

**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

August 2012

Abstract: During the Age of Mass Migration (1850-1913), the US maintained an open border and absorbed 30 million European immigrants. Prior cross-sectional 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, we show that, in fact, immigrants did not face a substantial initial earnings penalty and experienced occupational advancement at the same rate as natives. Cross-sectional patterns are driven by biases from declining arrival cohort quality and departures of negatively-selected return migrants. We show that these findings vary substantially across sending countries and explore potential mechanisms.

JEL Code: J61, N30

Keywords: Migration, return migration, assimilation, selection

* We are grateful for the access to Census manuscripts provided by Ancestry.com. 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, the Economic History Association, and the NBER Development of the American Economy Summer Institute. We also thank participants of seminars at Berkeley, Caltech, Duke, Hebrew University, Northwestern, Norwegian School of Economics, Tel Aviv, UC-Davis, UCLA, and UT-Austin. We benefited from conversations with Manuel Amador, Attila Ambrus, Pat Bayer, Doug Bernheim, Tim Bresnahan, Marianne Bertrand, David Card, Greg Clark, Dora Costa, Pascaline Dupas, Liran Einav, Joseph Ferrie, Erica Field, Doireann Fitzgerald, Bob Gordon, Avner Greif, Hilary Hoynes, Nir Jaimovich, Pete Klenow, Pablo Kurlat, Aprajit Mahajan, Daniel McGarry, Roy Mill, Joel Mokyr, Jean-Laurent Rosenthal, Seth Sanders, Izi Sin, Yannay Spitzer, Gui Woolston, Gavin Wright, and members of the UCLA KALER group. Roy Mill helped with data collection. 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.

I. Introduction

This paper assembles new panel data to study the assimilation of European migrants in the US labor market in the early twentieth century and the selection of return migrants from this migrant pool. Our findings challenge the conventional wisdom about immigrant assimilation in the Age of Mass Migration (1850-1913). Prior studies, which were based on cross-sectional data, found that immigrants held substantially lower-paid occupations than natives upon first arrival but experienced rapid convergence with natives over time. In contrast, the typical immigrant in our panel data did not face a large initial earnings penalty and moved up the occupational ladder at the same rate as natives.¹ Comparing assimilation patterns in the panel and cross sectional data, we conclude that the apparent convergence in the cross section is driven by a decline in the productivity of immigrant arrival cohorts over time and the departure of negatively-selected return migrants.

Gaining an accurate view of assimilation and selection of return migrants in the Age of Mass Migration is important for our understanding of both the past and the present. First, this era was one of the largest migration episodes in modern history, sizeable enough to influence labor supply and economic development on both sides of the Atlantic. European countries lost up to a third of their population through emigration and, by 1910, 22 percent of the US labor force was foreign born (compared with 17 percent today). Second, the US maintained an open border policy for European migrants in this period and had yet to develop a comprehensive welfare state, allowing us to assess how the processes of labor market assimilation would look like in the

¹ We follow the historical literature and focus on occupation-based earnings because individual earnings were not recorded in population Censuses before the mid-twentieth century. We thus match each individual's recorded occupation to the median earnings in his occupation in 1950. We note that this unavoidable reliance on occupation-based earnings suggests an important limitation: our measure only captures convergence between occupations and cannot speak to within-occupation income convergence.

absence of immigrant selection policies or government support. Finally, beliefs about immigrant assimilation during this period shaped migration policies that are still with us today. The widespread view that migrants from eastern and southern Europe could not assimilate triggered restrictive migration quotas, first passed in the 1920s, that select migrants on the basis of nationality and skill.

Our current understanding of assimilation in the Age of Mass Migration is based on cross-sectional data (Blau 1980; Hatton 1997; Minns 2000). While these studies contributed greatly to our understanding of this era, we illustrate in Figure 1 how inferring assimilation from a cross section is subject to bias due both to differences in the productivity of arrival cohorts and to return migration. The figure normalizes natives' earnings to 100 in all years and depicts four types of migrants. Migrants A and B arrived to the US in 1895 and earned 100 and 80 respectively throughout their stay in the US, while migrants C and D arrived to the US in 1915 and earned 60 and 40 respectively. Furthermore, migrant B is assumed to return to his home country in 1909, illustrating one possible case of negatively-selected return migration.

Now imagine that the researcher only has a single cross section of data, say the 1920 Census. In this case, she will compare migrant A, who arrived in the US in 1895 and remained in the US in 1920, to migrants C and D, who arrived in 1915, and conclude that, upon arrival migrants faced earning penalty relative to natives but completely closed the earning gap after 25 years in the US. However, this conclusion would mistake a combination of arrival cohort differences and negatively-selected return migration for true migrant assimilation (this point was first made by Douglas, 1919 and was developed by Borjas, 1985).² The bias due to differences in

² In an early paper in this literature, Chiswick (1978) found that immigrants in the 1970 cross section experienced faster wage growth than the native-born and overtook natives within 15 years of arrival. Borjas (1985) demonstrated that, in this period, half of the apparent convergence in a cross section is driven by changes in cohort quality over time.

arrival cohorts can be avoided if the researcher has access to repeated cross sections, say the 1900 and 1920 Censuses. However, Figure 1 demonstrates that, even with repeated cross sections, inferences on migrant assimilation can still be biased by the process of return migration (this point was made in Jasso and Rosenzweig, 1988). In the 1900 census, the 1895 migrant arrival cohort contains both the high-skilled migrant A and the lower-skilled migrant B. By the 1920 census, migrant B had returned to his home country, leaving only migrant A behind. The apparent increase in wages over time in the 1895 arrival cohort is driven solely by a compositional change in the migrant pool rather than by the true assimilation of those who remained in the US.

