

**A Nation of Immigrants:  
Assimilation and Economic Outcomes in the Age of Mass Migration\***

**Ran Abramitzky**  
Stanford University  
and NBER

**Leah Platt Boustan**  
UCLA  
and NBER

**Katherine Eriksson**  
UCLA

**[Incomplete draft]**

July 2011

**Abstract:** In the early twentieth century, over 20 percent of the US labor force was foreign born. Prior work finds that immigrants during this era earned less than natives upon first arrival but that their earnings converged over time. Comparisons with newly-assembled panel data of natives and immigrants from 16 European sending countries reveals that the apparent convergence from 1900 to 1920 is driven by a decline in the quality of immigrant arrival cohorts over time and the departure of negatively-selected return migrants. Permanent immigrants actually held higher-paid occupations to natives even upon first arrival, and experienced identical rates of occupational upgrading over time. However, these patterns vary substantially across sending countries.

JEL Code: J61, N30

Keywords: Migration, return migration, assimilation, selection

\* We benefited from the helpful comments of participants at the UC-Davis Interdisciplinary Conference on Social Mobility, the AFD-World Bank Migration and Development Conference, the Labor Markets, Families and Children conference at the University of Stavanger, and seminars at Caltech, Hebrew University, Norwegian School of Economics and UCLA. We especially appreciate conversations with Dora Costa, Joseph Ferrie, Daniel McGarry, Roy Mill, Jean-Laurent Rosenthal, Izi Sin and members of the UCLA KALER group. Roy Mill helped to collect data from Ancestry.com. We acknowledge financial support from the National Science Foundation, the California Center for Population Research and UCLA's Center for Economic History.

## **I. Introduction**

This paper studies the assimilation of European migrants in the US labor market in the early twentieth century. We focus on this period for two reasons. First, this Age of Mass Migration (1850-1913) is characterized by large migration rates, even by modern standards. By 1910, 22.0 percent of US labor force was foreign born. As a result, the pace of immigrant assimilation was an important factor in the development of the US labor force and in US economic growth during this period. Second, during this period, the US maintained an open border policy for European migrants and had only a rudimentary welfare state, allowing us to study the process of assimilation in the absence of government selection policies or social support.

We use newly-constructed panel data to address two questions. First, how rapidly and completely did European immigrants assimilate in the US labor markets? Specifically, how did migrants perform in the US labor markets upon arrival? And did migrants' (occupation-based) earnings converge on those of US natives? Second, are permanent migrants (those who settled in the US) positively or negatively selected from the migrant pool? Understanding the selection of temporary versus permanent migrants is important in this context because, while many migrants settled permanently in the US, up to 70 percent of the migrant flow by sending country returned home (Gould, 1980; Bandiera, Rasul and Viarengo, 2010). Moreover, the direction of selection for return migrants is theoretically ambiguous. Return migrants would be negatively selected if, for example, migrants who were not successful in the US went back to their home countries, and would be positively selected if migrants intended to go back home, and more productive migrants reached "saving targets" faster.

Despite an extensive literature in economic history on assimilation in the early twentieth century, addressing these fundamental questions has been a challenge because of a lack of

historical panel data. Inferring assimilation rates from a cross-section raises well-known biases that limit our ability to study migrant assimilation. First, a single cross-section does not allow the researcher to distinguish differences in the quality of immigrant arrival cohorts from the assimilation of immigrant cohorts over time (Borjas, 1985). For example, if the cohort of migrants that arrived in 1910 is less skilled than the one that arrived in 1900, then an apparent increase in migrants' skill with years in the US may be due to the higher quality of the earlier migrant cohort rather than due to the assimilation process of any particular migrant cohort. Second, even if repeated cross sections are available through which to follow migrant cohorts over time, 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 are more likely to return to their countries of origin, then apparent increases in skills across years may simply be due to negative selection of return migrants rather than to the assimilation of migrants who settled in the US.

To address this challenge, we construct a large panel dataset of 16,000 native born workers and immigrants from 16 sending countries by matching individuals by name, age and place of birth from 1900 to 1920.<sup>1</sup> In particular, we match the 5 percent sample of the 1900 Census from IPUMS (or the full 1900 population from small sending countries) to the 1910 and 1920 Census manuscripts using Ancestry.com. This panel dataset allows us to study the assimilation patterns of permanent migrants who settle in the US for at least twenty years (between 1900-1920). Moreover, comparing the assimilation patterns in the repeated cross-section and panel data allows us to infer the nature of selection of return migrants relative to migrants who permanently settled in the US. In particular, any difference between repeated cross

---

<sup>1</sup> We note that the data collection is not quite complete: so far we collected data on 14 sending countries; results reported in this paper are based on this sample.

sections and the panel is due to selective attrition, which likely reflects selective return migration.<sup>2</sup>

Consistent with the existing literature for this time period (Blau, 1980; Hatton, 1997; Minns, 2000), we find that, in the 1900, 1910 and 1920 cross-sections, immigrants initially held lower-paid occupations but converged upon natives over time. By following arrival cohorts through repeated cross sections from 1900 to 1920, the initial occupation-based earnings gap between immigrants and natives is cut in half. The earnings gap disappears entirely in the panel data; if anything, permanent immigrants held higher-paid occupations to natives upon first arrival and experience similar occupational upgrading over time. 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.

These patterns vary substantially by sending country. Permanent immigrants from eight sending countries, including the English-speaking sending countries of England, Scotland and Wales held higher-paid occupations than US natives upon first arrival, while immigrants from other sending countries started out in lower-paid occupations. However, permanent immigrants from two sending countries that started out with occupation scores below those of natives (Denmark and Portugal) experienced sizeable occupational convergence over thirty years.

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 results, and section 5 concludes.

---

<sup>2</sup> Selective attrition could also be driven by selective mortality. We discuss this possibility in section II.C.

## 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 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).<sup>3</sup> 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. For example, in 1900, only 51.2 of Italian immigrants could read and write, compared to 92.7 percent of the German born.<sup>4</sup>

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 regularly left school at young ages to work in textile factories and other manufacturing industries (Moehling, 1999). Nativist politicians and commentators saw these patterns as evidence that new arrivals would never be able to assimilate into American society (Higham, 1988; Jacobson, 1999). Progressive reformers instead 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).

---

<sup>3</sup> Authors' calculations using the 1910 Integrated Public Use Microdata Series (IPUMS) (Ruggles, 2010).

<sup>4</sup> Over 70 percent of German immigrants were literate as early as 1850.

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 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 failed attempts), which required potential immigrants to demonstrate the ability to read and write (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 20<sup>th</sup> century*

After a forty-year lull, immigration to the United States began again in earnest with the abolition of country-specific quotas in the 1965 Immigration and Nationality Act. Within a few years of this historic legislation, a literature emerged in economic history re-assessing the extent

of immigrant assimilation in the early twentieth century.<sup>5</sup> The earliest studies in this area (re-)analyzed the aggregate wage data published by the Immigration Commission (Higgs, 1971; McGoldrick and Tannen, 1977; Blau, 1980). Blau (1980) finds that immigrants from all sending regions caught up with and even overtook the earnings of the native-born after 10 to 20 years in the US.

