the four arrival cohorts are 1880-85; 1886-1890; 1891-95; and 1896-1900. Panel C interacts the single cohort indicator (1891-1900) with country fixed effects. In Panel E, we raise farmers' income by 20 percent. Panel F replaces the 1950 occupation score measure with occupation-based income from the 1900 Cost of Living Survey. Panel G drops immigrants who arrived in the US before age 10 or after age 40. Standard errors are in parentheses.

Table 6: Occupation-based earnings distribution, 1900-20
in 2010 dollars

	Cross-section		Panel	
	Immigrants	Natives	Immigrants	Natives
10th	$9,900	$8,100	$11,700	$8,100
25th	$18,000	$12,600	$18,000	$12,600
50th	$20,700	$20,700	$20,700	$20,700
75th	$22,500	$25,200	$22,500	$25,200
90th	$28,800	$34,200	$28,800	$34,200
99th	$37,800	$55,800	$37,800	$56,700

Notes: Occupation-based earnings for men aged 16-38. Immigrants restricted to men who have lived in the US for 10 years or less.

Table 7: Predicting cross-country differences in immigrants' initial occupation-based earnings relative to natives

Dependent variable = Initial difference in occupation-based earnings (immigrants versus natives)

Characteristic of sending country (RHS variable)	Mean/standard deviation of RHS variable	Univariate regression*	Multivariate regression: Add economic variable**	Multivariate regression: Add cultural variables***
Share in agriculture	0.466 (0.172)	-6526.86 (3113.67)	-7476.85 (3619.31)	3546.71 (4309.10)
Real wage	57.726 (25.636)	43.93 (24.67)	23.70 (23.77)	12.79 (17.28)
Natural increase	10.406 (3.635)	-7.62 (169.49)	-85.76 (156.14)	-206.82 (105.02)
Infant mortality rate	174.933 (54.934)	10.02 (10.15)	16.09 (9.48)	7.35 (8.19)
Linguistic distance	0.526 (0.344)	-3419.61 (1534.52)	-2229.88 (2540.56)	1090.03 (1860.67)
Cultural distance	1.053 (0.588)	-2999.37 (677.38)	-2610.20 (961.15)	-1848.62 (920.47)
Religious distance	0.148 (0.045)	-39,433.23 (8484.51)	-39,140.04 (11,222.16)	-22,244.94 (12,943.87)

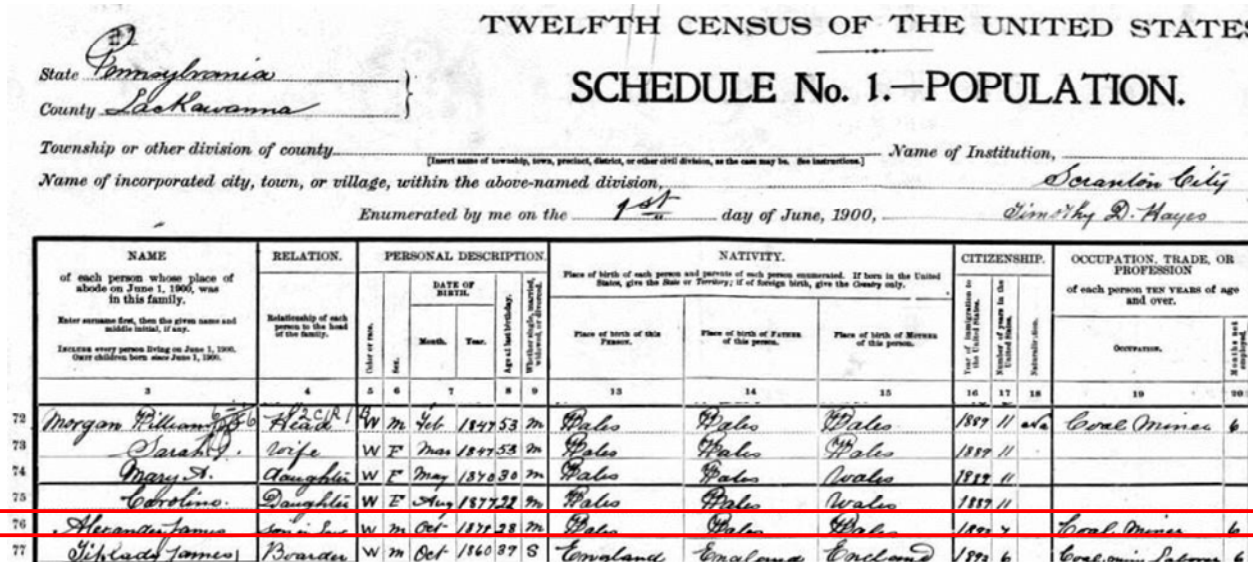
Notes: Initial earnings difference between permanent immigrants and US natives is measured using the coefficient on "0-5 years in US" (relative to natives) in panel sample. N = 16 except for the following RHS variables: infant death (missing for Portugal); cultural distance (missing for Russia) and real wage (missing for Finland, Russia and Switzerland). Economic characteristics are taken from Mitchell (2007) and real wage from Williamson (1995). We accessed the cultural distance measures from Sin (2011). These measures were originally collected by Alesina, et al. (2003); Fearon (2003) and Hofstede (1980). Real wage is indexed to US = 100. Natural increase is crude birth rate minus crude death rate. Infant mortality rate measured per 1000 live births. Linguistic distance theoretically varies between 0 (close) and 1 (far). Religious distance measures the chance that two randomly selected residents, one from the country of origin and one from the US, are of different religious backgrounds. Cultural distance varies between 0.3 (close) and 2.1 (far) in this sample. Standard errors are in parentheses.

* Column 3 regresses the immigrant earnings penalty on each of the RHS variables listed in the first column in turn.