Panel data, which follows individual migrants over time in the US, measures immigrant assimilation free from these two forms of bias (see: migrant A in Figure 1). Adopting this approach, we construct a large panel dataset of 24,000 native-born Americans and immigrants from 16 European 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.

Moreover, by contrasting the assimilation patterns in the repeated cross-section and panel data, we can infer the selection of return migrants relative to migrants who remained in the US long term. In particular, as is illustrated in the southeast corner of Figure 1, differences in the assimilation profile estimated from repeated cross-section and panel data are due to selective

return migration.³ Importantly, we are also able to compare this indirect approach, in which we infer the selection of return migrants, with a direct approach. Specifically, we compare return migrants in a special supplement to the 1910 Norwegian Census to Norwegian migrants who remain in the US; this direct evidence on the selection of return migrants validates our indirect approach.⁴

Over 25 percent of migrants returned to Europe during this era (Gould, 1980; Bandiera, Rasul and Viarengo, 2010). Return migrants may have been negatively selected, as in the example above, if migrants who were not successful in the US returned home. Furthermore, many migrants in this era employed a deliberate strategy of temporary migration to the New World (Piore, 1980; Wyman, 1996). These temporary migrants may have been negatively selected on observable characteristics if they remained in low-paid occupations during their short sojourn (Dustmann, 1993) or may have been positively selected if productive migrants reached their “target savings” goal faster (Galor and Stark, 1991).

Our comparison of assimilation patterns in cross-sectional and panel data reveals patterns akin to the hypothetical illustration in Figure 1. We find that immigrants in each cross section held lower-paid occupations than natives upon first arrival and appear to completely converge with natives over time. The occupation-based earnings gap between immigrants and natives is only half as large when using repeated, rather than single, cross sections, suggesting a decline in cohort quality. In the panel data, immigrants’ initial penalty disappears almost entirely

³ Differences in the assimilation profile in the repeated cross-sections and the panel could also be due to other sources of selective attrition, namely selective mortality or selective name changes. We discuss these possibilities in section III.D.

⁴ Our indirect approach of comparing panel and cross sectional data is similar to Lubotsky (2007), who 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. For other panel analyses of immigrant assimilation in the contemporary period, 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). Zakharenko (2008) provides descriptive evidence that return migrants leaving the US are negatively selected.

(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.

Our analysis is also the first to reveal a large degree of heterogeneity in the assimilation and selection of return migrants across sending countries. We show that immigrants 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. Yet, regardless of starting position, immigrants from all countries experience occupational upgrading at the same rate as natives, thereby preserving the initial gaps between immigrants and natives over time. We explore potential mechanisms for variation in immigrant performance and find suggestive evidence that immigrants from countries that were more economically developed or more culturally similar to the US are those who were most successful in the US labor market.

Finally, we study how the children of immigrants who came during this era performed in the US labor market. Occupational convergence between immigrants and natives may take more than one generation. 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 persistence in occupations across generations rather than extensive assimilation in the second generation: when migrants from a certain sending country outperformed US natives, so did second generation migrants, and vice versa.

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 on immigrant assimilation and the selection of return migrants. We estimate the occupation-based earnings penalty (or premium) as well as the earnings distribution for the typical immigrant. Section 5 contains direct evidence on the selection of return migrants from the 1910 Norwegian Census. In Section 6, we show country-by-country results on assimilation and return migration consider possible mechanisms for the cross-country variation in immigrant performance. Section 7 analyzes the performance of second generation immigrants relative to their parents, Section 8 provides a robustness analysis for our main findings, and Section 9 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).⁵ 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.

⁵ 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.⁶

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 short... the acquisition of the English language will be of little consequence...

⁶ Over 70 percent of German immigrants were literate as early as 1850.

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.⁷ 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).

A second generation of scholarship examined individual-level wage data from surveys conducted by State Labor Bureaus (Hannon, 1982; Eichengreen and Gemery, 1986; Hanes,

⁷ 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 sociological literature on immigrant assimilation.

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.⁸

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.⁹ 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 of the earnings gap with natives after 15 years in the US. However, all these analyses compare

⁸ 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.

⁹ 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.

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.¹⁰

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, we exclude native-born men residing in a southern state and all black natives regardless of place of residence.¹¹

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

¹⁰ We note that Minns (2000) acknowledges the potential bias from changes in the quality of immigrant arrival cohorts. 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).

¹¹ 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.

relevant age range in 1900 from the genealogy website Ancestry.com. Altogether, we identify immigrants from 16 sending countries.¹²

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 2 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:

- (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.¹³
- (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

¹² 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.

¹³ 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.

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 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.¹⁴ 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

¹⁴ 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. Our final sample has 1,426 native-born men and 18,249 immigrants with non-missing occupation data.

native born were more likely to be salesmen and clerks, two occupations with high returns to fluency in English. Other common occupations in both groups include managers, operatives, and general laborers.¹⁵

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.¹⁶ A further concern with the IPUMS ‘occupation score’ variable is its anchoring to occupation-based earnings in the year 1950. The 1940s and 1950s was 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.

¹⁵ 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.

¹⁶ 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.¹⁷ 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.

¹⁷ 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-sections, which is arguably mostly due to selective return migration.¹⁸ 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 a likely 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). Furthermore, 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; 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 a non-issue (at least for the native born).¹⁹ 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.

¹⁸ 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.

¹⁹ 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).²⁰ 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. We estimate:

$$Occupation_score_{ijmt} = \gamma_{t-m} + \lambda_m + \eta_t + \alpha_j + \beta_1 Age_{it} + \beta_2 I[Age_{it} \geq 35] + \beta_3 Age_{it} I[Age_{it} \geq 35] + \varepsilon_{ijmt} \quad (1)$$

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).

