

How do Existing Immigrants Respond to New Immigration? Evidence from the Age of Mass Migration.

Mike Feng*

September 15, 2026

Abstract

How do existing immigrants respond to new immigration? New arrivals may complement earlier immigrants or compete with them, especially when they share an origin. I study the effect of new immigration on the economic outcomes and assimilation of existing immigrants in the U.S. from 1900 to 1930 using a shift-share instrument and linked census records. Aggregate inflows have little effect on existing immigrants' occupational standing. Co-ethnic inflows, by contrast, lower incumbents' occupational rank and widen their gap with natives, with the largest losses where newcomers and incumbents hold similar occupations. Co-ethnic inflows also draw incumbents into the enclave, making them less likely to leave the county and more likely to give their children foreign names. These results highlight how the composition of new immigration shapes incumbents' labor market trajectories.

*I am grateful for the guidance and support from Brian Beach and Ariell Zimran. I thank Hanna Cao, Claire Chen, William Collins, Mika Inoue, Jonatan Moukh, Nathan Nunn, Marco Tabellini, João Tampellini, Reem Zaiour, and seminar participants at Vanderbilt University for helpful comments and discussions. All errors are my own.

1 Introduction

Whether immigrants assimilate economically is a central question in the study of immigration (Abramitzky et al., 2014; Collins & Zimran, 2023; Ward, 2025), and one that shapes native attitudes and debates over immigration policy (Alesina et al., 2023; Tabellini, 2020). Work on the determinants of assimilation has studied the traits of immigrants and the places where they settle (Abramitzky et al., 2020, 2025; Achard, 2022; Eriksson, 2020). A parallel literature asks how immigrants affect native labor market outcomes (Borjas, 2003; Card, 2001; Tabellini, 2020). However, both literatures largely overlook that every cohort of immigrants is followed by another, and that these later arrivals may affect those who came before them.

In this paper, I posit that these later arrivals change the labor market that existing immigrants face and alter their incentives to assimilate.¹ The direction of that change is ex ante ambiguous. Subsequent arrivals may complement existing immigrants, expanding local labor market demand and generating opportunities for advancement. They may also compete with incumbents in the labor market, especially when newcomers are from the same origin country, have similar skills, and enter labor markets through the same ethnic networks. Whether new immigration helps or hurts the assimilation trajectories of existing immigrants therefore depends on the composition of new immigration as well as its magnitude.

This paper studies how existing immigrants in the United States responded to new immigrant inflows between 1900 and 1930. Throughout, I make a distinction between the volume of new immigration and its composition. I first ask whether new arrivals raised or lowered existing immigrants' occupational rank and income scores, and whether the answer depends on the origins of the newcomers. Since natives were also exposed to the same inflow and made labor market gains (Tabellini, 2020), I ask whether any gains were large enough for existing immigrants to close the gap with natives. Finally, I consider how immigrants responded outside the labor market, in mobility, homeownership, and marriage.

¹I use “existing immigrants” and “incumbents” interchangeably throughout this paper.

I study these questions in the context of the Age of Mass Migration, during which over 30 million, mostly European, immigrants arrived in the U.S. between 1850 and 1924. As with more recent immigration, the volume and changing origin composition of arrivals sparked native backlash and congressional scrutiny, culminating in the quotas of the 1920s and the end of unrestricted mass migration. This setting is well suited to study this question as the sustained scale and shifting composition of immigration during this period from Northern and Western to Southern and Eastern Europe allow me to separate the effect of inflows in the aggregate from co-ethnic inflows.

I use complete-count census records from 1900 to 1930 ([Ruggles et al., 2024](#)). I link ten-year cohorts of working-age men at the county level using the Census Linking Project ([Abramitzky et al., 2021](#)). The linked structure lets me track individuals over time and reduces concerns about compositional change, but because I do not observe return migration, results should be interpreted as assimilation conditional on staying.

The empirical approach compares otherwise similar immigrants who were exposed to different amounts of new immigration. There are two threats to identification. The first is the sorting of existing immigrants to counties that have stronger labor markets or larger enclaves. I address this by exploiting the panel structure of the data to include county and state-by-year fixed effects, exploiting within-county variation over time so that identification comes from changes in immigrant inflows within a county over time. When considering co-ethnic inflows, which vary within a county, I include county-by-year fixed effects that absorb any local shock common to all immigrants in the county. Identification then compares immigrants from different origins living in the same county in the same decade.

The second and main threat is the non-random sorting of new immigrants to places due to the strength of the local economy. For instance, new immigrants may be drawn to counties with strong labor markets, which also improves the occupational status of existing immigrants. To address this form of endogeneity, I use a shift-share instrument for recent inflows. Since immigrants tend to settle near existing co-ethnic clusters, the instrument exploits vari-

ation in their 1900 distribution across the U.S. to predict new inflows. This identification strategy relies on the exogeneity of the initial country shares (Goldsmith-Pinkham et al., 2020). The identifying assumption is that historical county characteristics that affected immigration before 1900 are uncorrelated with location-specific economic trajectories between 1900 and 1930.

I follow Goldsmith-Pinkham et al. (2020) in conducting several robustness checks to assess the instrument’s validity. As the exclusion restriction hinges on changes rather than levels, using 1900 shares is plausibly exogenous since those shares are predetermined prior to new inflows. The restriction would be violated if pre-1900 settlement patterns predicted later changes in economic outcomes, conditional on fixed effects and other baseline controls. First, I show that the instrumented inflows do not predict pre-period changes in economic outcomes. Second, 1900 county characteristics explain little of the variation in immigrant settlement, and controlling for their correlates leaves the co-ethnic estimates unchanged. Third, the co-ethnic estimates are robust to interacting initial origin shares with year indicators, so they are not driven by particular origin groups settling in counties on differential growth paths. The aggregate estimate, by contrast, becomes more negative and reaches statistical significance under some of these controls.

I find that new inflows in the aggregate have little effect on the occupational standing or income scores of existing immigrants. Co-ethnic inflows, by contrast, have large negative effects. A one standard deviation increase in the share of recent co-ethnic immigrants implies a 29% decrease in occupational rank relative to the average occupational upgrading over the period, and widens the native-immigrant gap by 0.54 percentile points. Evaluated at the average co-ethnic inflow and cumulated over the three decades, the occupational rank penalty accounts for roughly a quarter of the gap in immigrants’ occupational advancement relative to natives that Collins and Zimran (2023) document for this period. These findings suggest that what matters for existing immigrants in the labor market is not the volume of new immigration but its composition, with co-ethnic competition hurting incumbents the most.

The contrast between aggregate and co-ethnic effects suggests that immigrants of different origins were not direct competitors, and that the co-ethnic penalty operates through competition within the ethnic network. Consistent with this mechanism, I find that greater co-ethnic immigration has more negative effects in places where the occupations that existing and new immigrants hold are similar. Co-ethnic inflows also make incumbents less likely to move into white-collar work.