** Column 4 adds "share in agriculture" to each regression (except in the "share in agriculture" row, in which case the real wage is added).

*** Column 5 adds the cultural distance and religious similarity measures to each regression. In "cultural distance" row, we only add religious similarity and in "religious similarity" row we only add cultural distance.

Figure 1: Sample Census manuscripts illustrating matching procedure, 1900-1910-1920

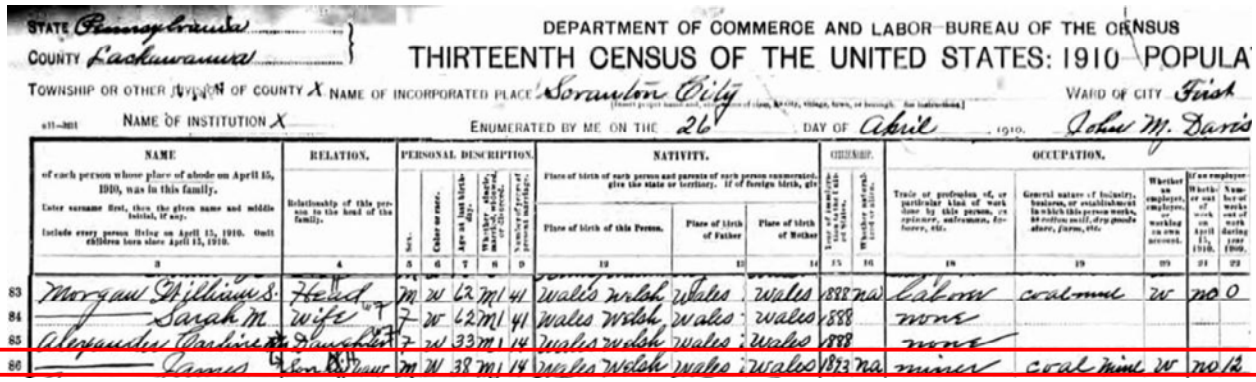


TWELFTH CENSUS OF THE UNITED STATES

SCHEDULE No. 1.—POPULATION.

State Pennsylvania
County Lackawanna
Township or other division of county
Name of Institution,
Name of incorporated city, town, or village, within the above-named division, Scranton City
Enumerated by me on the 1st day of June, 1900, Timothy D. Hayes

	NAME of each person whose place of abode on June 1, 1900, was in this family.	RELATION. Relationship of each person to the head of the family.	PERSONAL DESCRIPTION: Color or race.	Sex.	DATE OF BIRTH: Month.	Year.	Age at last birthday.	Whether single, married, widowed, or divorced.	NATIVITY: Place of birth of this Person.	Place of birth of Father of this person.	Place of birth of Mother of this person.	CITIZENSHIP: Year of immigration to the United States.	Number of years in the United States.	Naturalization.	OCCUPATION, TRADE, OR PROFESSION of each person TEN YEARS of age and over. Occupation.	Months not employed.
	3	4	5	6	7		8	9	10	11	12	13	14	15	16	17
72	Morgan Williams	Head	W	M	Feb	1847	53	M	Wales	Wales	Wales	1869	11	na	Coal Miner	6
73	Sarah J.	wife	W	F	Mar	1847	53	M	Wales	Wales	Wales	1869	11			
74	Mary A.	daughter	W	F	May	1870	30	M	Wales	Wales	Wales	1869	11			
75	Caroline	Daughter	W	F	Aug	1877	22	M	Wales	Wales	Wales	1869	11			
76	Alexander James	Son in law	W	M	Oct	1871	28	M	Wales	Wales	Wales	1891	7		Coal Miner	6
77	Richard James	Boarder	W	M	Oct	1860	39	S	England	England	England	1892	6		Coal mine Laborer	6



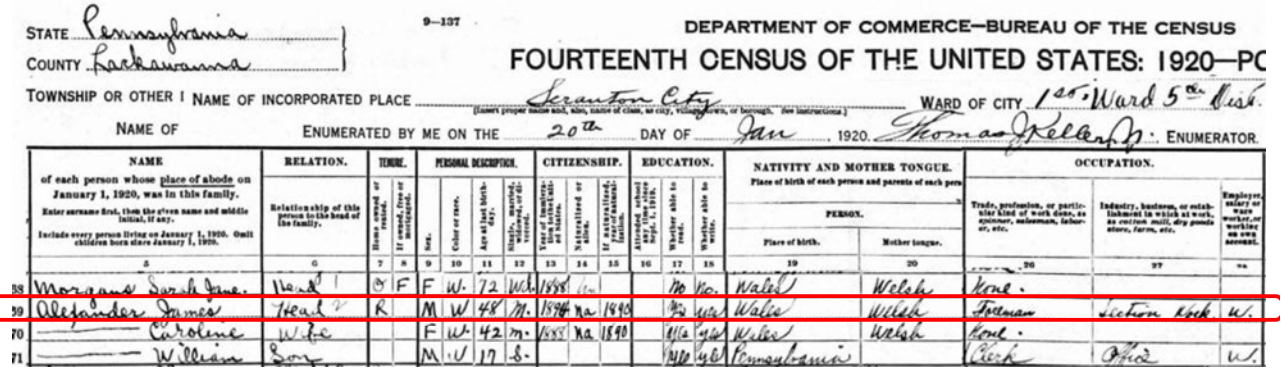
STATE Pennsylvania
COUNTY Lackawanna

DEPARTMENT OF COMMERCE AND LABOR—BUREAU OF THE CENSUS

THIRTEENTH CENSUS OF THE UNITED STATES: 1910 POPULA

TOWNSHIP OR OTHER DIVISION OF COUNTY X NAME OF INCORPORATED PLACE Scranton City WARD OF CITY First
NAME OF INSTITUTION X ENUMERATED BY ME ON THE 26 DAY OF April 1910. John M. Davis