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 arrive at age 25 fully erase the wage gap with natives within 13 years in the US.<sup>6</sup>

The most recent work on immigrant assimilation incorporates data from the federal Census of Population. Unlike the State Labor Bureau surveys, which are confined to specific manufacturing 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. Minns (2000) finds partial convergence between immigrants and natives outside of the agricultural sector in each of the

---

<sup>5</sup> In a related body of work, Ferrie (1997, 1999) measures immigrant assimilation in the Antebellum period. Lieberman (1980) and Alba and Nee (2003) are two core references in the parallel literature on immigrant assimilation in sociology.

<sup>6</sup> In particular, Hatton (1997) allows for differences in the return to experience for younger and older workers and separates immigrants who arrive as children from those who arrive 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.

1900 and 1910 Census cross-sections. Immigrants erase 30 to 40 percent of their (between-occupation) earnings deficit after spending 15 years in the US. Our paper differs from Minns (2000) by using the 1920 census (along with 1900 and 1910) to build a panel of immigrant and native workers.

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 of the earnings gap with natives after 15 years in the US. However, these data sources all measure 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.<sup>7</sup> 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. We say that immigrants “assimilate” into the US labor market when their earnings grow faster than those of the native-born for each year of labor market experience due to some positive return to time spent in the

---

<sup>7</sup> Minns (2000) acknowledges the potential role of changes in immigrant arrival quality. He compares the implied immigrant earnings growth with time spent in the US in a single cross section to that observed by following arrival cohorts from the 1900 to 1910 Census and concludes that changes in arrival quality is not a large concern.

US. The extent of immigrant assimilation is estimated using a standard age-earnings profile. We illustrate one such (stylized) profile in equation 1:

$$\ln(\text{earnings}_i) = \alpha + f(\text{experience}_i, B) + \gamma I(\text{ForeignBorn}_i) + g(\text{YearsInTheUS}_i, \Delta) + \varepsilon_i \quad (1)$$

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

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 with 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, recently-arrived immigrants in the year 1900 were likely to hail from the southern and eastern European countries, such as Italy and Poland, whereas long-standing migrants 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 representative samples of arrival cohorts over time.<sup>8</sup> 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 be randomly selected from the immigrant population; in some cases, the least successful migrants may leave after a trial period in the US while, in others, successful migrants may save up enough during a short stay in the US to return home. If return migrants are negatively selected, the immigrant population will lose its lowest-earning members over time, 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. In this case, the years in US parameters are identified by following the outcomes of the same group of immigrants as they spend an increasing number of years in the US. Assembling panel data from Social Security earnings records, Lubotsky (2007) 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.<sup>9</sup> In other

---

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

<sup>9</sup> For other panel data analyses, see Borjas (1989), Hu (2000), Edin, Lalonde and Aslund (2000), Duleep and Dowhan (2002) and Constant and Massey (2003). Lubotsky's conclusion is consistent with descriptive evidence

words, up to 80 percent of observed assimilation in a single cross-section may be capturing some combination of changes in cohort quality and selective return migration. The magnitude of these biases in the contemporary data motivates us to revisit patterns of immigrant assimilation in the early twentieth century.

#### *D. Other sources of selective attrition*

Note that the differences between the repeated cross section and the panel data could be due to any form of selective attrition. In addition to return migration, individuals may fail to appear in the 1910 or 1920 Census due to selective mortality or selective name changes. We doubt that selective mortality drives these results. First, return migration was quantitatively more important during this period. Estimates for decadal return migration rates range from 25 to 75 percent, while the average decadal mortality rate in 1900 for men throughout our age range (18 to 55) was only 10 percent (Haines, XX). Secondly, native born men may also experience selective mortality; yet, we find no differences in the economic outcomes of native born men in the repeated cross-section and the panel samples in 1900 or 1910 when men who may die by 1920 (and therefore fail to be included in the panel sample) are still present in the cross-sections. This pattern is apparent in Figure 1, which is discussed in more detail in Section 4B, and implies that, at least for the native born, selective mortality was unimportant for this age range and during this period.

We also doubt that the observed patterns are driven by selective name changes. First, many name changes occurred as immigrants entered the country and were processed by state or federal officials. All of these changes would have already taken place before our first observation

---

from Zakharenko (2008) documenting that return migrants leaving the United States are negatively selected from the immigrant population.

of the immigrant sample who are measured after entering the country. Second, following Fryer and Levitt (2004), we plan to use the complete 1880 US Census to construct indices of a name’s distinctively ethnic content.<sup>10</sup> We will then be able to test whether or not, by this metric, men in our matched sample were less likely than the typical migrant to have changed their name after spending some time in the US (as proxied by having a distinctively ethnic name).

### III. Data and matching

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

Our goal is to create a panel dataset following native-born workers and immigrants from 16 sending countries through the US Censuses of 1900, 1910 and 1920. We restrict our attention to men between the ages of 18 and 35 in 1900, who are both old enough to be employed in 1900 and young enough to still be in the workforce in 1920, and further limit immigrants to men who arrived in the US before 1900. For comparability with the foreign born, 95 percent of whom live outside of the South, we exclude native men residing in a southern state.

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

---

<sup>10</sup> We calculate a separate name index for each country of origin. The index ranges from zero to two, with a value of zero reflecting the fact that no men in the US with a certain first and last name were born in a given country and a value of two assigned to men whose first and last names are both unique to that country of origin. For example, the first name index for the Norwegian case is equal to:

$$\frac{\Pr[1^{st} \text{ name} | \text{Norwegian\_born}]}{\Pr[1^{st} \text{ name} | \text{Norw\_born}] + \Pr[1^{st} \text{ name} | \text{not\_Norw\_born}]}$$

and likewise for the last name index. The full measure adds these two indices together.

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 (2010) and Ferrie and Long (2011). Our matching procedure proceeds as follows:

- (1) We begin by standardizing the first and last names of men in our 1900 sample to address orthographic differences between phonetically equivalent names using the NYSIIS algorithm (see Atack and Bateman, 1992). Men who are unique by first and last name, birth year, and place of birth (either state or country) in 1900 become the candidates for our matching procedure. Between 28 percent (Scotland) and 98 percent (Belgium) of men meet this requirement. Table 1 presents information about the number of potential matches and the uniqueness rates by country.<sup>11</sup>
- (2) For small sending countries, we compile complete populations with the relevant sample characteristics in 1910 and 1920 from Ancestry.com to serve as potential matches.
- (3) For large sending countries and the native born, the full populations in 1910 and 1920 are too large to compile from Ancestry.com. Instead, we use the (expansive) Ancestry.com algorithm to search for these men in 1910 and 1920 by name, birth year and birth place; this search returns many potential matches for each case, which we cull using the iterative match procedure described in the next step.<sup>12</sup>

---

<sup>11</sup> Note that, for the moment, most of the observations from the large sending countries are considered unique because they are unique in the 5 percent IPUMS sample (panel A). We plan to confirm the uniqueness of these cases using the 1900 Census manuscripts on Ancestry.com.

<sup>12</sup> 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.