Outside the labor market, larger co-ethnic inflows strengthen existing immigrants' ties to the enclave. Incumbents are less likely to move out of the county, and the neighborhood they live in is more ethnically concentrated. They also give their children more foreign names and are more likely to become homeowners. Put together, these results suggest that rather than pushing incumbents to assimilate, new co-ethnic immigration reinforces ties to the home country and draws them deeper into the enclave.

This paper builds on the literature on immigrant economic assimilation (see [Abramitzky & Boustan, 2017](#), for a review). This literature finds that immigrants did not immediately catch up to natives in terms of economic outcomes ([Abramitzky et al., 2014](#)), but that there was substantial variation by cohort ([Collins & Zimran, 2023](#)). A separate literature assesses the economic effects of immigrants on natives ([Borjas, 2003](#); [Card, 2001](#); [Foged & Peri, 2016](#); [Ottaviano & Peri, 2012](#); [Tabellini, 2020](#)). Both literatures largely treat the existing stock of immigrants and new arrivals as a homogeneous group, leaving open whether new inflows themselves shape the economic outcomes of immigrants already in the destination country. I show that both the economic and social outcomes of existing immigrants are affected by the arrival of new immigrants, documenting another determinant of immigrant assimilation and the economic effects of immigration.

This paper contributes to the literature on the relationship between labor market outcomes of immigrants and future cohorts. [Beaman \(2012\)](#) studies the dynamic effect of ethnic network size on wages, finding that for new immigrants, a larger arrival cohort generates competition while a larger tenured cohort improves labor market outcomes. I study the reverse

direction, how a larger arrival cohort affects the incumbents already there, and find that co-ethnic arrivals lower their occupational rank. In a modern-day setting, [Sparber \(2019\)](#) studies the H-1B program and finds that new high-skill foreign labor substitutes for other recent foreign workers but not for established immigrants. Looking across all skill levels, I find that whether new arrivals substitute for incumbents depends on their origin composition. Unlike in the H-1B setting, established immigrants are not insulated from co-ethnic arrivals. Most closely related is [Albert, Glitz, and Llull \(2026\)](#), who show that rising immigrant inflows widen the immigrant-native wage gap, accounting for roughly one-fifth of the increase in that gap across U.S. arrival cohorts since the 1960s. I decompose that effect and show that during the Age of Mass Migration, aggregate inflows had little impact on the occupational standing of existing immigrants, while co-ethnic inflows lowered it.

This paper also speaks to the literature on ethnic enclaves and immigrant networks. Networks lower the cost of entering the labor market for new arrivals by transmitting job information ([Beaman, 2012](#); [Munshi, 2003](#)), and they concentrate immigrants into a narrow set of occupations. [Locke \(2025\)](#) finds that 46% of immigrants between 1850 and 1940 worked in ethnic occupational niches due to ethnic networks. I show that while ethnic networks can assist newcomers in entering the labor market, new co-ethnic inflows concentrating in the same occupations hinder the assimilation of existing immigrants. Existing work documents that living in an ethnic enclave can slow economic assimilation ([Abramitzky et al., 2024, 2025](#); [Blau et al., 2013](#); [Damm, 2009](#); [Eriksson, 2020](#)). I extend this work by showing that the arrival of new co-ethnics, and not only the presence of an established enclave, slows assimilation.

2 Historical Background

During the Age of Mass Migration between 1850 and 1924, over 30 million immigrants, mostly European, moved to the U.S. This episode of unrestricted migration can be broadly

divided into two periods. The first, up to 1890, saw most immigrants arrive from Western and Northern Europe, from places such as the British Isles, France, and Germany. The second, between 1890 and 1924, was driven mostly by immigrants from Southern and Eastern Europe, such as Italy and Russia. Contemporaneous accounts differentiated between the two groups, describing the “new immigrants” from Southern and Eastern Europe as “inferior” to the “old immigrants” from Northern and Western Europe (Kraut, 1982, p. 152).

Upon arrival, ethnic enclaves and networks helped new immigrants gain a foothold. New immigrants who already had family members living in the U.S. kept in close contact, as relatives and friends provided “steady” information about labor market conditions and housing, allowing them to make “informed decisions” (Bodnar, 1985, p. 57). In other cases, labor recruiters directed flows of co-ethnic immigrants to specific industries or cities for a fee (Bodnar, 1985, p. 58). The way most immigrants found work led to significant concentrations of ethnic groups in particular occupations (Locke, 2025). For instance, Slavic groups worked in mining while Italians worked in factories (Kraut, 1982, p. 77). These networks therefore steered new immigrants toward the occupations that incumbents already held, placing new arrivals in potential competition with them in the local labor market.

While some immigrants were transient, homeownership was not an unusual outcome. Ownership was rarely financed out of a single immigrant’s wages. Rather, immigrants purchased homes with the support of ethnic financial institutions, which provided more favorable loan terms. Chicago alone had over 70 Polish savings and loan associations by 1916, whose main purpose was financing mortgages for homeownership (Bodnar, 1985, p. 182). Homes also served as the center of family life, places where second-generation children were raised and continued to live well into adulthood (Bodnar, 1985, p. 182). Property therefore served as a way to deepen ties to the enclave across generations, even where local labor market opportunities were limited.

Immigrants assimilated in some aspects but not others. While they improved their occupational outcomes over time, there was generally slow or no convergence toward those of

natives ([Abramitzky et al., 2014](#)). Even so, there were differences across immigrant cohorts, with the first and last cohorts assimilating rapidly and the intermediate cohorts converging less with natives ([Collins & Zimran, 2023](#)). The extent of assimilation also depended on whether an individual lived in an immigrant enclave, as many did ([Eriksson & Ward, 2019](#)). In general, living in an enclave harmed economic assimilation. Immigrants in enclaves had lower earnings and were less likely to work in white-collar jobs than those outside them ([Abramitzky et al., 2024](#); [Eriksson, 2020](#)).

While living in an enclave is known to have slowed economic assimilation, it is not clear how the strengthening or weakening of enclaves through new immigration affected the immigrants already there. It is in this context that this paper examines how new inflows, and their origin composition, shaped the economic assimilation of existing immigrants.

3 Data

This paper uses the 1900-1930 full count Censuses made available by [Ruggles et al. \(2024\)](#). I include individuals from any country that had at least 100 people living in U.S. cities in 1900 and at least 1,000 new immigrants in each decade so that I have a sufficient stock of immigrants from the same country.² After doing so, I have 32 sending countries.³ The sample consists of immigrant men who arrived in years preceding each census and are not living in the South or in group quarters.⁴ I observe these individuals when they are aged 18-50, and observe them again when they are aged 28-60. I restrict the sample to counties that appear in each Census between 1900-1930.

I use the Census Linking Project to link first-generation immigrants ([Abramitzky et al., 2021](#)). The Census Linking Project (CLP) uses an algorithm based on each individual's NYSIIS standardized name, age, and place of birth to generate matches and is only able to match men. I use the conservative links which require an additional unique match on names

²I drop individuals with birthplaces listed as broad geographic regions or continents.