²⁰ 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.²¹ *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 1 includes a dummy variable for each time interval, with the native born constituting the omitted category.²² 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 a cross-section version of equation 1 with data from the 1900, 1910 and 1920 IPUMS and Census year fixed effects (results are similar when looking at each single cross-section separately, see online Appendix B). We then add the arrival cohort dummy (λ_m) to estimate the repeated cross-section model. The coefficient on the arrival cohort dummy

²¹ 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.

²² 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).

indicates changes in cohort quality over time. Comparing 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.²³ 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.

We emphasize that measuring the assimilation process requires a panel dataset that follows individuals 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

²³ 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).

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 observe occupational outcomes for the typical migrant in the US labor market, 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-section and panel data

In this section, we estimate equation 1 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 somewhat 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.

Table 4 presents estimates of equation 1 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 3. 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.

In section VIII, we explore a series of alternative specifications and measures of occupation-based earnings. Overall, we show that specification choice has little effect on either

the estimated rate of convergence or the implied selection of return migrants. Adjusting the occupation-based earnings measure increases the initial earnings gap that immigrants face, sometimes substantially, but has little effect on either estimated convergence or selection of return migrants.

C. Earnings distribution of natives and immigrants

On average, immigrants earn less than natives upon first arrival in the cross-section. Table 5 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.

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 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 below).

VI. Heterogeneity by sending country

A. Variation in convergence and selectivity of return migration

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 4 illustrates cross-country variation in the occupation-based earnings of immigrants relative to the native born, both upon first arrival and after 30 or more years in the US. The grey bars indicate that 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 We explore possible explanations for these cross-country differences in Section VI.

The black bars in Figure 4 demonstrate that, on the whole, permanent immigrants experience little occupational growth relative to natives after spending time in the US. That is, permanent immigrants move up the occupational ladder 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.

²⁴ Consistent with Table 7, 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).

Immigrants 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.

The average immigrant in the cross-section and panel samples differ both because of declines in arrival cohort quality and negatively-selected return migration. Figures 5 and 6 report variation in the direction and magnitude of these two biases by country-of-origin. We begin by estimating a version of equation 1 with four arrival cohorts (see Table 7, 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 this 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.²⁵

²⁵ The height of the bars in Figure 6 represent the product of the selectivity of return migrants and the return migration rate. We use return migration rates by country reported either in Gould (1980) or in Bandiera, Rasul and Viarengo (2011) to back out the actual selection term. Gould (1980) reports return migration rates for Russian Jews and non-Jews separately (7.1 percent and 87 percent); we use the weighted average. 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 in both cases (not shown). The one exception is that return migrants to Russia look even more negatively selected when we use the Bandiera, et al. (2011) return migration rates.

Russia is a particularly interesting case. Figure 4 shows that Russian migrants performed well in the US upon first arrival and Figure 6 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.

B. Explaining cross-country variation in immigrant performance

Figure 4 documents 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 6.

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 market

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.²⁶ 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:

$$Y_{it} = \left\{ \begin{array}{l} \alpha_i + \beta_a Age_{it} + \beta_{a2} Age_{it}^2 + \beta_{a3} Age_{it}^3 + \beta_{a4} Age_{it}^4 + \\ Migrant_i * \sum_k \gamma_k YearsIUS_{itk} + \varepsilon_{it} \end{array} \right\} \quad (2)$$

As before, our outcome variable is occupation-based earnings converted to 2010 dollars. In Figure 7, we illustrate the results from equation 2 for a person who is 25 years old in either 1910

²⁶ 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.

(first generation versus natives) or in 1930 (second generation versus natives). We assume the first generation migrant moved to the US in 1890.²⁷

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.

VIII. Alternative specifications and earnings measures

We return in this section to our main analysis, which compared the age-earnings profile of immigrants and natives from 1900 to 1920. Table 7 assesses the sensitivity of our findings to a series of alternative specifications and substitute measures of occupation-based earnings.²⁸ The first section of Table 7 considers alternative specifications for equation 1. 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, permanent immigrants fare even better, earning \$600 more than natives upon first arrival. This modification does not alter the comparison of permanent and temporary immigrants nor the degree of convergence observed 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

²⁷ Results are robust to alternative specifications of the age effects.

²⁸ 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 3.

eliminate the initial earnings gap between migrants and natives in the repeated cross-section; yet, permanent immigrants continue to earn \$600 more than typical migrant upon first arrival. In Panel C, we interact the country-of-origin fixed effects with the arrival cohort dummy. The premium earned by permanent immigrants is preserved.

The next section of Table 7 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, again consistent with negatively-selected return migration. 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.

Our main dependent variable is based on occupation-based earnings in 1950. In that year, farmers earned incomes below the median, whereas, in 1900, 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. Because natives were more likely than immigrants to be owner-occupier farmers, this modification increases the immigrant earnings penalty. However, the degree of convergence and the comparison between immigrants in the cross-section and panel are unchanged. Panel F replaces the 1950-based earnings measure with mean earnings by occupation from the 1900 Cost of Living survey.²⁹ Given the greater income inequality in 1900

²⁹ We note that this measure has several disadvantages relative to our main measure based on the 1950 occupation score. 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 have to infer from the US Census of Agriculture).

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, suggesting that return migrants were negatively selected.