	NAME of each person whose place of abode on April 15, 1910, was in this family. Enter surname first, then the given name and middle initial, if any. Include every person living on April 15, 1910. Omit children born since April 15, 1910.	RELATION. Relationship of this person to the head of the family.	PERSONAL DESCRIPTION: Sex.	Color or race.	Age at last birthday.	Whether single, married, widowed, or divorced.	Number of years of present marriage.	NATIVITY: Place of birth of this Person.	Place of birth of Father	Place of birth of Mother	CITIZENSHIP: Year of immigration to the U.S.	Whether naturalized or alien.	OCCUPATION: Trade or profession of, or particular kind of work done by this person, as spinner, salesman, laborer, etc.	General nature of industry, business, or establishment in which this person works, as cotton mill, dry goods store, farm, etc.	Whether an employer, employee, or working on own account.	If an employer, whether out of work on April 15, 1910.	If an employer, number of weeks out of work during year 1909.
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
83	Morgan Williams S.	Head	M	W	62	M	41	Wales	Welsh	Wales	Wales	1888	na	Laborer	coal mine	W	no	0
84	Sarah M	wife	F	W	62	M	41	Wales	Welsh	Wales	Wales	1888		none				
85	Alexander Caroline	Daughter	F	W	33	M	14	Wales	Welsh	Wales	Wales	1888		none				
86	James	Son in law	M	W	38	M	14	Wales	Welsh	Wales	Wales	1873	na	miner	coal mine	W	no	12



STATE Pennsylvania
COUNTY Lackawanna

9—137

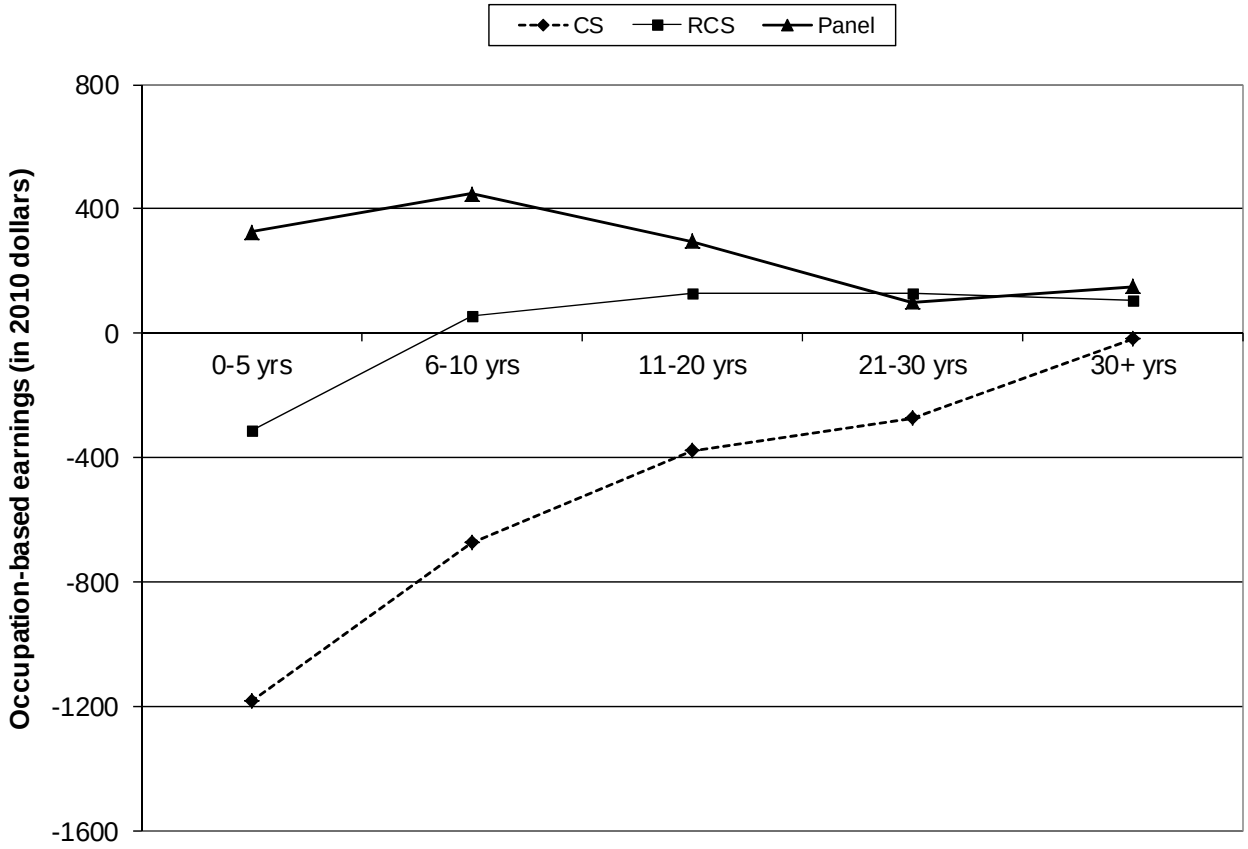
DEPARTMENT OF COMMERCE—BUREAU OF THE CENSUS

FOURTEENTH CENSUS OF THE UNITED STATES: 1920—PO

TOWNSHIP OR OTHER I NAME OF INCORPORATED PLACE Scranton City WARD OF CITY 1st Ward 5th Dist.
NAME OF ENUMERATED BY ME ON THE 20th DAY OF Jan 1920. Thomas J Kelly Jr. ENUMERATOR.