- (4) We match unique observations in 1900 forward to the full population (for small countries) and 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. If we find a *unique* match here, we stop and consider the observation “matched.” If we find multiple matches for the same birth year, the observation is thrown out. If we do not find a match at this first step, we try matching first within a one-year band (older and younger) and then with a two-year band around the reported birth year. If neither of these attempts produces a match, the observation is considered to be “unmatched.”
- (5) 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 final columns of 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 14,848 immigrants and 1,436 natives. We achieve a forward double match rate of 19 percent for natives who are unique by name and age in 1900 and an equivalent rate of between 4 percent (Finland) and 28 percent (Scotland) for the foreign born. These double match rates compare favorably with Ferrie’s (1996) match rate of 19 percent for the native born with “uncommon names” (names shared by fewer than 10 others).<sup>13</sup>

### *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

---

<sup>13</sup> As one would expect, Ferrie’s single match rate is larger than but not twice as large as our double match rate. The probability of matching from 1910 to 1920 conditional on being successfully matched between 1900 and 1910 is almost surely higher than the unconditional match rate.

individuals the median income in their reported occupation. For the native born and immigrants from large sending countries, we use the occupation recorded in the IPUMS 1900 digitized sample. 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.

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 are less likely to be farmers (16.6 versus 22.6 percent) and more likely to be common laborers (10.7 versus 6.7 percent). The native born are more likely to be salesmen and clerks, two occupations with high returns to fluency in English. Other common occupations in both groups include carpenters, machinists and merchants.

Our primary source of income data is the ‘occupational score’ variable constructed by IPUMS. This score, measured in 100’s of 1950 dollars, assigns an occupation the median income of all individuals in that job category in 1950. Using this measure, our dataset contains individuals representing XX occupational categories. Our unavoidable reliance on average 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. 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.<sup>14</sup>

### *C. Comparing matched samples with the full population*

Our matched samples may not be fully representative of the immigrant and native born populations from which they are drawn. In 1900 and 1910, matched samples could differ from the full population for two reasons: non-random attrition or selective aspects of the matching procedure. The first concern arises because individuals in the matched sample must survive until 1920 and be resident in the US in that year, while some proportion of the 1900 and 1910 population will die or emigrate before 1920. If either mortality or return migration is correlated with socio-economic status, this attrition could generate discrepancies between the population and the matched sample in 1900 or 1910. However, by 1920, differences between the cross-section and panel samples must be due to features of the matching procedure. In particular, men with uncommon names are more likely to be successfully linked between Censuses. The commonness of one's name could potentially be correlated with socio-economic status.

Table 3 addresses this concern by comparing the mean occupation score of men in our matched samples to a representative sample of the 1920 population in the relevant age range. We consider natives and the foreign born separately and re-weight the matched foreign-born sample to reflect the distribution of country of origins in the 1920 foreign-born population. For the foreign born, we also report country-by-country comparisons between the matched sample and the population. Among natives, the difference in the mean occupation score in the matched sample and the population in 1920 is small (0.05 of an occupation score point) and statistically indistinguishable from zero. In contrast, immigrants in the matched sample have a 0.8 occupation

---

<sup>14</sup> To address this concern, we plan to match our occupations to the 1901 and 1919 Cost of Living surveys.

score point advantage over immigrants in the representative sample. However, the matched sample premium falls to 0.3 occupation score points upon adding country-of-origin fixed effects, which are included in the main analysis. The country-specific comparisons reveal that this gap is generated by five sending countries: Belgium, France, Germany, Italy and Norway. Results are robust to dropping these five countries from the analysis.

#### IV. Immigrant assimilation in panel data

##### A. Estimating equation

We compare the occupational mobility of native-born and immigrant workers by estimating a modified version of equation 1:

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

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 in the US. Occupation score is a proxy for labor market earnings that varies between (but not within) occupations. The coefficients  $\beta_1$  through  $\beta_3$  relate years of labor market experience to the average 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, followed by a long period in which experience does not translate into higher earnings.<sup>15</sup>

We separate the foreign-born into five categories according to their 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

---

<sup>15</sup> Results are robust to instead parameterizing age using a higher-order polynomial.

dummy variable for each interval, with the native born constituting the omitted category. The sign and magnitude of the coefficient on the first dummy variable (0-5 years) indicates whether immigrants arrive with greater or lesser earnings capacity than natives, whereas the difference between this indicator and the remaining dummy variables reveal whether immigrants eventually catch up with or surpass the earnings of natives. We also separate the foreign born into two year-of-arrival cohorts (those who arrived in 1890 or earlier and those who arrived after 1890) to allow for differences in earnings capacity by arrival cohort.

Our preferred specification estimates equation 2 using the balanced panel following individuals from 1900 to 1920. We re-weight the panel regressions by country of birth to be representative of the full population, both native- and foreign-born, in 1920. For comparison, we also estimate equation 2 using separate IPUMS cross-sections from 1900, 1910 and 1920 or using pooled repeated cross-sections from these years.

In comparing the estimates obtained from repeated cross sections with those from panel data, we can infer whether and to what extent return migrants were positively or negatively selected from the immigrant population. Repeated cross-sectional data follows arrival cohorts across Censuses. In the first snapshot of the data (here, 1900), the sample includes both temporary and permanent migrants. Over time, the temporary migrants return home, leaving only permanent migrants 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.

We note that, in the panel sample, we are only able to estimate an assimilation profile for these “permanent” migrants – defined here as migrants who remain in the US for at least 20

years. The rate of assimilation in this subgroup may not generalize to the full immigrant population. However, the assimilation patterns of permanent immigrants are arguably the most interesting precisely because of their persistent settlement in the US.

### *B. Occupational convergence in cross-sectional and panel data*

This section estimates equation 2 with the full sample of immigrant and native-born workers. We show that: (1) In the 1900, 1910 and 1920 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 weakens the initial migrant disadvantage. (3) Finally, permanent immigrants represented in the panel data hold higher-paid occupations to natives upon first arrival and experience similar occupational upgrading over time. We conclude that the apparent convergence in a single cross-section is driven by a combination of changes in cohort quality and selective return migration. 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 separate estimates of equation 2 for the 1900, 1910 and 1920 cross-sections. The coefficients on the ‘years in the US’ dummy variables indicate the gap between immigrants of a given vintage and the native born. Recall that, in this specification, ‘years in the US’ dummies are coincident with ‘year of arrival’ cohort dummies, which therefore cannot be separately included. For each year, we first report a specification that omits country-of-origin fixed effects, mirroring the existing literature, and then add a set of sending country dummy variables. Regressions without country fixed effects rely in part on variation in typical sending

countries across arrival years, while adding fixed effects narrows the comparison to immigrants from the same country who arrived in different years.

In the models without country fixed effects, immigrants who recently arrived in the US earned an occupation score around 1.2 points below that of native-born workers of the same age in every cross-section (a gap of \$850 in 2010 dollars). Within 15 years, immigrants appear to have drawn even with natives and within 20-30 years significantly surpass them. Adding country fixed effects reduces the occupational standing of recent immigrant arrivals by around 0.5 of an occupation score point; in this specification, immigrants appear to catch up with natives but do not surpass them. Overall, though, given the lower starting point, immigrants experience a similar amount of convergence relative to natives in both models.