³Refer to [subsection 9.1](#) for the full list of sending countries.

⁴I exclude immigrants living in the South since the stock of immigrants living in the South is low.

within a 5-year band. The linkage rates range from 30-53%.⁵

Since there may be non-random selection into linking (Bailey et al., 2020), I use inverse probability weights to re-weight the linked sample to be representative of the restricted sample. I use the starting year variables of age, birthplace,⁶ occupational score, literacy, Census region, and nativity to predict the probability of an individual being linked across Censuses.⁷ The longitudinal structure of the data ensures that results are not being biased by compositional changes or due to selective internal migration (Abramitzky & Boustan, 2017; vom Lehn et al., 2024). Nonetheless, there are several limitations to the linked data. For instance, immigrants who passed away or selectively left the U.S. are not captured in the data. Therefore, my results should be considered as the effect of new immigration on existing immigrants who chose to stay, rather than for all or the average immigrant.

3.1 Treatment: the share of recent immigrants

The main explanatory variable is the share of a county’s population that is comprised of recent immigrants. I define recent as having immigrated within the last decade (immigrated between $t - 9$ and t) and scale this number by the county population at time t . Therefore, by construction, the sample of linked individuals is different from the sample of new inflows. This ensures that the effect is not being mechanically driven by compositional changes from the arrival of new immigrants. I follow the definition of Ruggles et al. (2024) in defining ancestry of first-generation immigrants by using the birthplace variable.

⁵1900-1910: 30.17%, 1910-1920: 35.78%, 1920-1930: 53.29%.

⁶I bin the country of origin into geographical regions.

⁷In addition, I use urban and labor force status to predict linkage.

3.2 Outcomes

3.2.1 Occupational ranks and income scores

Information on income was not collected in the Census until 1940. To study the effect of new immigration on prior immigrants' wages and labor market assimilation, I use two measures of occupational status and income. I consider the effects of new immigration on men who are currently in the labor force. The first is a measure of occupational ranking following [Collins and Zimran \(2023\)](#). This rank-based occupational score is constructed using information on wealth from the 1870 Census and estimates of 1900 annual income by occupation from [Haines and Preston \(1991\)](#). These measures are then used to rank individuals relative to non-Southern white men in the labor force. Therefore, an individual's occupational measure can change over time either due to a change in occupation, or due to a change in the distribution of occupations of men in the labor force.

The second measure is a predicted income score using the 1940 Census ([Abramitzky et al., 2024](#); [Saavedra & Twinam, 2020](#)), which has the advantage of being detailed enough to construct income scores that vary within occupations.⁸ Formally, I regress log income on fixed effects for age, detailed occupation codes, birthplace, region of residence, and all of their interactions.⁹ Then, I assign individual incomes in 1900-1930 at the age, birthplace, occupation, and region level.¹⁰ The predicted income scores allow for within-occupation variation in earnings stemming from differing birthplaces, age, and geography. Therefore, I view the two measures as complementary. The occupational rankings are a relative measure that allows me to study the occupational standing of immigrants relative to that of natives and accounts for changes in that distribution over time. Meanwhile, the predicted income

⁸This measure is similar to a LIDO score.

⁹I use reported incomes of white men between the age of 18-60 who are not living in group quarters or in the South. I follow the literature in multiplying top-coded income by 1.4. Since farmers did not always report farm income in 1940, I follow [Collins and Wanamaker \(2022\)](#) in imputing income for farmers using the 5% sample of the 1960 census. Specifically, I take the ratio of earnings of farmers and farm laborers and multiply that by the income of farm laborers to impute the income of farmers. Income scores in 1960 are predicted using ownership in addition to the same covariates used for the 1940 income score.

¹⁰Missing cells are assigned values from a specification without geographic indicators.

scores allow me to study absolute earnings rather than occupational status.

3.2.2 Foreignness of children’s names

I construct an index of the foreignness of names following [Abramitzky et al. \(2020\)](#). The foreignness index ranges between 0 and 1, with higher values denoting that a name is more likely to be given to foreigners than to natives. Formally, the foreignness index is constructed as follows

$$R_{name} = \frac{\frac{\#Foreigners_{name}}{Totpop_{foreign}}}{\frac{\#Natives_{name}}{Totpop_{native}}} \quad (1)$$

For each name, sex and birth cohort, I compute the frequency of that name among the foreign born and natives, using all individuals born 20 years prior to that cohort’s birth year.¹¹ This index is then normalized to $\frac{R}{1+R}$. A higher average F-index for an individual indicates that on average, their children had more foreign-sounding names, relative to other same-sex children born in the same cohort.

3.2.3 Residential segregation

To study residential segregation, I construct a representation index, or location quotient that captures how concentrated a neighborhood is by a specific ethnic group, relative to the concentration of the county ([Dai & Schiff, 2023](#); [Locke, 2025](#)). Formally,

$$RI_{ebct} = \frac{\frac{Pop_{be}}{Pop_e}}{\frac{Pop_{bc}}{Pop_c}} \quad (2)$$

where the RI varies at the enumeration district by birthplace level. The numerator is the share of enumeration district e ’s population born in country b , and the denominator is the share of county c ’s population born in country b . For instance, an RI of 1.2 means that an enumeration district is 1.2 times as concentrated by an ethnic group relative to the county.

To facilitate interpretation across census years and birthplaces, I rank each individual’s

¹¹Names with less than 100 observations are dropped. I construct the index using the NYSIIS phonetic encoding, so that results are not driven by transcription errors by enumerators.

RI relative to others from the same birthplace and census year. This allows me to make comparisons across years, as the boundaries of enumeration districts change. Changes to the RI can happen due to an individual moving to a different neighborhood, or from changes in the distribution of ethnic neighborhoods.

[Table 7](#) reports summary statistics for the main outcomes and covariates used in this paper. The linked sample contains 1,150,610 first-generation men, observed in 1,547 counties and drawn from 32 sending countries. Over a decade, the average man gains 2 percentile points of occupational rank and 0.15 log points of income score, but there is significant variation in outcomes. There is large variation across counties in the share of recent inflows. Recent arrivals average 2.39% of the county population with a standard deviation of 3.33%, and the co-ethnic component is an average of 0.44% with a standard deviation of 1.02%. Recent co-ethnic arrivals average 0.8% of county population against 6.5% for the aggregate inflow, so for the median man fewer than one in ten recent arrivals in his county share his origin. The two shares correlate at 0.28 across individuals, once fixed effects are partialled out.¹² Thus, the co-ethnic measure is not a rescaling of the aggregate but a distinct source of variation that this paper also exploits.