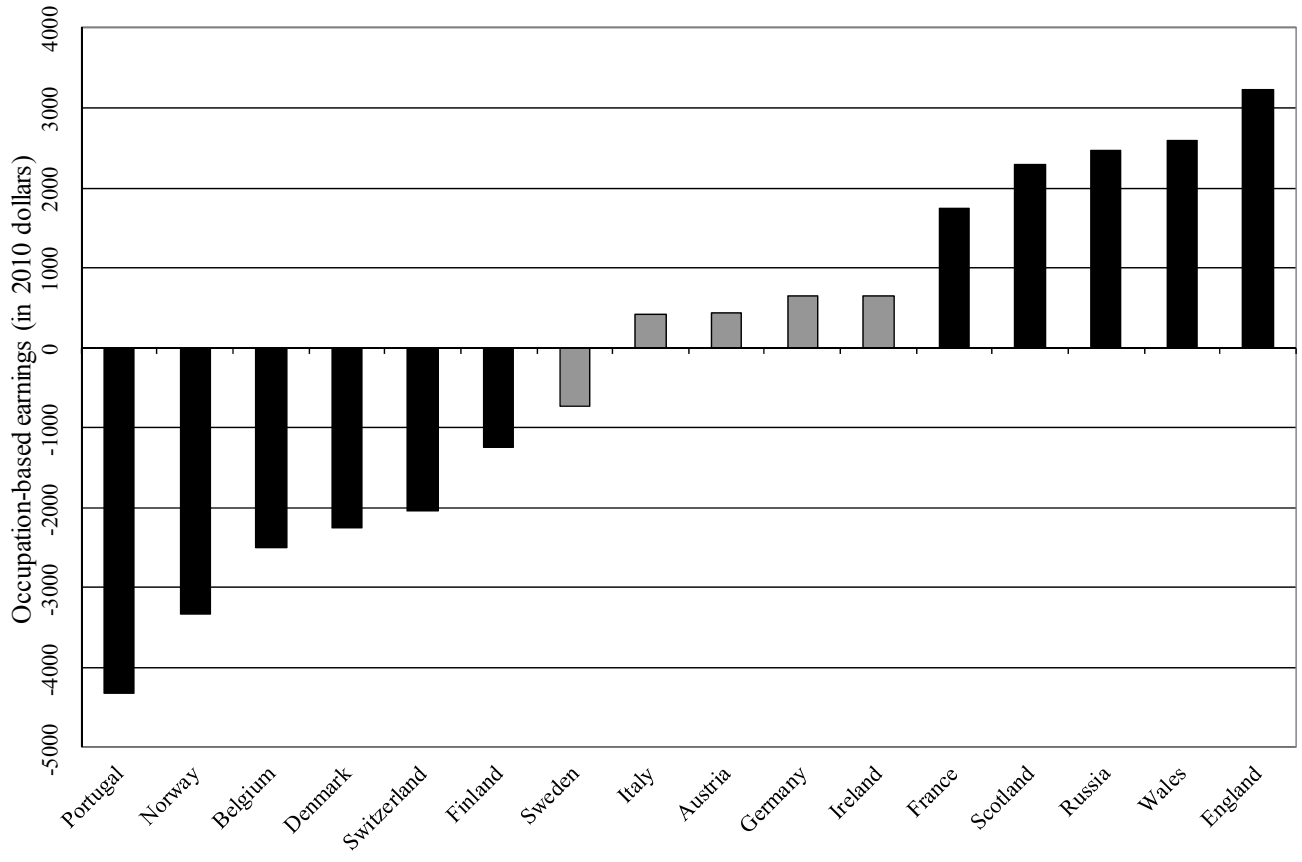
	NAME of each person whose place of abode on January 1, 1920, was in this family. Enter surname first, then the given name and middle initial, if any. Include every person living on January 1, 1920. Omit children born since January 1, 1920.	RELATION. Relationship of this person to the head of the family.	TENURE: Home owned or rented.	If owned, free or mortgaged.	PERSONAL DESCRIPTION: Sex.	Color or race.	Age at last birthday.	Whether single, married, widowed, or divorced.	CITIZENSHIP: Year of immigration to the United States.	Naturalized or alien.	If naturalized, year of naturalization.	EDUCATION: Attended school any time since Sept. 1, 1919.	Whether able to read.	Whether able to write.	NATIVITY AND MOTHER TONGUE. Place of birth of each person and parents of each person. PERSON. Place of birth.	Mother tongue.	OCCUPATION: Trade, profession, or particular kind of work done, as spinner, salesman, laborer, etc.	Industry, business, or establishment in which at work, as cotton mill, dry goods store, farm, etc.	Employer, salary or wage worker, or working on own account.
	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	26	27	na
68	Morgan Sarah Jane	Head	O	F	F	W.	72	Wd.	1888				No	No.	Wales	Welsh	None.		
69	Alexander James	Head	R		M	W	48	M.	1896	na	1898		Yes	Yes	Wales	Welsh	Foreman	Section Work	W.
70	Caroline	Wife			F	W.	42	M.	1888	na	1890		Yes	Yes	Wales	Welsh	None.		
71	William	Son			M	W.	17	S.					Yes	Yes	Pennsylvania		Clerk	Office	W.