More accurate estimates of convergence can be achieved by following either arrival cohorts or individuals as they spend time in the US. Table 5 presents estimates of equation 2 for repeated cross-sections and for our newly-constructed panel sample. For comparison, we also report coefficients from a cross-sectional regression that pools data from 1900 to 1920 and focuses on men who meet the panel sample selection criteria. All regressions in Table 5 contain country-of-origin fixed effects.

As before, in the cross-section, new immigrants hold occupations that earn 1.5 occupation score points below natives of similar ages and appear to make up this gap over time (column 1). However, immigrants who arrived after 1890 have significantly lower occupation-based earnings than do earlier arrivals (0.8 of an occupation score point). Thus, simply by controlling for arrival cohort in column 2, the occupation score gap between recently-arrived immigrants and natives shrinks to half of an occupation score point. In other words, *even within sending countries*, around two-thirds of the initial 1.5 point gap in the pooled cross-section is due

to the lower occupational skills of immigrants who arrive after 1890. In the repeated cross-section framework, immigrants again appear to completely close this (smaller) occupation gap with natives after spending time in the US.

Column 3 contains estimates of equation 2 for the panel sample. Unlike the two cross-sectional approaches, the panel data compare natives with permanent immigrants (that is, foreign born men who remained in the US for at least 20 years). In this subsample, we find *no initial occupation score gap* between immigrants and natives. If anything, immigrants start out one occupation score point ahead of natives. Immigrants and natives also experience a similar rate of occupational upgrading over time. The difference between the repeated cross-section and the panel is driven by selective return migration. As immigrants with lower occupational status return home, the composition of the immigrant population in a cross-section becomes increasingly weighted toward men with higher occupational standing, thereby providing an illusion of convergence that is notably absent in the panel data.

Figure 1 uses the coefficients for the repeated cross-section and panel datasets in Table 5 to predict occupation scores for natives and immigrants for men in the birth cohort of 1875 and immigrants in the arrival cohort of 1896. In the repeated cross-section, graphed in grey, 25 year old natives earned an occupation score of 22 in 1900, equivalent to annual earnings of \$20,500 in 2010 dollars. After 20 years in the labor market, these native-born men had increased their occupation scores to 25 (or by \$3,000). The typical immigrant in the repeated cross section held an occupation with a score 1.5 points below that of natives in 1900, both because he would have faced an initial arrival penalty and because of the lower quality of men in this later arrival cohort. After 20 years in the US, immigrants in this cohort reduce the gap by 0.7 of an occupation point.

The predicted occupation scores of immigrants and natives in the panel sample are graphed in black. The permanent immigrants in the panel sample do not face an occupation-based earnings penalty relative to natives upon first arrival, but neither do they surpass natives over time. Both immigrants and natives in the birth cohort of 1875 earn an occupation score of around 21.5 in 1900 and increase their occupation score to around 26 points by 1920. Note also that native-born men in the repeated cross-section and panel samples have very similar occupation scores throughout the period, implying that selective mortality (at least among natives) is not contributing to bias between the two samples.

The differences in the initial immigrant-native gaps and implied rates of convergence between the cross-section and panel samples are underscored in Figure 2a. 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 pooled cross-section, immigrants face an occupation score gap with 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.

Table 6 reports the pooled and repeated cross-section and panel regressions using a series of alternative samples or specifications. Panel A excludes immigrants who arrived before the age of 10; 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). 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). Panel C interacts the country-of-origin fixed effects with the initial arrival cohort dummy (arrival after 1890). In all cases, the pattern outlined above whereby the apparent convergence in the cross-section can be attributed to declining arrival cohort quality and negatively selected return migration is preserved.

The final panel of Table 6 instead uses the natural log of occupation score as a dependent variable. In so doing, the initial occupation score gap between immigrants and natives that was present in the level specification disappears, even in the pooled cross-section. This modification occurs because, in this period, the highest-paid occupations are disproportionately held by natives and the levels specification weights these lucrative occupations heavily. In log terms, immigrants appear to hold equally-paid occupations to natives, even upon first arrival, and eventually to earn four percent more than natives in occupation-based terms after spending 30 years in the US. However, the differences in assimilation patterns between the cross-sections and the panel remain the same as before, as can be seen most readily when comparing the two panels of Figure 2. In comparison with the repeated cross-section, we again see that initial immigrant earnings are biased downward by the lower skills of immigrants who arrived after 1890. Furthermore, in the log specification, permanent immigrants in the panel sample arrive with a sizeable occupation-based advantage relative to natives (12 log points). Over time, natives converge upon permanent immigrants from below so that this immigrant advantage is cut in half after 30 years in the US.

### *C. Heterogeneity by sending country*

In the panel sample, we find that the full population of permanent immigrants holds higher-paid occupations than the average natives, even upon first arrival. However, this pattern

masks substantial heterogeneity across sending countries. Figures 3 and 4 depict variation in both initial occupation score and the extent of occupational growth and convergence with natives by place of birth.

Figure 3 illustrates cross-country variation in the initial occupation score of immigrants relative to the native born. Six of the 14 countries in the current sample hold occupations with scores equal to or lower than those of the native born upon first arrival. The size of this occupation-based earnings penalty varies from one (Switzerland) to five occupation score points (Portugal). In contrast, immigrants from three English-speaking countries (England, Scotland and Wales), two developed countries in Western Europe (France and Germany) and three countries from the new immigrant stock (Austria, Italy and Russia) arrived with more occupation-based skill than the typical native-born worker.<sup>16</sup>

Figure 4 compares the degree of convergence relative to natives across the 14 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. Two sending countries (Denmark and Portugal) whose immigrants started out in lower-paid occupations experienced substantial convergence with natives over time. Danish and Portuguese immigrants gained around 1.5 occupation score points relative to natives after spending 30 years in the US. In addition, two countries that started out with higher occupational standing relative to natives gained even further ground (France and Scotland), though neither did so by a sizeable amount. Immigrants from nine of the remaining sending

---

<sup>16</sup> Given their low literacy rates, we were surprised to find that Italian immigrants hold (slightly) higher-paid occupations than natives upon first arrival. We speculate that Italian immigrants who arrived before 1900 may have been disproportionately from the North and may not have been representative of the later southern Italian migrant wave. Italian immigrants who arrived before 1900 are less likely to be illiterate (21.4 percent) than those who arrived between 1901 and 1920 (25.7 percent). Yet, even these lower rates are higher than nearly all other sending countries in our sample.

countries experienced minor *divergence* from natives, although none of these countries lost more than a full occupation score. Finnish immigrants were the only group to begin with lower-paid occupations than natives and fall even further behind over time.

Taken together, we conclude that permanent immigrants from the United Kingdom, Germany and France, and the new sending areas of Austria, Italy and Russia are positively selected relative to the native born; because of superior levels of education, training or health, immigrants from these countries held higher-earning occupations than native born workers upon first arrival in the US. Immigrants from Denmark and Portugal appear to be “weakly positively selected” relative to the native born. That is, although immigrants from these sending countries held lower-earning occupations upon first arrival, they were able to acquire US-specific skills at a fast enough rate to entirely converge with, or at least substantially close the gap with, US natives. Immigrants from the remaining four sending countries (Belgium, Finland, Norway and Switzerland) display no evidence of positive selection and may, in fact, be negatively selected relative to the native born.