4 Empirical Strategy

I estimate the effect of new immigrant inflows on existing immigrants using the following specification on the linked sample of first-generation immigrants

$$y_{icbt} = \alpha + \beta Inflow_{ct} + \delta D_{icbt} + \theta_{st} + \phi_c + \gamma_b + \varepsilon_{icbt} \quad (3)$$

y_{icbt} is the change in an outcome between $t - 10$ and t for individual i from country of origin b who lived in county c at $t - 10$. The main labor market outcomes are the change in occupational rank and the change in log income score. I also consider geographic mobility,

¹²Fixed effects in [Equation 3](#) include birthplace, state by year, and county

homeownership, residential segregation, marriage, and children’s names. The main variable of interest is *Inflow*, the share of county c ’s population that consists of recent immigrants.¹³ D is a vector of individual controls including age and its square, the year of immigration, starting occupational rank, and literacy. State by year fixed effects allow each state to follow its own economic trajectory. Birthplace fixed effects absorb time-invariant differences across origins, such as the transferability of skills. County fixed effects absorb time-invariant county characteristics, including time-invariant sorting. Standard errors are clustered by county, the level at which the treatment varies. Identification of β comes from within-county changes in immigrant inflows over time, relative to other counties in the same state.

To examine the effect of recent co-ethnic immigration, I replace $Inflow_{ct}$ with the share of a county’s population that consists of recent immigrants from individual i ’s birthplace b . I instrument this variable with the country- b term of the sum in Equation 6, which predicts the share of recent immigrants from country b .¹⁴ Since the co-ethnic share varies across birthplaces within a county, I can include county by year fixed effects.¹⁵ Identification then comes from comparing, within the same county and year, immigrants whose own origin group received a larger predicted inflow with immigrants whose origin group received a smaller one. These fixed effects absorb common shocks to the county in each decade, including the aggregate inflow itself. In these specifications, standard errors are two-way clustered by county and birthplace.

I next assess whether these inflows speed or slow the economic assimilation of existing immigrants by comparing their labor market trajectories with those of natives.¹⁶ Specifically, I compare each immigrant’s change in occupational rank with the average change among

¹³Recent immigrants are those who arrived between t and $t - 9$. Each individual is assigned the inflow into the county he lived in at the start of the decade.

¹⁴Inflows from the same birthplace as defined by Ruggles et al. (2024).

¹⁵County by year fixed effects subsume the state by year and county fixed effects.

¹⁶Natives are individuals born to two native parents, i.e., the third-plus generation.

natives in the same county by estimating

$$Gap_{icbt} = \alpha + \beta Inflow_{ct} + \delta D_{icbt} + \theta_{st} + \phi_c + \gamma_b + \varepsilon_{icbt} \quad (4)$$

where Gap_{icbt} is constructed as follows:

$$Gap_{icbt} = g_{ct} - y_{icbt} \quad (5)$$

where y_{icbt} is the change in occupational rank for immigrant i and g_{ct} is the average change in occupational rank among natives in the same county. β therefore measures how recent inflows affect immigrants relative to natives, that is, the effect on the rate of occupational assimilation. Since the outcome subtracts the immigrant's own change from the native change, a positive β means the native-immigrant gap widens as natives pull ahead, and a negative β means that immigrants catch up.

4.1 Instrument for immigration

The key threat to identification is the non-random sorting of new immigrants into counties in a way that is correlated with location-specific labor trends. For instance, new immigrants may be attracted to counties that have strong labor markets, which would confound the estimated effect of inflows on the labor market outcomes of existing immigrants.

To address endogeneity concerns, I construct a shift-share instrument to predict the inflow of new immigrants, denoted in [Equation 6](#). I take the share of immigrants from country b living in county c , relative to the stock of immigrants from country b living across the U.S. in 1900 to construct the shares s_{cb} . M_{bt} is the national inflow of migrants from country b between t and $t - 9$. To make comparisons across counties of different sizes, I scale predicted inflows by the population in each county at time t .

$$Z_{ct} = \frac{1}{P_{ct}} \sum_b s_{cb}^{1900} \cdot M_{bt} \quad (6)$$

This instrument exploits chain migration, the empirical regularity that new immigrants settle where co-ethnics already live (Card, 2001, 2009). The instrument therefore compares counties with higher and lower 1900 shares s_{cb}^{1900} . The identifying assumption is that conditional on controls, historical features of counties that attracted immigrants before 1900 are uncorrelated with the evolution of the labor market between 1900 and 1930. Since the outcomes are measured in changes and the specification includes county fixed effects, the 1900 shares may be correlated with the level of county outcomes. They need only be uncorrelated with subsequent changes.¹⁷ The restriction would be violated if pre-1900 settlement patterns predicted later changes in economic outcomes, conditional on fixed effects and baseline controls. I provide evidence in subsection 4.3 to suggest that this is not the case.

4.2 First stage results

First stage results are shown in Table 8. Each column regresses the actual share of recent immigrants on the predicted share from Equation 6. Column (1) presents results using the aggregate shift-share instrument, and columns (2) and (3) use the co-ethnic immigration instrument. The aggregate instrument strongly predicts actual inflows and is significant at the 1% level in column (1). The co-ethnic instrument also predicts actual co-ethnic inflows, both in column (2) and in the preferred specification with county by year fixed effects in column (3). The Kleibergen-Paap F-statistics range from 26 to 47, well above the conventional threshold of 10.

¹⁷The assumption is analogous to parallel trends in a difference-in-differences design.

4.3 Instrument validity

In this section, I probe the validity of the instrument for immigration. Recent literature has highlighted the conditions under which the shift-share instrument is valid (Borusyak et al., 2022; Goldsmith-Pinkham et al., 2020; Jaeger et al., 2024). In this setting, identifying the effect of new immigration on the economic assimilation of existing immigrants relies on the “shares approach” (Goldsmith-Pinkham et al., 2020). That is, I rely on the exogeneity of 1900 immigrant shares for identification, which are pooled together by the instrument to capture differential exposure to new immigrant inflows. This period is particularly well-suited to a shift-share instrument. The “trend-break” in the composition of sending countries around the turn of the century alleviates the dynamic adjustment concern raised by Jaeger et al. (2024). Furthermore, WWI and the Quota Acts impose additional exogenous shocks to the level and composition of immigration (Ager et al., 2024; Tabellini, 2020).

To see what countries are driving the identifying variation, I compute the Rotemberg weights for each sending country (Goldsmith-Pinkham et al., 2020). A summary of the Rotemberg weights is presented in Table 9. The vast majority of Rotemberg weights are positive (93%). The countries with the largest Rotemberg weights are Russia, Austria, Hungary, Italy, and Finland. Therefore, the identifying variation behind the aggregate inflow estimates can be thought of as comparing counties with high shares of Russian, Austrian, Hungarian, Italian, and Finnish immigrants to counties with low shares of immigrants from these groups and observing how labor market outcomes evolve between 1900 and 1930.