Figure 2: Convergence in occupation score between immigrants and native-born workers by time spent in the US, cross-sectional and panel data, 1900-1920



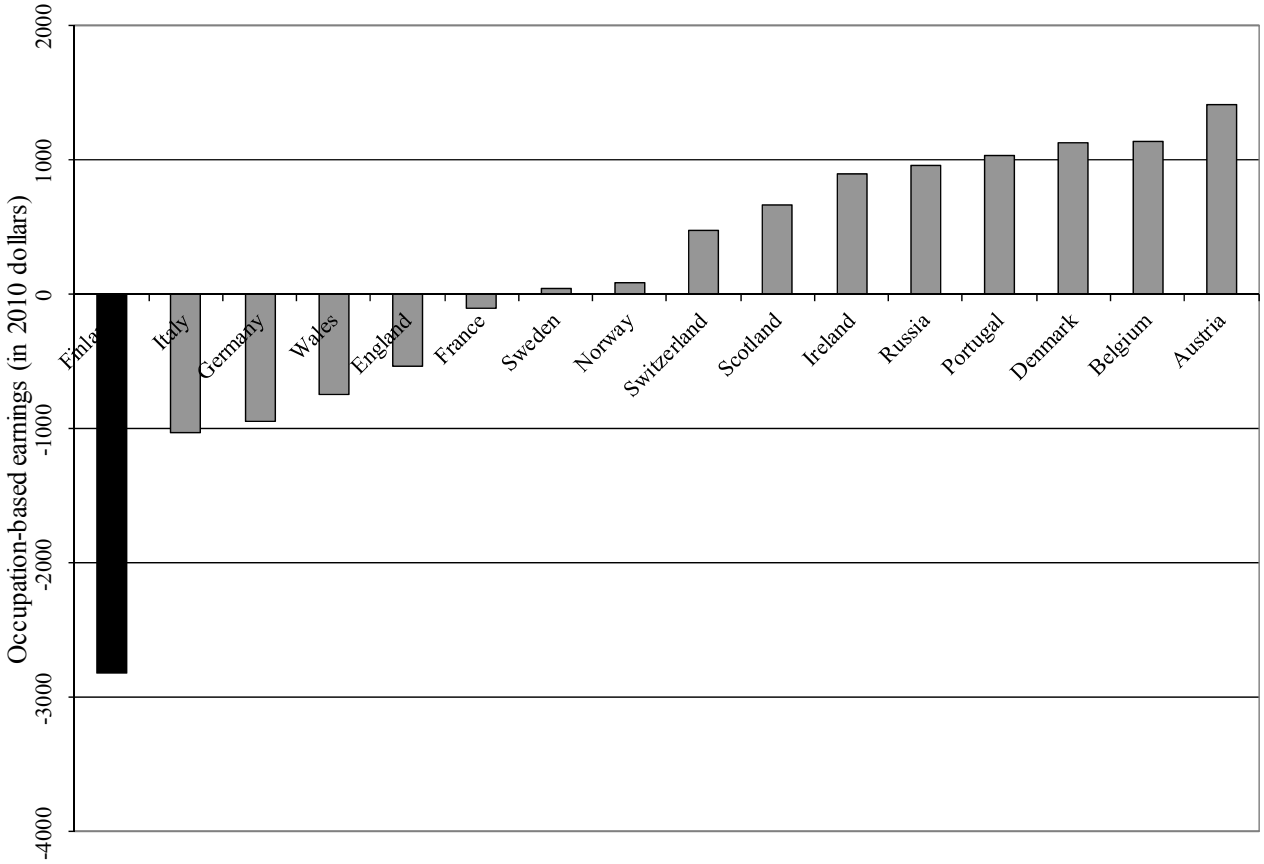
Notes: Plot of coefficients for "years spent in the US" indicators in equation 2. See Table 4 for coefficients and standard errors.

Figure 3: Initial earnings gap between native- and foreign-born in panel sample. Natives versus immigrants upon first arrival (0-5 years in US), By country of origin



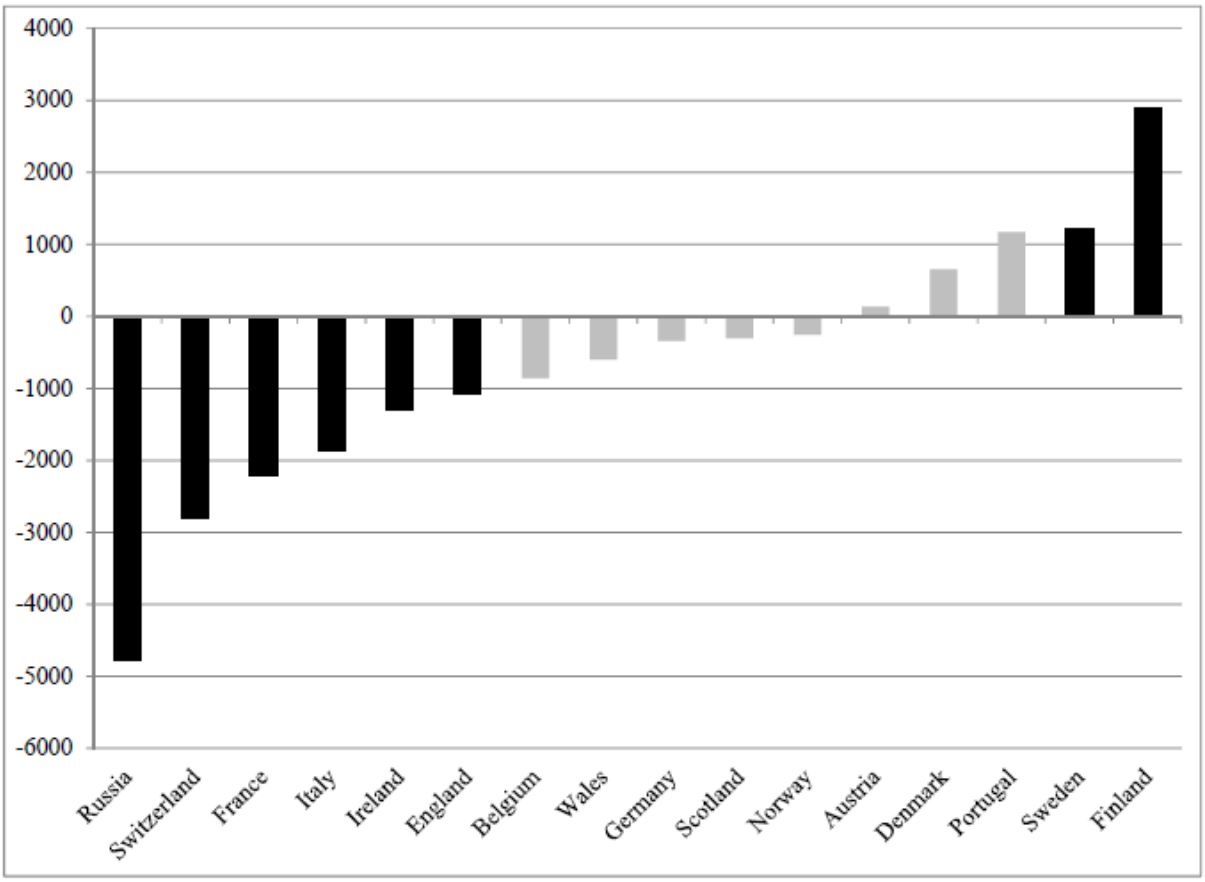
Notes: Graph reports coefficients on interaction between country-of-origin fixed effect and dummy variable for being in the US for 0-5 years from regression of equation 2 in the panel sample. Coefficients that are significantly different from zero are in black.