Finally, Figure 5 explores heterogeneity in the implied selection of return migrants by sending country. As we argued above, comparing estimated convergence in the repeated cross-section and panel data indicates the direction of selection in the return migrant flow. Figure 5 reports the difference between estimated convergence in the panel versus repeated cross-sections after 30 years in the US; recall that a negative value indicates that return migrants are negatively selected, thereby biasing upward the estimated convergence in the cross-section. The figure reveals positive selection in the return migration flow back to two sending countries (Finland and France); neutral return migration to four sending countries (Belgium, Germany, Scotland and Wales) and negative selection in the return migration flow to eight sending countries (Austria,

Denmark, England, Italy, Portugal, Norway, Russia and Switzerland).<sup>17</sup> In other words, we find evidence of both kinds of return migration: positively-selected return migrants who may have followed a target savings strategy in the US before returning home to invest in their home country and negatively-selected return migrants who may have been unlucky in their ability to secure a good job or move up their occupational ladder in the US.

The height of the bars in Figure 5 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 Gould (1980) and normalized the coefficients by the ratio between the average migration rate (26.3%) and the country's actual migration rate. Thus, for a country with average return migration we simply drew a bar identical to that in Figure 5, and for a country with an above- (below-) average return migration rate we drew a shorter bar than in Figure 5.<sup>18</sup> 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 5 and therefore not shown.

## **V. Intergenerational assimilation: Comparing second generation migrants to native-born men of native parentage**

**\*\*Section under construction\*\***

## **VI. Conclusion**

**\*\*Section under construction\*\***

---

<sup>17</sup> Russia stands out as having particularly negatively selected return migration, perhaps because few of its out-migrants, many of whom were Jews, were saving up with the intention of returning. Therefore, those migrants who *did* return were more likely to be those who faced negative economic shocks in the US.

<sup>18</sup> In particular, we used the observed selection term in Figure 5 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 assumed that the migration flow was 80 percent Jewish, who had a return rate of 7.1 percent, and 20 percent non-Jewish, who had the much higher rate of 87 percent.

## 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, NBER Working Paper 15684, January 2010.
- R. Alba and V. Nee, *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. (Cambridge, MA: Harvard University Press, 2003).
- 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, pp. XX.
- 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, The Economics of Immigration. *Journal of Economic Literature* **32** (1994), pp. 1667-1717.
- 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.

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. 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).

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

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

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.)

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

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.

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. Lieberman, *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.

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 20<sup>th</sup> Century. *Explorations in Economic History* **37** (2000), pp. 326-350.

C. M. Moehling, State Child Labor Laws and the Decline of Child Labor. *Explorations in Economic History* **36.1** (1999), pp. 72-106.

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

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.

Wyman, Mark. *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 Potential Universe	Unique in 1900	Number matched	Match rate 1900-1910-1920
<b>A. 1900 source: IPUMS*</b>				
Austria	4,722	4,720	397	0.084
England	7,296	7,296	916	0.126
France	11,615	11,615	728	0.062
Germany	19,855	19,833	2,891	0.146
Ireland**	9,737	9,730	1,115	0.114
Italy	6,649	6,625	1,076	0.162
Norway	3,541	3,362	575	0.171
Russia	5,641	5,610	771	0.137
Sweden**	6,164	X	X	X
US natives	10,000	9,870	1,891	0.192
<b>B. 1900 source: Ancestry.com</b>				
Belgium	6,060	5,962	545	0.091
Denmark	34,594	17,425	1,980	0.114
Finland	23,843	22,197	828	0.037
Portugal	12,585	8,362	584	0.070
Scotland	53,091	15,529	4,349	0.280
Switzerland	22,276	20,588	3,311	0.161
Wales	17,767	9,876	1,342	0.135

\* Need to run men from the 1900 IPUMS through Ancestry.com to confirm uniqueness in 1900 in the full population

\*\* Not yet incorporated into dataset

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

	Natives			Foreign-born		
	Occupation	Freq.	Percent	Occupation	Freq.	Percent
1.	Farmer	325	22.63	Farmer	2,739	16.59
2.	Laborer	96	6.69	Laborer	1,498	10.66
3.	Salesman	46	3.20	Miner	506	3.31
4.	Carpenter	44	3.06	Carpenter	468	2.62
5.	Machinist	40	2.79	Machinist	369	2.58
6.	Clerk	38	2.65	Foreman	364	2.23
7.	Manager	32	2.23	Proprietor	276	1.71
8.	Foreman	28	1.95	Manager	273	1.63
9.	Merchant	21	1.46	Merchant	226	1.58
10.	Engineer	21	1.46	Tailor	211	1.50
Total		691	48.12		6,930	44.25

Notes: The native column includes only men born outside of the South.

**Table 3: Comparing occupation scores of matched sample and population, 1920**

Country	Mean occupation score, Matched sample	Difference in occupation score (Population - Matched sample)
Native born	25.77	-0.055 (0.333)
All foreign, weighted	26.65	-0.832 (0.177)
With country FE	26.65	-0.278 (0.175)
Austria	25.71	0.711 (0.812)
Belgium	24.02	<b>-1.917</b> <b>(1.149)</b>
Denmark	23.96	-0.448 (0.631)
England	28.10	0.413 (0.594)
Finland	20.40	0.027 (0.669)
France	27.20	<b>-1.594</b> <b>(0.865)</b>
Germany	25.60	<b>-0.622</b> <b>(0.339)</b>
Italy	26.10	<b>-1.045</b> <b>(0.366)</b>
Norway	21.52	<b>1.322</b> <b>(0.587)</b>
Portugal	21.93	0.364 (0.883)
Russia	29.45	-0.239 (0.665)
Scotland	28.76	-0.040 (0.534)
Switzerland	23.38	-0.484 (0.699)
Wales	27.57	0.398 (0.815)

**Table 4: Age-earnings profile for natives and the foreign-born, Cross-sections by year**

	Dependent variable = Occupation score					
	1900		1910		1920	
0-5 yrs in US	-1.208 (0.196)	-1.553 (0.254)	-1.106 (0.101)	-1.697 (0.148)	-1.330 (0.295)	-2.019 (0.313)
6-10 yrs in US	-0.104 (0.164)	-0.399 (0.260)	-0.500 (0.127)	-1.022 (0.167)	-1.168 (0.126)	-2.375 (0.161)
11-20 yrs in US	0.258 (0.114)	0.153 (0.253)	0.472 (0.122)	0.027 (0.171)	-0.045 (0.101)	-1.081 (0.140)
21-30 yrs in US	0.485 (0.181)	0.428 (0.296)	0.411 (0.122)	0.172 (0.187)	0.707 (0.155)	-0.189 (0.191)
30+ yrs in US	0.591 (0.215)	0.401 (0.325)	0.077 (0.211)	-0.245 (0.260)	0.695 (0.159)	-0.117 (0.215)
Age	0.383 (0.009)	0.384 (0.009)	0.359 (0.008)	0.361 (0.008)	0.337 (0.008)	0.337 (0.008)
Age > 35	14.263 (0.420)	14.317 (0.419)	13.345 (0.358)	13.249 (0.358)	12.443 (0.345)	12.504 (0.345)
Age * Age > 35	-0.441 (0.011)	-0.443 (0.011)	-0.409 (0.010)	-0.407 (0.010)	-0.385 (0.009)	-0.386 (0.009)
Constant	12.153 (0.228)	12.118 (0.228)	13.697 (0.198)	13.665 (0.198)	15.317 (0.200)	15.265 (0.199)
Country FE?	N	Y	N	Y	N	Y
<i>N</i>	119,538	159,092	169,296	119,538	159,092	169,296

Notes: IPUMS, men aged 18-55, occupation score non-zero. Foreign born includes same set of countries in matched sample (see Table 1). For columns 2, 4 and 6, omitted country = Italy.