The final section of Table 7 presents estimates of equation 1 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 10.³⁰ 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). 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

³⁰ We choose the age of 10 because it is an age at which most people did not work, even in this historical period. Results are similar at cutoffs of age 12 or 14 as well.

permanent immigrants and the total migrant pool, consistent with negative selection of return migrants, as well as a similar amount of convergence between immigrants and natives.³¹

IX. Conclusion

We construct a new panel dataset of native- and foreign-born men in the US labor market at the turn of the twentieth century, an era in which US borders were open to all European migrants. This Age of Mass Migration is not only of interest in itself, as one of the largest migration waves in modern history, but is also informative about the process of immigrant assimilation in a world without migration restrictions. Most of the previous research on this era relies on a single cross-section of data and finds that immigrants started with lower-paid occupations than natives but caught up with natives after spending some time in the US.

In our panel dataset, we instead find that immigrants who remained in the US did not hold lower-paid occupations than US natives, even upon first arrival, and moved up the occupational ladder at the same rate as natives. We conclude that the apparent convergence in a single cross-section reflects a substantial decline in the quality of migrant cohorts over this period as well as a change in composition of the migrant pool as negatively-selected return migrants leave the US over time. Our paper further demonstrates the importance of accounting for differences in migration patterns across sending countries. Permanent migrants from countries with cultural proximity to the US performed far better than natives upon first arrival, while permanent migrants from other countries performed far worse. Yet immigrants from all countries, regardless of their starting position, experienced little occupational convergence with

³¹ 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.

natives. Moreover, we find that these differences in performance across sending countries are persistent across generations.

Contemporaries questioned the ability of European immigrants to assimilate in the US economy and called for strict migration restrictions that favored countries with highly-skilled residents. Our results indicate that these concerns were unfounded: the average permanent immigrant in this era arrived with skills similar to those of natives and experienced identical rates of occupational upgrading over their lifecycle. These successful outcomes suggest that migration restrictions are not necessary to ensure migrant assimilation. At the same time, we also note that migrants that arrived with low skill levels did not manage to close their skill gap with natives over time. This finding undercuts the commonly-held view that, unlike today's migrants, past waves of European immigrants, even those who arrived without the ability to read or to speak English, were able to quickly catch up with natives.

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

- C. Dustmann, Earnings Adjustment of Temporary Migrants. *Journal of Population Economics* **6** (1993), pp. 153-168.
- 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_immormpm.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. 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.
- 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.
- M. Piore, *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.	%	Mean income	Occupation	Freq.	%	Mean income
1.	Farmer	352	24.82	\$12,702	Farmer	3,301	18.09	\$12,702
2.	Manager	129	9.10	36,991	Manager	1,999	10.95	36,991
3.	Laborer	117	8.25	18,070	Laborer	1,791	9.81	18,070
4.	Salesman	75	5.28	22,253	Operative	1,102	6.04	20,931
5.	Operative	71	5.00	20,931	Foreman	603	3.30	34,194
6.	Clerical	45	3.17	22,579	Mine operative	596	3.27	21,854
7.	Carpenter	45	3.17	21,728	Machinist	578	3.17	28,961
8.	Machinist	45	3.17	28,961	Carpenter	529	2.90	21,728
9.	Farm laborer	39	2.75	12,674	Salesman	495	2.71	22,253
10.	Foreman	27	1.90	34,194	Clerical	326	1.79	22,579
Total (top 10)		945	66.61	\$20,460		11,320	62.03	\$22,542
Outside top 10		481	33.39	\$29,015		6,929	37.97	\$25,258

Notes: See notes to Table 1 for sample restrictions. Occupations based on 'OCC1950' IPUMS variable. Mean incomes reported in 2010 dollars.

**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 (301.546)	0.010 (0.013)
Foreign born	\$23,471	368.75 (127.42)	0.024 (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

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

Notes: See Table 1 notes for sample restrictions. Columns report coefficients from estimation of equation 1. 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: Occupation-based earnings distribution, 1900-20
in 2010 dollars**

	Cross-section		Panel	
	Immigrants	Natives	Immigrants	Natives
10 th	\$9,900	\$8,100	\$11,700	\$8,100
25 th	\$18,000	\$12,600	\$18,000	\$12,600
50 th	\$20,700	\$20,700	\$20,700	\$20,700
75 th	\$22,500	\$25,200	\$22,500	\$25,200
90 th	\$28,800	\$34,200	\$28,800	\$34,200
99 th	\$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 6: 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).