The negative weights are quantitatively small. Under an assumption of homogeneous effects, negative weights are not a conceptual problem and do not pose a threat to identification. However, negative weights are susceptible to misspecification if there are heterogeneous treatment effects (Goldsmith-Pinkham et al., 2020). Such a concern about heterogeneous effects is unlikely in this case as the just-identified estimates using the shares of each country with the largest positive weights are tightly clustered around the overall estimate as shown

in Figure 1.¹⁸

I conduct several checks suggested by Goldsmith-Pinkham et al. (2020). Since the exogeneity restriction requires that different counties are not already trending differently, I conduct a falsification check for pre-trends in Table 10 and Table 11. I regress the change in outcomes in the “pre-period” 1880–1900 on the post-1900 instrumented immigrant inflow.¹⁹ I find that the 1900–1930 immigrant inflows are not predictive of economic changes in the county, suggesting that the results are not driven by different counties already being on different economic trajectories. There are, however, significant pre-trends in specifications that use the probability of various types of marriage as the outcome. Therefore, these results should be interpreted cautiously.

To assess balance in levels, I regress the shares of the high Rotemberg weight countries on plausible correlates of changes in labor market outcomes over time. The exclusion restriction fails if the initial shares of these countries are correlated with determinants of later changes in labor market outcomes. Table 12 presents results of this exercise. Each column is a seemingly unrelated regression of the county share of the noted sending country on the county share of manufacturing, agriculture, literacy, and urban shares and log population. Since every equation uses the same county characteristics, estimating them jointly as an SUR leaves the coefficients unchanged but allows a joint test across all five shares.

The R^2 is small in every column (0.05 to 0.09), so these county characteristics explain little of the variation in immigrant shares. The coefficients on the manufacturing share are small and insignificant in every column, and the agriculture share is significant only for Finland. Counties with higher Russian, Austrian, Italian, and Hungarian shares are more urban and more populated, and all five shares are higher in less literate counties. As an additional robustness check, I interact each of these characteristics with year indicators and

¹⁸Goldsmith-Pinkham et al. (2020) show that the shift-share instrument is equivalent to pooling together b just-identified GMM estimates using each country’s share as an instrument and the inflows as weights.

¹⁹The sample is constructed by linking individuals from 1880–1900 using the same criteria as described in the Data section. I collapse the panel to be at the county level for aggregate inflows, and county by birthplace level for co-ethnic inflows.

control for them in [Figure 2](#). The first row of each panel reports the baseline estimate, and the “Share correlates” row adds these controls. The aggregate estimate turns positive but remains statistically insignificant, and the co-ethnic estimate is essentially unchanged.

A more direct test of the shares approach asks whether counties settled by particular ethnic groups were on different economic trajectories for reasons unrelated to subsequent inflows. Since identification rests on the 1900 shares of the high Rotemberg weight sending countries, I control for each of those shares interacted with year indicators, which lets counties with larger initial Russian, Austrian, Italian, Hungarian, or Finnish populations follow their own time path. Rows 3 to 8 of [Figure 2](#) add these controls one country at a time and then jointly. The co-ethnic estimate in Panel B is essentially unchanged in every specification, including the one that controls for all five shares at once. The aggregate estimate in Panel A is more sensitive. The baseline coefficient is a tightly estimated null, but once initial settlement patterns are allowed to trend differently, the estimate becomes substantially more negative and is significant at the 5% level for most of the high Rotemberg weight origins as well as in the joint specification. The aggregate estimate thus takes the same sign as the co-ethnic estimate.

The shares used in that exercise are county-wide, so they apply to everyone living in the county regardless of their own origin. For the co-ethnic specification, a sharper version of the test controls instead for the 1900 county share of the individual’s own origin group interacted with year indicators, which allows origin groups that were more concentrated in their county in 1900 to follow different trajectories. [Table 13](#) reports these estimates. Unlike [Figure 2](#), these regressions include county by year fixed effects. Both coefficients are close to, and slightly larger in magnitude than, the baseline estimates in columns (3) and (4) of [Table 1](#), and both remain statistically significant. This suggests that the co-ethnic estimates are not driven by differential trends in the counties where each origin group had already clustered by 1900.

5 Results

This section reports how existing immigrants responded to new immigrant inflows between 1900 and 1930. I estimate the effect of aggregate and co-ethnic inflows on two measures of labor market standing, occupational rank and predicted income scores. I then ask whether these effects were large enough to change the rate of economic assimilation, comparing the gains of existing immigrants against those of natives exposed to the same inflow.

5.1 Labor market outcomes

Table 1 presents the results of OLS regressions of Equation 3 in Panel A, and 2SLS regressions using the instrument for immigration from Equation 6 in Panel B. Columns (1) and (2) show that aggregate inflows have statistically and economically insignificant effects on the occupational rank and income scores of incumbents.

The effect of new inflows in the aggregate can mask heterogeneity caused by the ethnic composition of the inflow. Co-ethnics are more likely to be labor market substitutes, as they come from a similar skill distribution in the home country, receive similar information about jobs, and are pulled to the same enclaves that host existing immigrants. To study this, I replace $Inflow_{ct}$ with the share of a county's population that consists of recent immigrants from the same birthplace as each immigrant. Columns (3) and (4) of Table 1 present results using this specification. Incumbents living in counties that received a larger share of recent co-ethnic immigration fared worse in the labor market. Column (3) shows that a one standard deviation increase in the co-ethnic share lowers occupational rank by 0.61 percentile points. This effect is around 29% of the average occupational upgrading over this period. Column (4) shows that the same increase lowers income scores by 1.89%, although this estimate is significant only at the 10% level. These results suggest that immigrant labor was more substitutable within origin groups than across them during this period. Together with the null effect of aggregate inflows, this suggests that what matters for existing immigrants is

not the volume of new arrivals but their origin composition.

Table 1: Effect of new immigration on the economic outcomes of existing immigrants, 1900–1930

	All inflows		Co-ethnic inflows	
	(1) Occ rank	(2) Log income	(3) Occ rank	(4) Log income
Panel A: OLS				
Share of recent immigrants	0.011 (0.019)	0.085 (0.092)		
Share of recent co-ethnic immigrants			-0.407*** (0.114)	-0.925 (0.693)
Adjusted R^2	0.301	0.064	0.302	0.068
Panel B: 2SLS				
Share of recent immigrants	-0.013 (0.037)	0.207 (0.161)		
Share of recent co-ethnic immigrants			-0.599*** (0.190)	-1.856* (0.925)
Decadal mean (Dep var)	0.021	0.149	0.021	0.149
County FE	Yes	Yes	No	No
State by year FE	Yes	Yes	No	No
County by year FE	No	No	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes
KP F-Stat	46.48	47.48	26.34	26.69
1 SD effect (pctile pts / %)	-0.04	0.69	-0.61	-1.89
Observations	1,148,276	1,075,883	1,148,211	1,075,814

Notes: Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, starting occupational rank, and literacy. Robust standard errors in parentheses. Columns 1–2 cluster at the county level, columns 3–4 are two-way clustered by county and birthplace. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2 Economic assimilation

This section considers the effect of new immigration on the economic assimilation of existing immigrants relative to natives in the same county. Because natives in the same county face the same inflow, a gain or loss for immigrants changes assimilation only if natives respond differently. [Table 2](#) presents results from [Equation 4](#). A positive coefficient means the gap widens as natives pull ahead, while a negative coefficient means immigrants catch up. Column (1) provides suggestive evidence that aggregate inflows narrow the gap between existing immigrants and natives, although the estimate is significant only at the 10% level.