**Table 5: OLS estimates, Age-earnings profile for natives and the foreign-born, 1900-1920**

Dependent variable = Occupation score			
	Cross 1900-20	Repeated cross	Panel
0-5 yrs in US	-1.413 (0.264)	-0.462 (0.318)	0.741 (0.730)
6-10 yrs US	-0.799 (0.234)	-0.008 (0.277)	0.909 (0.637)
11-20 yrs US	-0.308 (0.194)	0.230 (0.219)	1.028 (0.489)
21-30 yrs US	-0.255 (0.203)	0.142 (0.216)	0.546 (0.496)
30 yrs in US	-0.185 (0.235)	-0.037 (0.237)	0.243 (0.542)
Year = 1910	0.909 (0.082)	0.964 (0.082)	1.488 (0.173)
Year = 1920	1.746 (0.117)	1.833 (0.118)	3.499 (0.235)
Arrive 1891+	---	-0.842 (0.158)	-0.304 (0.361)
Age	0.368 (0.009)	0.366 (0.009)	0.315 (0.017)
Age > 35	13.772 (0.397)	13.714 (0.398)	14.272 (0.784)
Age x (Age >35)	-0.419 (0.011)	-0.417 (0.011)	-0.440 (0.022)
Denmark	-2.365 (0.388)	-2.518 (0.389)	-2.387 (1.197)
Finland	-2.004 (0.475)	-2.040 (0.475)	-4.143 (1.423)

Norway	-2.673 (0.308)	-2.838 (0.314)	-3.957 (0.909)
England	2.551 (0.249)	2.344 (0.252)	2.061 (0.559)
Scotland	2.922 (0.377)	2.714 (0.379)	2.515 (1.113)
Wales	2.863 (0.586)	2.656 (0.586)	1.343 (1.924)
Belgium	-2.180 (0.773)	-2.257 (0.773)	-1.828 (2.113)
France	0.409 (0.515)	0.288 (0.515)	1.132 (1.406)
Switzerland	-2.269 (0.463)	-2.445 (0.463)	-2.427 (1.496)
Portugal	-3.745 (0.652)	-3.818 (0.652)	-5.059 (1.521)
Austria	0.787 (0.269)	0.732 (0.270)	0.085 (0.600)
Germany	0.236 (0.205)	0.022 (0.209)	-0.138 (0.492)
Russia	3.266 (0.245)	3.223 (0.245)	2.818 (0.492)
Constant	12.523 (0.225)	12.553 (0.225)	13.681 (0.441)
<i>N</i>	199,906	199,906	53,115

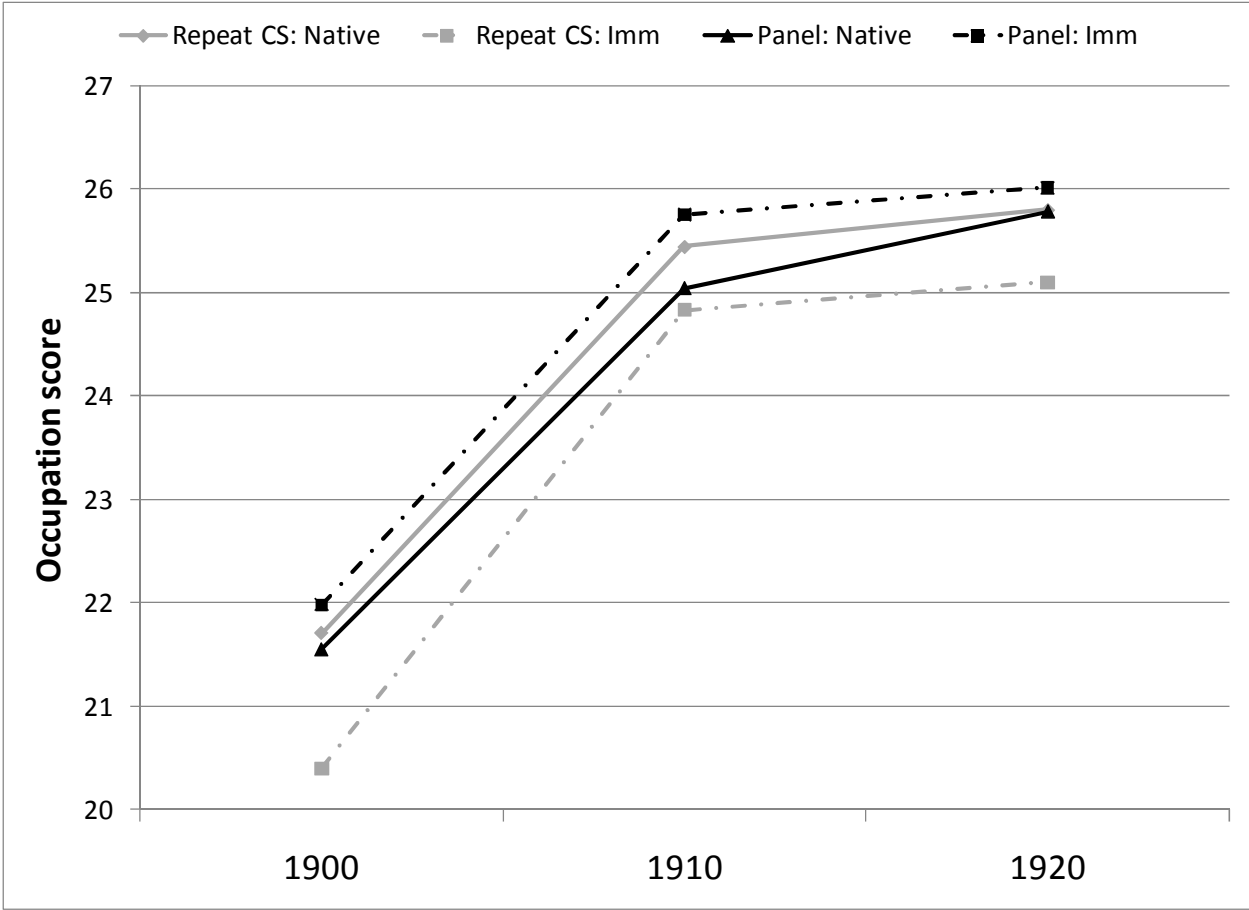
Notes: Sample restricted to men age 18-35 in 1900; 28-45 in 1910 and 38-55 in 1920 and, for the foreign born, men who arrived in the US in or before 1900. Excludes southern states and men for whom occupation score equals zero. Columns (1) and (2) use pooled IPUMS data; column (3) uses matched panel sample.