Figure 4: Convergence between immigrants and natives in panel sample. 30+ years in US vs. 0-5 years in the US, relative to natives, By country of origin



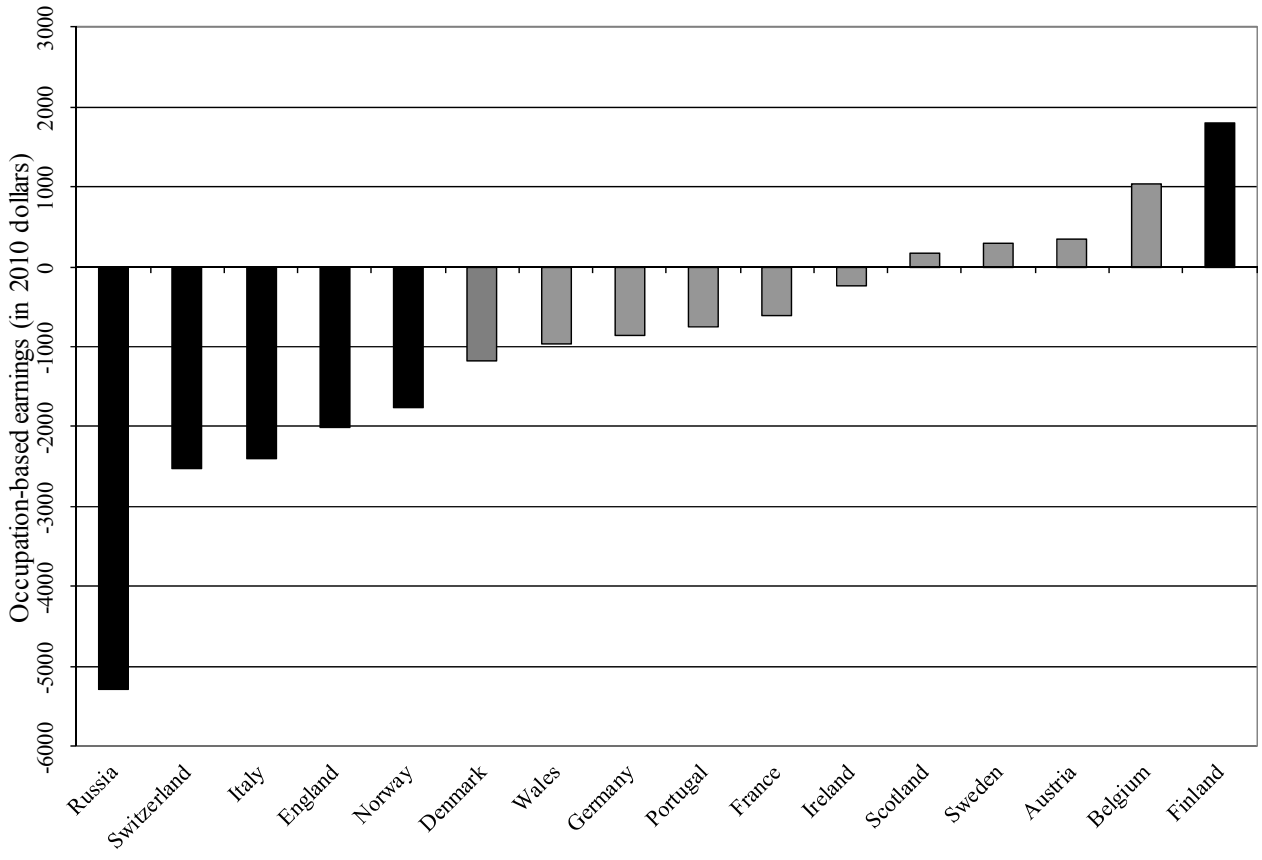
Notes: Graph reports the difference between two coefficients: one interacts a country-of-origin fixed effect with a dummy variable for being in the US for 30+ years and the other interacts the country fixed effect with a dummy variable for being in the US for 0-5 years. Results from regression of equation 2 in the panel sample. Differences that are significantly different from zero are in black.