Column (2) shows that a one standard deviation increase in the co-ethnic share widens the native-immigrant gap by 0.54 percentile points, or 7.4% of the starting gap of 7.2 percentile points. Column (3) decomposes the effect of column (1) into co-ethnic and other-origin inflows.²⁰ Co-ethnic inflows still widen the gap while other-origin inflows narrow it, and the two effects differ significantly. This suggests that the origin of new immigration mattered for incumbents' assimilation.

5.3 Attachment to the enclave

Existing immigrants who faced worse labor market prospects need not have responded through the labor market. This section documents how incumbents adjusted along non-market margins. I examine geographic mobility, homeownership, residential segregation, marriage, and the names given to children.

5.3.1 Geographic mobility and homeownership

One response to worsening local labor market conditions is to move somewhere with better economic opportunity. During the Age of Mass Migration, immigrants tended to move across counties, but stayed within urban areas (Zimran, 2022). Column (1) of Table 3 tests whether co-ethnic inflows pushed incumbents to leave their county. I find that a one standard deviation increase in the co-ethnic share makes them 2.3 percentage points less likely to move out of the county. Existing immigrants therefore did not respond to co-ethnic competition by moving to other labor markets. If anything, they became more likely to stay.

Column (2) turns to homeownership.²¹ Immigrants in counties that received a one standard deviation increase in the co-ethnic share are 0.95 percentage points more likely to be homeowners, which is 1.62% relative to the mean. This may seem at odds with the finding

²⁰County by year fixed effects cannot be included because the outcome is measured against the native mean in each county by year cell.

²¹I follow Collins and Margo (2011) in restricting the sample to heads of households in both starting and ending censuses, since the question on homeownership was only asked of the household head.

Table 2: Effect of new immigration on the native-immigrant occupational rank gap, 1900–1930

	Gap in occ rank		
	(1)	(2)	(3)
Panel A: OLS			
Share of recent immigrants	-0.023 (0.052)		
Share of recent other-origin immigrants			-0.077 (0.049)
Share of recent co-ethnic immigrants		0.331*** (0.119)	0.295** (0.125)
Adjusted R^2	0.302	0.302	0.302
Panel B: 2SLS			
Share of recent immigrants	-0.134* (0.072)		
Share of recent other-origin immigrants			-0.250*** (0.044)
Share of recent co-ethnic immigrants		0.524*** (0.169)	0.374** (0.161)
Starting gap to natives (percentile points)	7.20	7.20	7.20
State by year FE	Yes	Yes	Yes
County FE	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes
KP F-Stat	46.49	25.87	14.35
Std. beta	-0.02	0.02	0.02
1 SD effect (percentile points)	-0.446	0.535	0.381
p-value: equal 1 SD effects			0.000
Observations	1,148,254	1,148,254	1,148,254

Notes: The outcome is the change in the occupational rank gap to natives in the same county, taken as the native mean change less the individual's own, so a positive coefficient means the gap widens. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, starting occupational rank, and literacy. Robust standard errors in parentheses. Column 1 clusters at the county level, columns 2–3 are two-way clustered by county and birthplace. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

that co-ethnic immigration lowers occupational standing. As discussed in [section 2](#), homeownership was within reach for many immigrants. [Bodnar \(1985, p. 180\)](#) documents that

some immigrants of this period attained homeownership at higher rates than natives did. The pattern is also consistent with family formation and with incumbents bringing relatives over from the country of origin. In the same sample of household heads, a one standard deviation increase in the co-ethnic share raises the number of newly arrived immigrants living in the household by 46% relative to the mean, mostly relatives of the head (Table 24).

5.3.2 Residential segregation

New immigration can also affect the social integration of incumbents. Increased co-ethnic immigration may lower the social cost of remaining in the enclave, and so weaken incumbents' incentives to assimilate. In an era where most immigrants were residentially segregated, new co-ethnic inflows may have contributed to further segregation. Column (3) of Table 3 uses the change in each incumbent's percentile rank on the representation index (RI) as the outcome. A one standard deviation increase in the co-ethnic share moves incumbents 1.4 percentile points up the RI distribution, into neighborhoods where their group is more over-represented. Together with the mobility and homeownership results, this suggests that new co-ethnic immigration embedded incumbents further in the enclave despite its negative labor market effects.

5.3.3 Marriage

Another pull of the enclave is its marriage market. This section considers whether co-ethnic inflows changed whether incumbents married and whom they married. Table 4 presents results for men never married at the start of the decade. Column (1) shows that a one standard deviation increase in the co-ethnic share makes them 1.1% more likely to marry over the decade, relative to the mean. Columns (2) and (3) show whom they married. The same increase makes them 2.4% less likely to marry a U.S.-born spouse and 7.4% more likely to marry within their ethnic group, both relative to the mean, consistent with an expansion of the marriage market (Angrist, 2002). However, the pre-period falsification test fails for the

Table 3: Effect of new co-ethnic immigration on mobility, homeownership and enclave sorting, 1900–1930

	(1) Moved county	(2) Homeowner	(3) Δ RI rank
Panel A: OLS			
Share of recent co-ethnic immigrants	-1.011* (0.519)	0.313 (0.197)	0.650*** (0.147)
Adjusted R^2	0.084	0.239	0.011
Panel B: 2SLS			
Share of recent co-ethnic immigrants	-2.224*** (0.584)	0.931*** (0.267)	1.341*** (0.230)
Mean (Dep var)	0.370	0.585	-0.024
County by year FE	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes
KP F-Stat	26.36	33.59	39.91
Std. beta	-0.05	0.02	0.06
1 SD effect (pp / rank pts)	-2.27	0.95	1.37
1 SD effect (% of mean)	-6.12	1.62	
Observations	1,149,320	721,440	1,139,701

Notes: Column 3 is the change over the decade in the household’s enumeration district representation index rank, the group’s share of the enumeration district relative to its share of the county, ranked within birthplace by year; a positive coefficient means sorting into a more co-ethnic neighborhood. Column 2 is restricted to household heads in both censuses and column 3 to men matched to an enumeration district in both, and both add starting home ownership status to the controls. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, starting occupational rank, and literacy. Robust standard errors, two-way clustered by county and birthplace, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

marriage outcomes ([subsection 4.3](#)), so these estimates should be interpreted with caution.