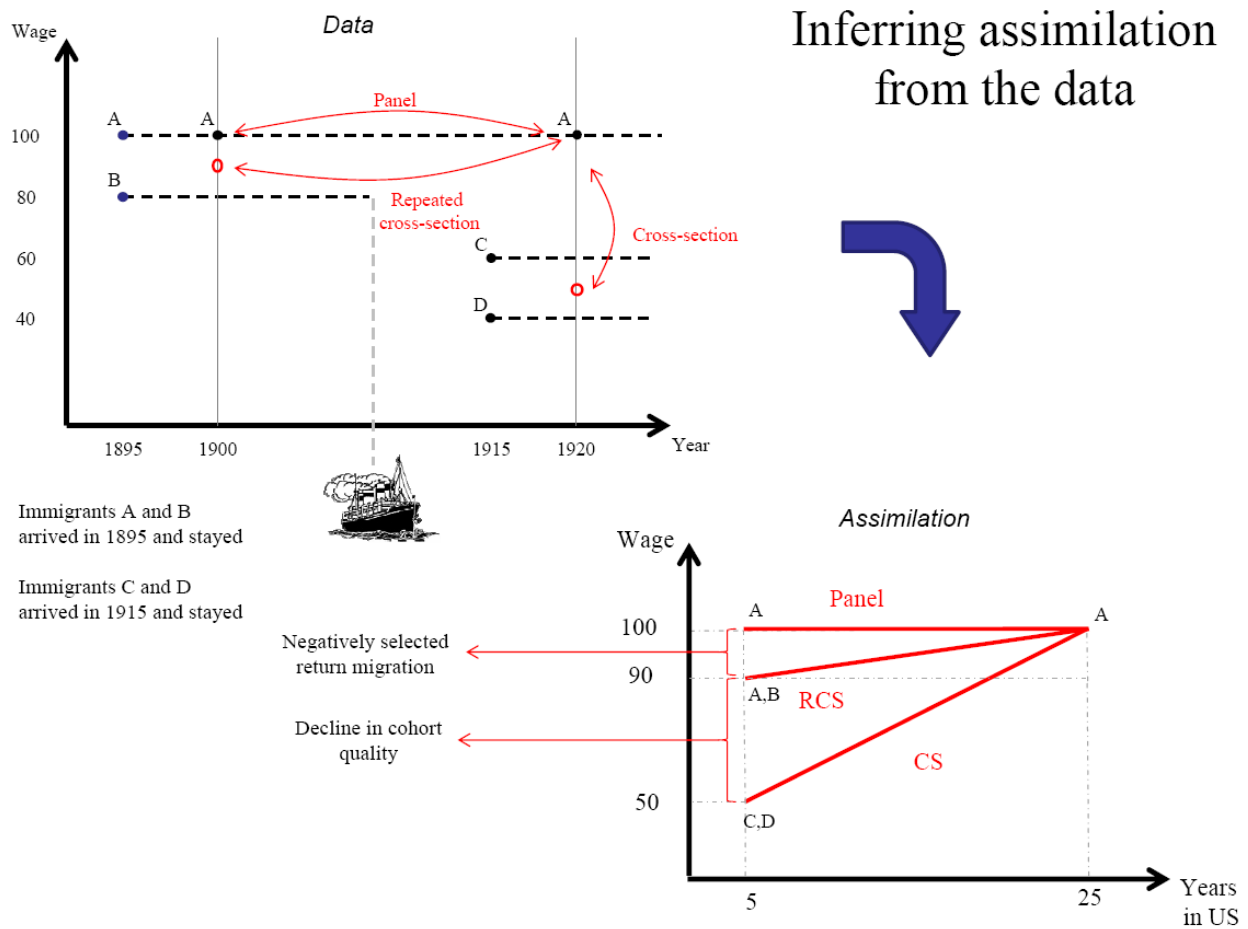
*** Column 5 adds the cultural distance and religious similarity measures to each regression. In “cultural distance” row, we only add religious similarity and in “religious similarity” row we only add cultural distance.

Table 7: Robustness for age-earnings profile in panel sample, 1900-1920

	<i>A. Without country FE</i>		<i>B. 4 arrival cohorts</i>		<i>C. Country x cohort FE</i>	
	<i>RCS</i>	<i>Panel</i>	<i>RCS</i>	<i>Panel</i>	<i>RCS</i>	<i>Panel</i>
0-5 years in US	-132.53 (169.89)	632.18 (214.37)	44.52 (219.02)	588.60 (262.28)	-6.806 (247.55)	640.95 (277.98)
6-10 yrs in US	362.45 (145.07)	736.20 (180.92)	86.587 (196.75)	399.58 (226.47)	332.24 (228.31)	702.95 (250.96)
11-20 yrs in US	474.73 (92.613)	569.06 (110.09)	226.58 (155.93)	347.85 (167.73)	436.95 (203.68)	586.86 (211.28)
21-30 yrs in US	436.39 (100.04)	359.30 (108.00)	196.66 (157.90)	136.77 (165.04)	417.65 (206.22)	387.28 (211.53)
30+ yrs in US	300.30 (148.23)	398.08 (147.04)	184.00 (187.25)	198.59 (189.01)	414.38 (236.21)	457.35 (239.64)
	<i>D. ln(occupation score)</i>		<i>E. Raise farmer income</i>		<i>F. 1900 income</i>	
	<i>RCS</i>	<i>Panel</i>	<i>RCS</i>	<i>Panel</i>	<i>RCS</i>	<i>Panel</i>
0-5 years in US	0.051 (0.010)	0.087 (0.011)	-670.49 (182.15)	26.20 (221.63)	-3241.10 (148.88)	-2717.00 (186.45)
6-10 yrs in US	0.066 (0.008)	0.083 (0.009)	-340.30 (166.45)	120.39 (196.15)	-2694.03 (142.32)	-2033.67 (164.90)
11-20 yrs in US	0.063 (0.006)	0.069 (0.007)	-248.11 (128.91)	-36.01 (140.41)	-2257.03 (113.37)	-1972.49 (119.66)
21-30 yrs in US	0.053 (0.006)	0.060 (0.006)	-227.43 (133.38)	-206.59 (139.12)	-2055.00 (115.48)	-1945.88 (119.79)
30+ yrs in US	0.044 (0.008)	0.057 (0.008)	-221.87 (170.33)	-129.70 (169.90)	-1828.41 (137.24)	-1755.42 (138.07)
	<i>G. Drop child migrants</i>		<i>H. State FE</i>		<i>I. State * urban FE</i>	
	<i>RCS</i>	<i>Panel</i>	<i>RCS</i>	<i>Panel</i>	<i>RCS</i>	<i>Panel</i>
0-5 years in US	-419.28 (191.28)	212.13 (232.69)	-1654.14 (198.71)	-1138.32 (301.55)	-2218.65 (197.41)	-1430.21 (301.41)
6-10 yrs in US	-95.30 (177.25)	259.10 (210.17)	-1285.28 (190.91)	-759.39 (257.23)	-2009.73 (189.70)	-1132.61 (257.31)
11-20 yrs in US	-50.80 (144.09)	67.75 (160.15)	-1193.63 (157.29)	-960.18 (201.16)	-1860.04 (156.93)	-1032.53 (199.08)
21-30 yrs in US	120.75 (148.59)	-81.71 (161.72)	-1073.68 (163.78)	-1139.49 (209.51)	-1659.34 (163.46)	-1111.17 (210.92)
30+ yrs in US	118.75 (201.37)	62.60 (209.79)	-1008.40 (196.08)	-577.14 (229.82)	-1539.02 (193.29)	-478.06 (230.63)

Notes to Table 7: 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. Sample size for Panels A-E is 262,248. Panel F has 264,338 observations; Panel G has 246,365; Panel H has 228,793 and Panel I has 227,930.