Figure 5: Changing quality of arrival cohorts,
Difference between immigrant penalty for early and late arrivals in panel sample,
By country of origin



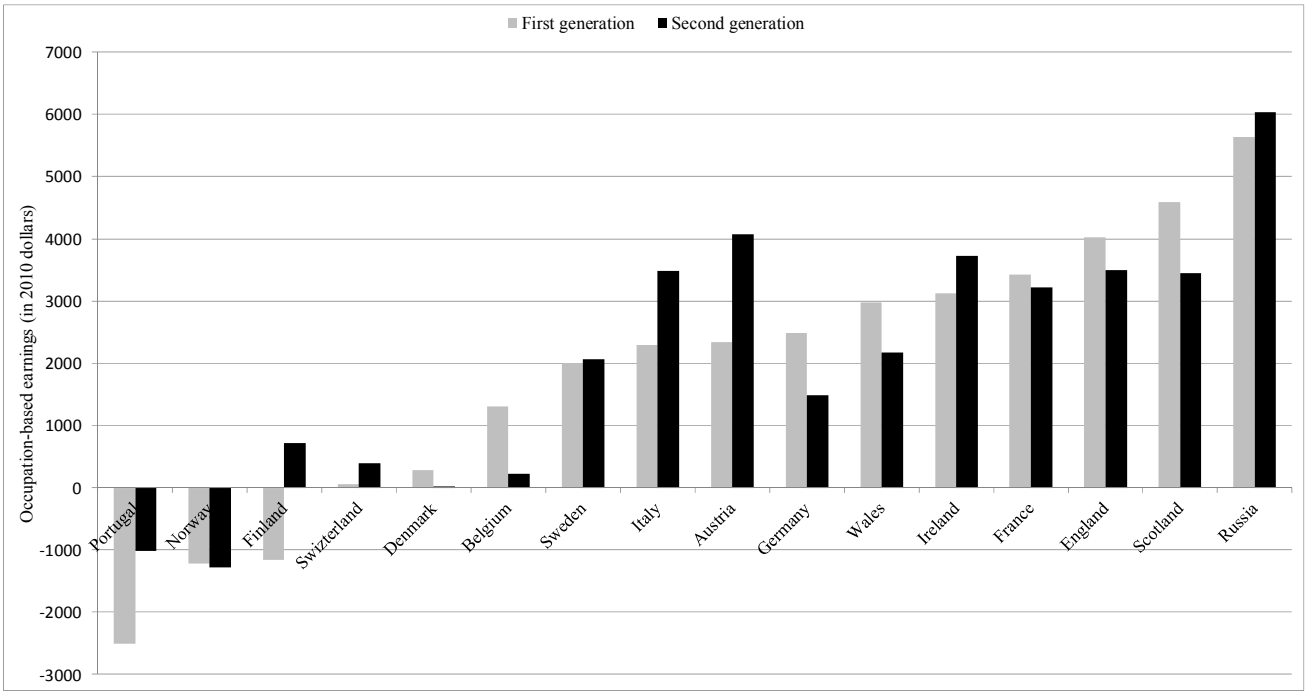
Notes: Estimates based on a version of equation 2 with four dummy variables for arrival cohorts in the panel sample (see Table 5, Panel B). The graph reports the difference between two coefficients: one interacts a country-of-origin fixed effect with the dummy variable for arriving in the US between 1880 and 1884 and the other interacts the country fixed effect with a dummy variable for arriving in the US between 1895 and 1900. Differences that are significantly different from zero are in black.

Figure 6: Implied selection of return migrants.
Difference between estimated convergence in panel and repeated cross-section data,
By country of origin



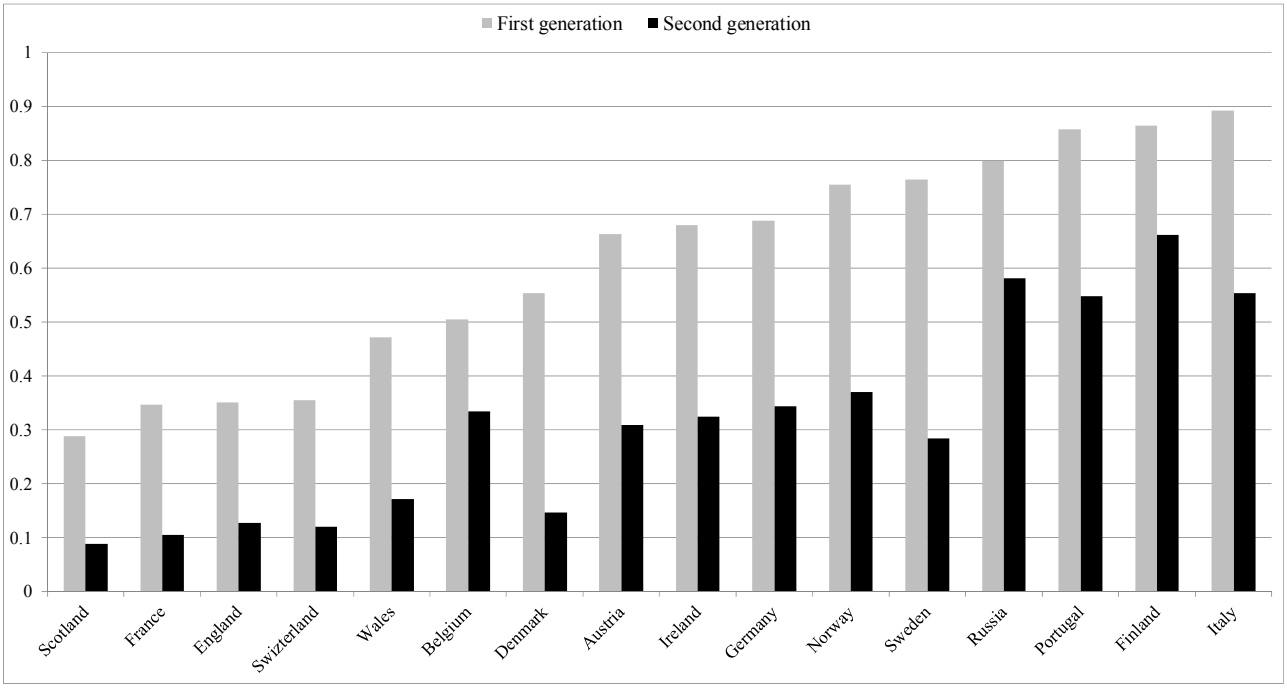
Notes: Figure reports the difference between immigrants' occupational upgrading relative to natives (defined as the difference between occupation-based earnings after 30+ years and after 0-5 years) in the cross-section versus the panel sample, by sending country. Results are from regression of equation 2 pooling the panel and cross-section samples. Coefficients that are significantly different from zero are in black.