**Table 6: Robustness for age-earnings profile in panel sample, 1900-1920**

	Drop child immigrants			4 arrival cohorts		
	CS	RCS	Panel	CS	RCS	Panel
0-5 years in US	-1.518 (0.267)	-0.665 (0.333)	0.559 (0.871)	-1.437 (0.263)	-0.182 (0.353)	1.049 (0.837)
6-10 yrs in US	-0.936 (0.242)	-0.217 (0.289)	0.598 (0.773)	-0.778 (0.234)	-0.095 (0.306)	0.402 (0.706)
11-20 yrs in US	-0.572 (0.206)	-0.034 (0.237)	0.762 (0.618)	-0.309 (0.194)	0.229 (0.245)	0.842 (0.549)
21-30 yrs in US	-0.306 (0.218)	0.119 (0.237)	0.578 (0.647)	-0.249 (0.203)	0.119 (0.236)	0.377 (0.548)
30+ yrs in US	-0.006 (0.290)	0.113 (0.291)	0.603 (0.779)	-0.188 (0.235)	-0.021 (0.245)	0.164 (0.568)
Arrive 1886-90					0.145 (0.171)	0.594 (0.391)
Arrive 1891-95					-0.480 (0.202)	0.372 (0.459)
Arr. 1896-1900					-1.218 (0.225)	-0.758 (0.576)
<i>N</i>	193,066	193,066	39,831	199,906	199,906	53,115
	Country x arrival cohort FE			ln(occupation score)		
	CS	RCS	Panel	CS	RCS	Panel
0-5 years in US	-1.437 (0.263)	-0.124 (0.407)	1.223 (0.904)	0.005 (0.011)	0.042 (0.013)	0.122 (0.029)
6-10 yrs in US	-0.778 (0.234)	0.341 (0.361)	1.259 (0.811)	0.032 (0.009)	0.063 (0.011)	0.108 (0.025)
11-20 yrs in US	-0.309 (0.194)	0.600 (0.323)	1.462 (0.727)	0.044 (0.008)	0.065 (0.009)	0.101 (0.020)
21-30 yrs in US	-0.249 (0.203)	0.515 (0.321)	0.974 (0.727)	0.036 (0.009)	0.052 (0.009)	0.072 (0.019)
30+ yrs in US	-0.188 (0.235)	0.331 (0.338)	0.685 (0.759)	0.032 (0.010)	0.038 (0.009)	0.051 (0.021)
<i>N</i>	199,906	199,906	53,115	199,906	199,906	53,115

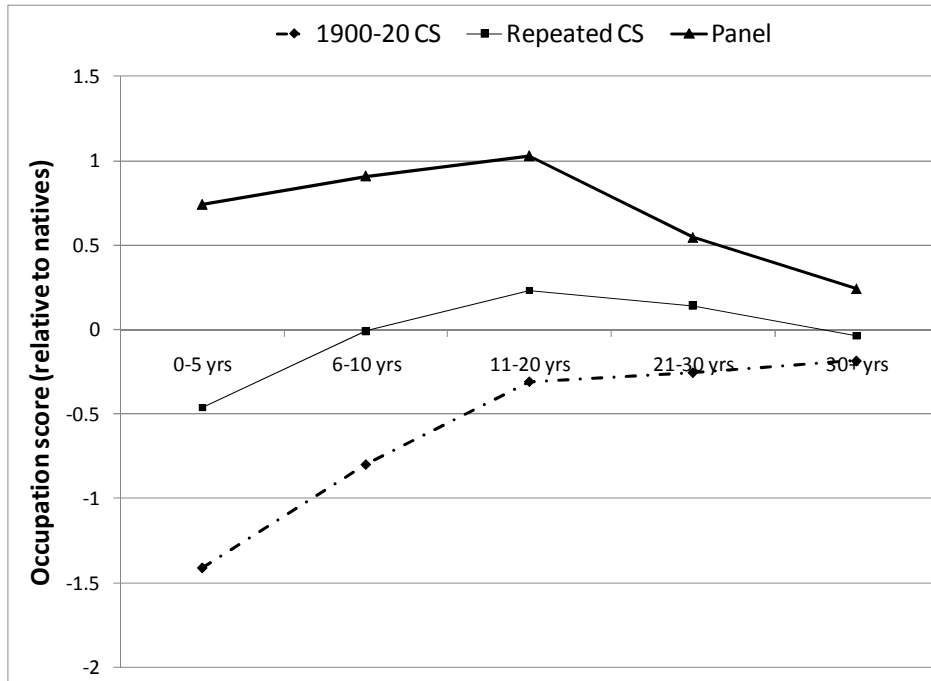
See notes to Table 5 for sample restrictions. All regressions include country fixed effects. Panel A drops immigrants who arrived in the US before age 10 or after age 40. In Panel B, the omitted arrival cohort is 1896-1900. In Panel C, we interact country-of-origin with a single arrival cohort dummy (arrived after 1890).

**Figure 1: Estimated occupation score for immigrant and native-born workers in repeated cross-section and panel data, 1900-1920**

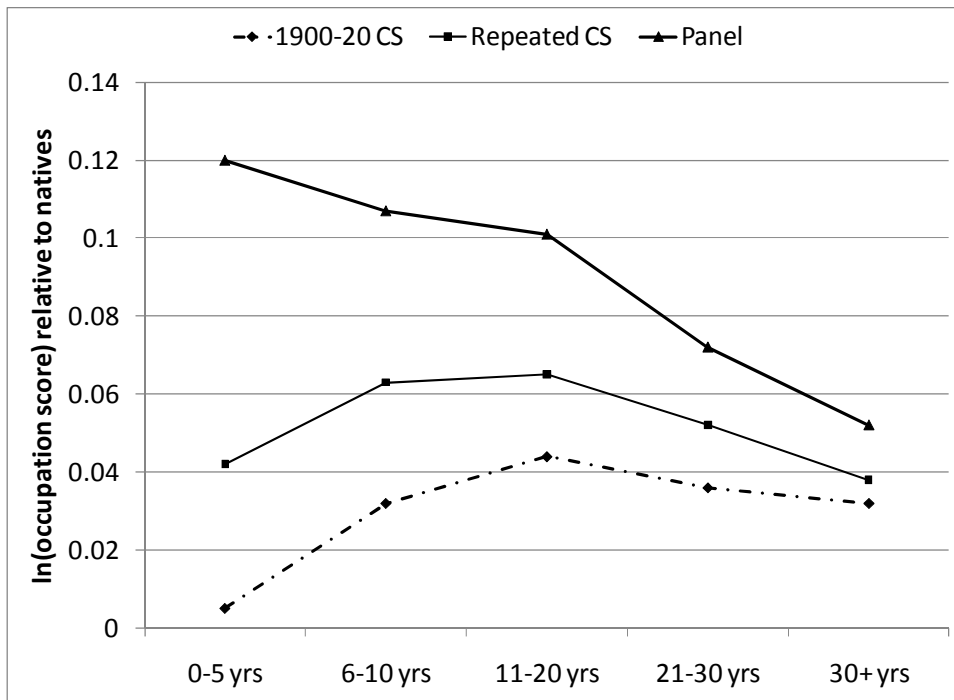


**Figure 2: Convergence in occupation score between immigrants and native-born workers by time spent in the US, cross-sectional and panel data, 1900-1920**