Figure 1: Inferences about immigrant assimilation in cross-sectional and panel data



Notes: The graph in the northwest corner depicts earnings for four hypothetical migrants. For illustrative purposes, we assume that natives earn 100 in every year. Migrants A and B arrived in 1895 and earn 100 and 80 respectively. Migrant B returns to his home country in 1909. Migrants C and D arrived in 1915 and earn 60 and 40 respectively. The graph in the southeast corner illustrates inferred assimilation profiles from a series of hypothetical datasets. The line labeled 'CS' refers to the 1920 cross section, wherein a researcher would compare migrants C and D to migrant A. The line labeled 'RCS' refers to two repeated cross sections (1900 and 1920), with which a researcher could compare migrants A and B in 1900 to migrant A in 1920. The line labeled 'Panel' refers to a panel data that follows migrant A over time.

Figure 2: 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 _____
(Insert name of township, town, precinct, district, or other civil division, as the case may be, in the instructions.)

Name of incorporated city, town, or village, within the above-named division, Scranton City
(Insert proper name and city, town, or village, or borough, in the instructions.)

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. <small>Enter surname first, then the given name and middle initial, if any. Exclude every person living on June 1, 1900. Omit children born since June 1, 1900.</small>	RELATION. <small>Relationship of each person to the head of the family.</small>	PERSONAL DESCRIPTION				NATIVITY.			CITIZENSHIP.			OCCUPATION, TRADE, OR PROFESSION <small>of each person TEN YEARS OF AGE AND OVER.</small>		
		Color or race.	DATE OF BIRTH. <small>Month. Year.</small>	Age at last birthday. <small>Whether single, married, widowed, or divorced.</small>	Place of birth of this Person.	Place of birth of Father of this person.	Place of birth of Mother of this person.	Year of immigration to this country. <small>Year of birth in this country.</small>	Number of years in the United States.	Married in this country.	Occupation.	Wages or salary received.		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
73 Morgan Williams	Head	W	M	Feb 1847	53	M	Wales	Wales	Wales	1857	11	W	Coal Miner	6
74 Sarah	Wife	W	F	Mar 1847	53	W	Wales	Wales	Wales	1857	11			
75 Mary A.	Daughter	W	F	May 1878	22	W	Wales	Wales	Wales	1887	11			
76 <u>Alexander James</u>	<u>Son in law</u>	W	M	Oct 1872	28	W	Wales	Wales	Wales	1887	11		Coal Miner	6
77 Elizabeth James	Boarder	W	M	Oct 1860	37	S	England	England	England	1890	6		Coal Miner	6

DEPARTMENT OF COMMERCE AND LABOR—BUREAU OF THE CENSUS

THIRTEENTH CENSUS OF THE UNITED STATES: 1910—POPULATION

STATE Pennsylvania
 COUNTY Lackawanna

TOWNSHIP OR OTHER DIVISION OF COUNTY Scranton City WARD OF CITY First
(Insert proper name and city, town, or village, or borough, in the instructions.)

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. <small>Enter surname first, then the given name and middle initial, if any. Exclude every person living on April 15, 1910. Omit children born since April 15, 1910.</small>	RELATION. <small>Relationship of this person to the head of the family.</small>	PERSONAL DESCRIPTION				NATIVITY.			CITIZENSHIP.		OCCUPATION.										
		Color or race.	DATE OF BIRTH. <small>Age at last birthday.</small>	Whether single, married, widowed, or divorced.	Place of birth of this Person.	Place of birth of Father of this person.	Place of birth of Mother of this person.	Year of immigration to this country. <small>Year of birth in this country.</small>	Whether naturalized in this country.	Trade, profession, or particular kind of work done by this person, or occupation, profession, laborer, etc.	General nature of industry, business, or establishment in which this person works, or on which he is employed, or on which he is working on any account.	Whether an employer, or working on any account.	Whether an employer or not.	Wages or salary received.	If an employer, value of work done.						
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
83 Morgan Williams	Head	M	M	Feb 1847	63	M	Wales	Wales	Wales	1857	NA		laborer	coal miner	W	70.0					
84 Sarah M.	Wife	W	F	Mar 1847	63	W	Wales	Wales	Wales	1857	NA		none								
85 Alexander James	Daughter	W	F	May 1878	32	W	Wales	Wales	Wales	1887	NA		none								
86 <u>James</u>	<u>Son in law</u>	M	M	Oct 1872	38	M	Wales	Wales	Wales	1887	NA		miner	coal miner	W	70.12					