5.3.4 Children’s names

One proxy for cultural assimilation is the names that immigrants give to their children ([Abramitzky & Boustan, 2017](#)). This section examines whether incumbents changed naming patterns for their children as a result of new immigration. New immigration may have reinforced incumbents’ ties to the home country, which could be reflected in the names

Table 4: Effect of new co-ethnic immigration on marriage outcomes, 1900–1930

	(1) Ever marries	(2) US-born spouse	(3) Same ethnicity
Panel A: OLS			
Share of recent co-ethnic immigrants	0.305** (0.144)	-1.191*** (0.202)	2.087*** (0.557)
Adjusted R^2	0.072	0.108	0.104
Panel B: 2SLS			
Share of recent co-ethnic immigrants	0.706** (0.273)	-0.651** (0.315)	2.748*** (0.753)
Mean (Dep var)	0.644	0.282	0.381
County by year FE	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes
KP F-Stat	14.44	14.44	14.44
Std. beta	0.02	-0.02	0.07
1 SD effect (pp)	0.72	-0.66	2.80
1 SD effect (% of mean)	1.12	-2.36	7.35
Observations	291,034	291,034	291,034

Notes: The sample is men never married at the start of the decade; columns 2–3 code men who did not marry as zero. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, starting occupational rank, and literacy. Robust standard errors, two-way clustered by county and birthplace, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

given to children. Because children’s birth years give annual variation in exposure, I use a different identification strategy. Rather than the decade-level county design above, I ask whether recent immigrant arrivals affect naming within the family. For child i born in year t of father f , I measure exposure as the immigrant inflow received by the father’s county in the five years preceding the birth year, scaled by the county’s 1900 population.

The sample consists of the co-resident children of the first-generation immigrant men in the linked panel. I take all children younger than 16 observed living with their father in the 1910, 1920, and 1930 censuses.²² I further restrict attention to births from 1905 onward,

²²I drop those that report a different birthplace than the father.

which ensures that the five-year window preceding every birth begins no earlier than 1900, so that the shares used to construct the instrument remain predetermined. Since identification

Table 5: Effect of new immigration on the foreignness of children's first names, 1900–1930

	All inflows (1)	Co-ethnic (2)	Decomposition (3)
Panel A: OLS			
Inflow (5yr pre-birth)	0.111*** (0.025)		
Co-ethnic inflow (5yr pre-birth)		0.488*** (0.143)	0.468*** (0.141)
Other-origin inflow (5yr pre-birth)			0.037* (0.020)
Adjusted R^2	0.287	0.287	0.287
p (co-ethnic = other)			0.038
Panel B: 2SLS			
Inflow (5yr pre-birth)	0.312*** (0.030)		
Co-ethnic inflow (5yr pre-birth)		1.045*** (0.319)	1.003*** (0.312)
Other-origin inflow (5yr pre-birth)			0.131** (0.054)
Mean (Dep var)	0.493	0.493	0.493
Father FE	Yes	Yes	Yes
Birth-year FE	Yes	Yes	Yes
KP F-Stat	46.09	38.15	22.03
Std. beta	0.08	0.08	0.07
Std. beta (other-origin)			0.03
p (co-ethnic = other)			0.117
1 SD effect relative to mean	3.45	3.11	2.99
Fathers	408,197	408,197	408,197
Observations	1,325,842	1,325,842	1,325,842

Notes: The unit of observation is a child of immigrants observed in the main analysis, and identification is within family. The treatment is the local inflow over the five years before the child's birth. The outcome is the foreignness of the NYSIIS phonetic encoding of the child's first name, where a higher value is more distinctively foreign. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted and control for birth order within the father. Robust standard errors in parentheses. Column 1 clusters at the county level, columns 2–3 are two-way clustered by county and birthplace. * p < 0.10, ** p < 0.05, *** p < 0.01.

exploits within-family variation, fathers observed with a single birth are dropped. The estimation sample consists of 1,325,842 children born to 408,197 fathers. I estimate

$$FNI_{ift} = \alpha + \beta Inflow_{ct} + \delta D_{ift} + \theta_f + \lambda_t + \varepsilon_{ift} \quad (7)$$

where $Inflow$ is instrumented by

$$Z_{ct} = \frac{1}{P_c^{1900}} \sum_{y=t-5}^{t-1} \sum_b s_{cb}^{1900} \cdot M_{by} \quad (8)$$

where FNI is the F-index for child i born to father f at year t . $Inflow$ is the immigrant inflow into county c over the five years before year t , scaled by the county's 1900 population, where c is the county in which the father is first observed. I include controls for birth order D , father fixed effects θ , and birth-year fixed effects λ . The instrument in [Equation 8](#) sums national inflows over the five years before the birth year. I report results for the aggregate inflow and for the co-ethnic inflow from the father's birthplace. Because the aggregate inflow varies only at the county by birth-year level, county by birth-year fixed effects would absorb it, so I also report a specification that enters co-ethnic and other-origin arrivals jointly. Standard errors are clustered at the county level, and at the county and birthplace level in the co-ethnic specifications.

[Table 5](#) presents results using this specification. Column (1) considers the effect of inflows in the aggregate, column (2) on co-ethnic inflows, and column (3) includes both. Existing immigrants responded to new inflows by giving their children more foreign-sounding names. Column (1) finds that a one standard deviation increase in recent immigration implies a 3.45% increase in the F-index, relative to the mean. Column (2) shows a 3.11% increase relative to the mean for co-ethnic inflows. Because column (2) does not absorb county by birth-year shocks, the co-ethnic estimate may partly pick up the aggregate inflow, so column (3) enters both jointly. I find that both result in fathers giving children more foreign names, but co-ethnic arrivals matter roughly twice as much per standard deviation (standardized

coefficients of 0.07 and 0.03). The difference is significant under OLS but not under 2SLS ($p = 0.117$), where the standard error on the co-ethnic term more than doubles. Unlike occupational standing, naming responds to aggregate as well as co-ethnic inflows.

5.4 Mechanisms

This section examines two mechanisms behind the results above. I first test whether competition within co-ethnic networks can account for the negative effect of co-ethnic inflows, using the degree of occupational overlap between incumbents and new arrivals. I then ask along which occupational margins existing immigrants moved, and whether co-ethnic inflows shifted those same margins.