Figure 7: Convergence in occupation-based earnings across immigrant generations. First-generation and second-generation migrants versus natives, By country of origin



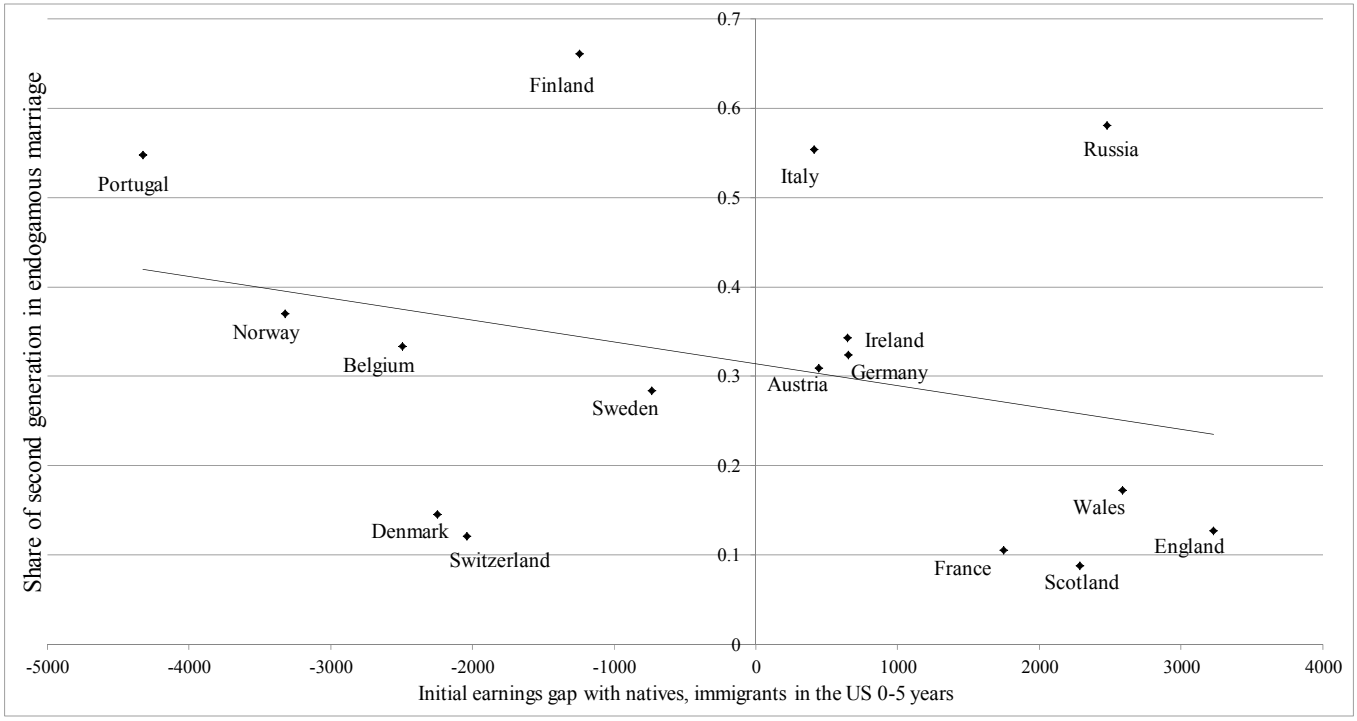
Notes: We estimate the regression equations (3) separately for each group and for each country – immigrants (1st generation), US natives in the same Censuses and ages as the immigrants, sons of immigrants (2nd generation), US natives in the same Censuses and ages as the 2nd generation sample. The bars for the first generation represent the difference in the predicted occupation-based earnings of an immigrant who came in 1890 and is 35 years old in 1910, relative to a 35-year old native. The bars for the second generation represent the difference in the predicted occupation-based earnings of a man born in the US to immigrant parents relative to a man born in the US to native parents, both of whom were 35 years old in 1930. First generation immigrants are taken from the panel sample. Natives and second generation immigrants come from IPUMS data in the respective Census year.

Figure 8: Share of first and second generation immigrants in endogamous marriage, By country of origin



Notes: Endogamous marriage defined as marriage to first or second generation immigrant from same country of origin. Shares calculated from IPUMS samples. First generation includes men who migrated between 1900 and 1920 in the 1920 IPUMS and second generation includes men in 1930 IPUMS who were born in the US and both of whose parents were born in country of origin.

Figure 9: Relationship between first generation immigrant earnings gap and second generation endogamy rates



Notes: The x-axis reports coefficients on the interaction between country-of-origin fixed effect and dummy variable for being in the US for 0-5 years from Figure 3. The y-axis graphs the share of second generation male migrants who are married to either first or second generation immigrants from the same country of origin in 1930. Endogamy rates are taken from the IPUMS 1930 sample.