DEPARTMENT OF COMMERCE—BUREAU OF THE CENSUS

FOURTEENTH CENSUS OF THE UNITED STATES: 1920—POPULATION

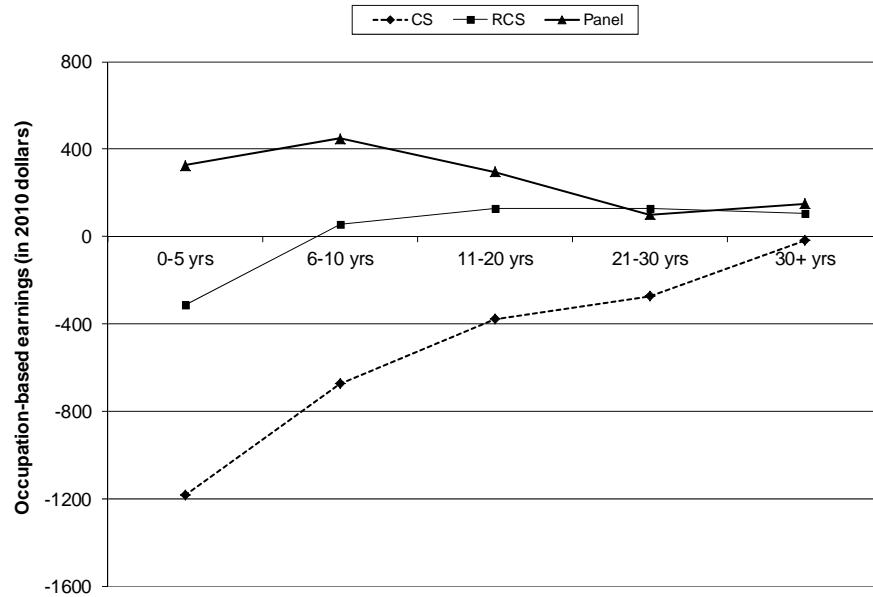
STATE Pennsylvania
 COUNTY Lackawanna

TOWNSHIP OR OTHER DIVISION OF COUNTY Scranton City WARD OF CITY 1st Ward 5th Dist.
(Insert proper name and city, town, or village, or borough, in the instructions.)

NAME OF INSTITUTION _____ ENUMERATED BY ME ON THE 20th DAY OF Jan, 1920, Thomas J. Keller

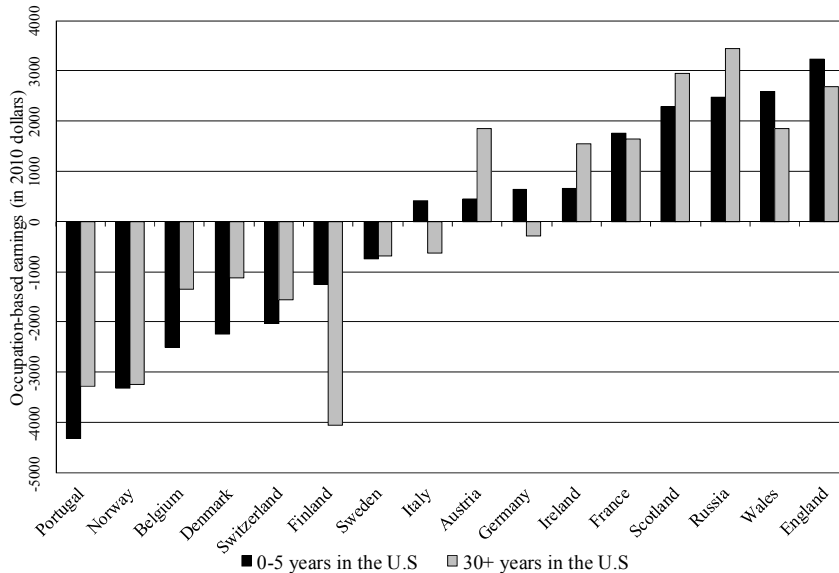
NAME of each person whose place of abode on January 1, 1920, was in this family. <small>Enter surname first, then the given name and middle initial, if any. Exclude every person living on January 1, 1920. Omit children born since January 1, 1920.</small>	RELATION. <small>Relationship of this person to the head of the family.</small>	SEX.	DATE OF BIRTH. <small>Month. Year.</small>	PERSONAL DESCRIPTION. <small>Whether single, married, widowed, or divorced.</small>	CITIZENSHIP.	EDUCATION.	NATIVITY AND MOTHER TONGUE.		OCCUPATION.												
							Place of birth of this Person.	Place of birth of Father of this person.	Trade, profession, or particular kind of work done, or occupation, profession, laborer, etc.	Industry, business, or establishment in which at work, or on which he is employed, or on which he is working on any account.	Whether an employer, or working on any account.	Wages or salary received.	If an employer, value of work done.								
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
88 <u>Margaret Sarah Jane</u>	<u>Head</u>	F	W	72	W	1848	U	Wales	Wales	Wales	1848	NA	none								
89 <u>Alexander James</u>	<u>Head</u>	M	W	48	M	1874	NA	Wales	Wales	Wales	1874	NA	Fireman	Scranton	W	14.00					
90 <u>Caroline</u>	<u>Wife</u>	F	W	42	M	1858	NA	Wales	Wales	Wales	1858	NA	none								
91 <u>William</u>	<u>Son</u>	M	W	17	S			Pennsylvania	Pennsylvania	Pennsylvania	1903	NA	none								

Figure 3: Convergence in occupation score between immigrants and native-born workers by time spent in the US, cross-sectional and panel data, 1900-1920



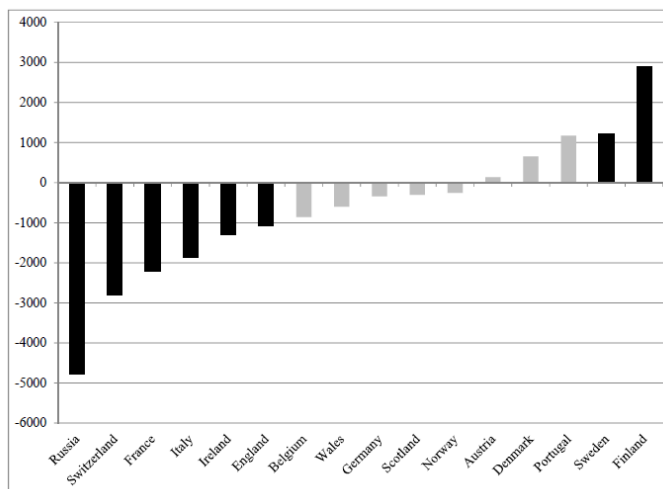
Notes: Plot of coefficients for “years spent in the US” indicators in equation 1. See Table 4 for coefficients and standard errors.