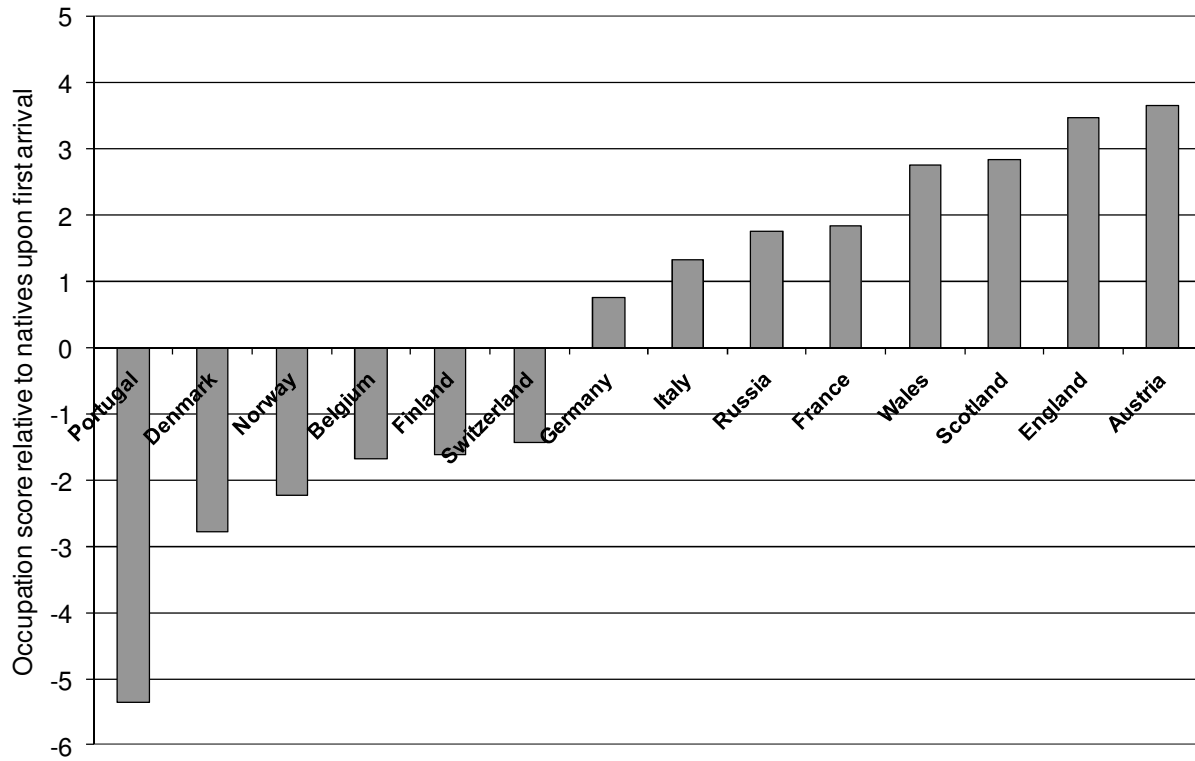
**a. Occupation score**



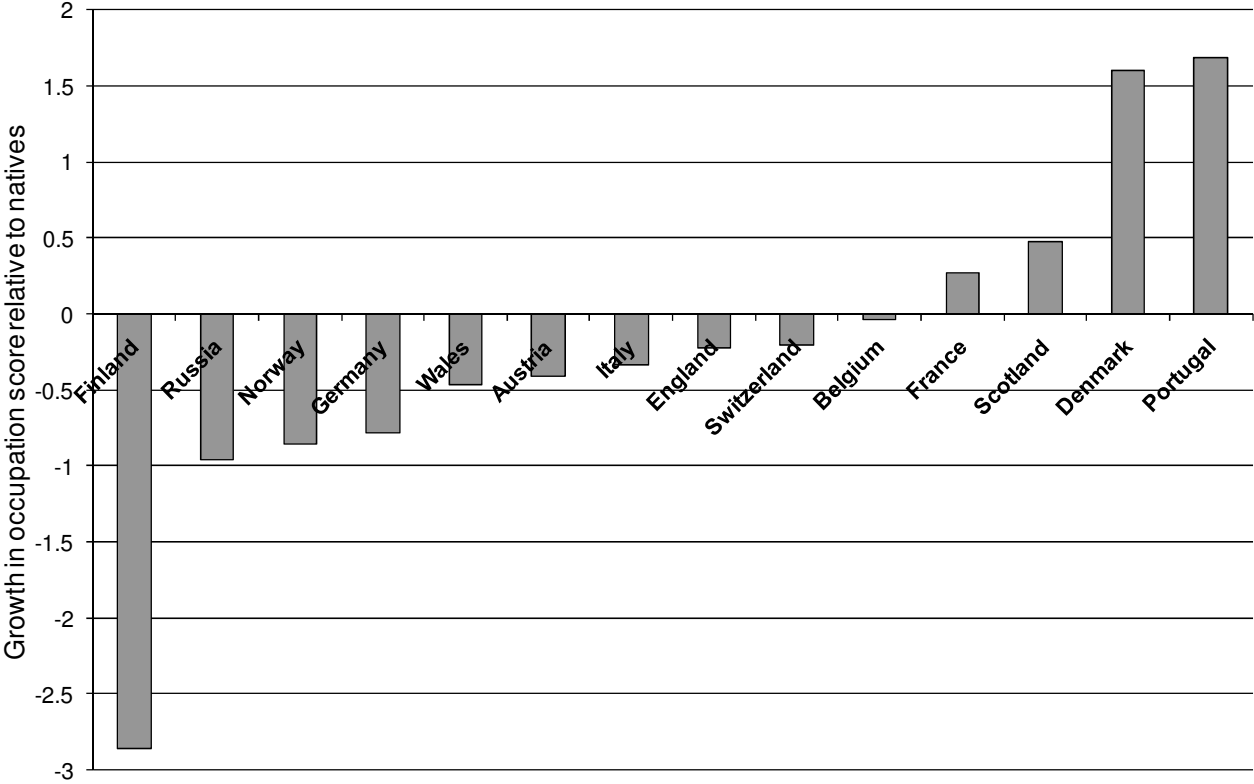
**b. ln(occupation score)**



**Figure 3: Occupation score. Gap between immigrants and natives upon first arrival in the US, by country of origin**



**Figure 4: Convergence in occupation score between immigrants and natives. New arrivals (0-5 years in US) versus longstanding immigrants (30 years in US), by country of origin**



**Figure 5: Implied selectivity of return migrants. Difference between estimated convergence in panel and repeated cross-section data, by country of origin**