Figure 4: Initial earnings gap between native- and foreign-born in panel sample. Natives versus immigrants upon first arrival (0-5 years in US) and after time in the US (30+ years in US), by country of origin



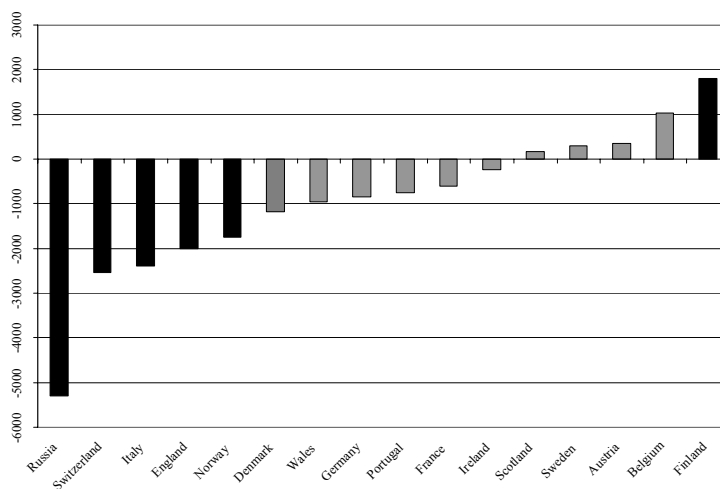
Notes: Graph reports coefficients on interaction between country-of-origin fixed effect and dummy variables for being in the US for 0-5 years or for 30+ years from regression of equation 1 in the panel sample. All coefficients for the 0-5 year interaction are significant except those for Austria, Germany, Ireland, Italy and Sweden. No differences between the 0-5 year and 30+ year coefficients are significant except that for Finland.

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



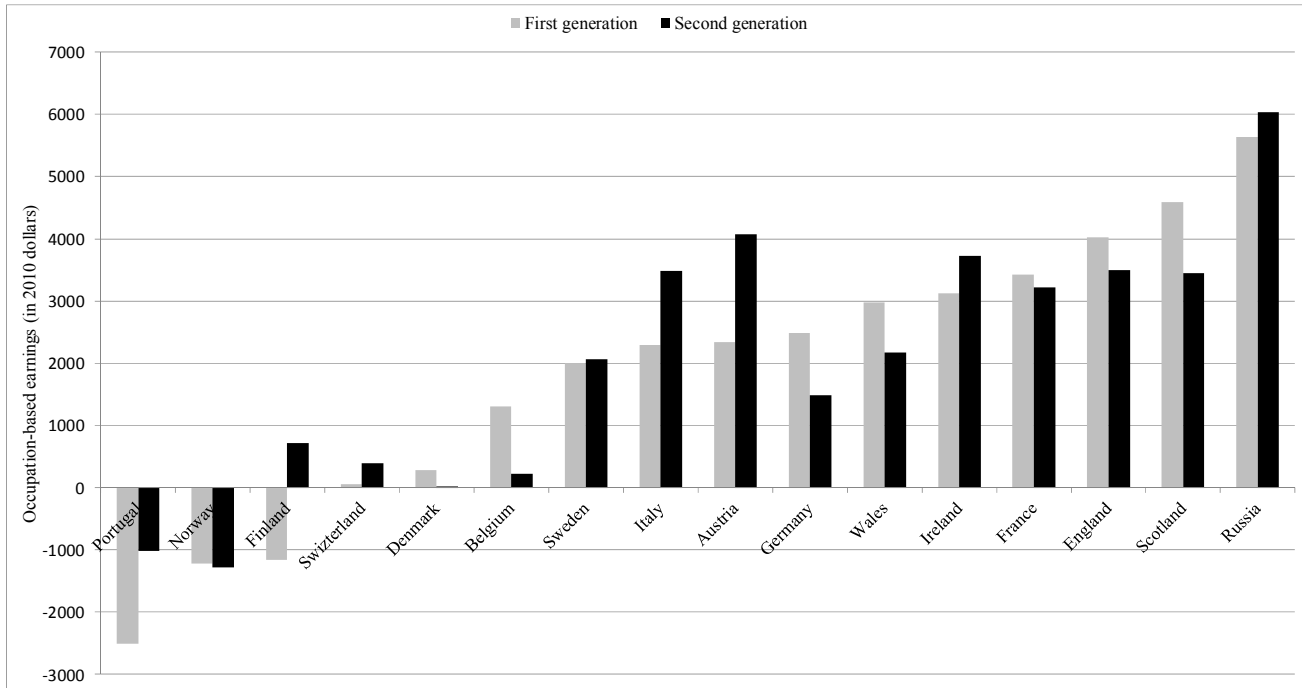
Notes: Estimates based on a version of equation 1 with four dummy variables for arrival cohorts in the panel sample (see Table 7, Panel B). The graph reports the difference between two coefficients: one interacts a country-of-origin fixed effect with the dummy variable for arriving in the US between 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 1 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 in equation 2 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.