5.4.1 Competition among co-ethnics

One explanation for why co-ethnic immigration lowers the occupational standing of existing immigrants while aggregate immigration leaves it unchanged is labor market competition within co-ethnic networks. Co-ethnic networks are important in helping new arrivals enter the labor market (Beaman, 2012; Munshi, 2003). A larger co-ethnic share may benefit new arrivals, but it can also create labor market competition that hurts existing immigrants. To assess whether existing immigrants and new arrivals are competing in similar occupations, I construct a Duncan dissimilarity index of the occupations of existing and new immigrants at the county by birthplace by decade level.²³ I then estimate:

$$y_{icbt} = \alpha + \beta_1 Inflow_{cbt} + \beta_2 (Inflow_{cbt} \cdot Dissim_{cbt}) + \beta_3 Dissim_{cbt} + \delta D_{icbt} + \phi_{ct} + \gamma_b + \varepsilon_{icbt} \quad (9)$$

Results using this specification are presented in Table 6. Column (1) reports occupational

²³I use two-digit occupation codes for both existing immigrants and the realized occupations of new immigrants. I drop (county $\times$ birthplace $\times$ decade) cells with fewer than 100 incumbents or fewer than 100 new arrivals, since the index is mechanically larger when computed using small populations.

rank while column (2) reports log income scores. I find evidence in support of this mechanism. The coefficient on the interaction between the share of recent co-ethnic inflows and the dissimilarity index is positively signed in both columns, significant at the 10% level for occupational rank and the 5% level for log income scores, indicating that the negative effect of co-ethnic inflows attenuates where new arrivals enter occupations distinct from those held by incumbents. Evaluated at the mean level of dissimilarity, the effect of co-ethnic inflows on occupational rank remains negative, at -0.431 . The point estimates imply that the effect crosses zero at a dissimilarity of 0.314, roughly 1.1 standard deviations above the sample mean of 0.238. Competition therefore binds across most of the observed distribution of occupational overlap, with 87.7% of the sample lying below that threshold. Only counties where new arrivals sort into markedly different occupations are insulated from it.

5.4.2 Occupational margins of adjustment

An increase in occupational ranks and income scores can come from changes in occupations either within or across broad occupational categories. For instance, an individual who starts as an unskilled laborer could move to a better-paying occupation within the unskilled category, or transition into a higher-ranked category such as white-collar work. To see which occupational moves lie behind the labor market results, [Table 14](#) presents results where the outcome variable is an indicator for transitioning between broadly defined occupational categories.²⁴ Columns (1), (3), and (4) show that aggregate inflows moved existing immigrants across occupations. They were more likely to enter white-collar work, more likely to leave farming, and less likely to remain in the same occupation, even though average occupational rank and income scores did not change. Columns (5) to (8) turn to co-ethnic inflows. Co-ethnic inflows reduced the probability that incumbents entered white-collar work (column (5)) and the probability that they left farming (column (7)). Neither type of inflow shifted

²⁴Column (1) is an indicator equal to one if an individual moved into a white-collar occupation between censuses, conditional on not being in a white-collar occupation in the prior census. Columns (2) and (3) are defined analogously for leaving an unskilled occupation and leaving farming, respectively. Column (4) is an indicator equal to one if an individual remains in the same detailed occupation code between censuses.

the probability of leaving an unskilled occupation (columns (2) and (6)).

Table 6: Heterogeneous effects of co-ethnic immigrant inflows on existing immigrants by occupational similarity, 1900–1930

	(1) Occ rank	(2) Log income
Panel A: OLS		
Share of recent co-ethnic immigrants	-1.032*** (0.366)	-3.309*** (1.187)
Co-ethnic share *	0.188* (0.103)	0.794*** (0.280)
Dissimilarity (SD)	0.005** (0.002)	0.011** (0.005)
Co-ethnic occ dissimilarity (SD, start)		
Adjusted R^2	0.279	0.132
Panel B: 2SLS		
Share of recent co-ethnic immigrants	-1.792** (0.661)	-5.461*** (1.929)
Co-ethnic share *	0.387* (0.199)	1.447** (0.572)
Dissimilarity (SD)	0.003 (0.002)	0.004 (0.006)
Co-ethnic occ dissimilarity (SD, start)		
Mean (Dep var)	0.015	0.166
County by year FE	Yes	Yes
Birthplace FE	Yes	Yes
KP F-Stat	33.17	31.89
Std. beta	-0.08	-0.08
Marginal effect at mean dissimilarity	-0.431	-0.373
Standard error	0.239	0.846
1 SD effect at mean dissimilarity (pctile pts / %)	-0.44	-0.38
Dissimilarity at which effect is zero	0.314	0.256
% of sample below	87.7	69.1
Observations	675,936	632,609

Notes: Dissimilarity is the Duncan index of occupational overlap between incumbents and new arrivals in the same county by birthplace by decade cell. Cells with fewer than 100 incumbents or fewer than 100 new arrivals are dropped. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, starting occupational rank, and literacy. Robust standard errors, two-way clustered by county and birthplace, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Robustness checks

6.1 1890 country shares

One concern with using shares from 1900 is that they are too temporally close to the shock of new immigrant inflows. If the determinants of immigrant shares in 1900 are correlated with factors that also affect the labor market outcomes of individuals, then the exclusion restriction would not hold. To address these concerns, I conduct a robustness check using country shares from the 1890 census instead.²⁵ Results using this version of the instrument are presented in [Table 15](#). The KP F-stat from the first stage falls from about 47 to 11 for aggregate inflows and from about 26 to 8 for co-ethnic inflows, which is expected as immigrants from “new” sending countries only started arriving in large numbers post 1890, so the 1890 shares are a weaker predictor of the inflows that drive the main result. Reassuringly, Anderson-Rubin confidence intervals suggest that the instrument using 1890 shares is robust to weak IV concerns. All estimates keep their sign and significance. The co-ethnic effect on occupational rank is larger than in the baseline (-0.90 versus -0.60), but the baseline estimate lies well inside the Anderson-Rubin interval.

6.2 WWI and Quota shocks

The shift-share instrument would not be valid if the shares are not exogenous, that is, if the determinants of immigrant shares are correlated with the evolution of labor market outcomes. An alternative is to rely on the exogeneity of the shifts, or immigrant inflows ([Borusyak et al., 2022](#)). In this section, I follow [Tabellini \(2020\)](#) and directly leverage the exogenous immigration shocks caused by World War I and the Quota Acts. Specifically, I replace [Equation 6](#) with the following:

²⁵The 1890 census tabulated population counts at the county by birthplace level. Several states report Norway and Denmark only as a combined count, so I treat the two as a single origin in every county, and combine the national inflows the same way so that the shares and the shifts align.

$$Z_{ct} = \frac{1}{P_{ct}} \sum_b s_{cb}^{1900} \cdot (\mathbb{1}(Allied) \cdot M_{bt}) \quad (10)$$

and

$$Z_{ct} = \frac{1}{P_{ct}} \sum_b s_{cb}^{1900} \cdot (Q_b) \quad (11)$$

for 1910-1920 and 1920-1930, respectively. The intuition behind the first instrument is that immigration from non-Allied countries was shut off during and after World War I, so only Allied countries should have inflows of new immigrants.²⁶ Similarly, the second instrument uses the total number of immigrants allowed into the U.S. by the two quota acts, Q_b . I estimate two specifications using these instruments. First, I combine these instruments with the instrument for 1900-1910 to estimate the effect on the full sample.²⁷ Second, I restrict the sample to 1910-1930 to only utilize the WWI and Quota shocks.

Results from this exercise are presented in [Table 16](#) and [Table 17](#). Odd numbered columns present the baseline results and even numbered columns present results using only the WWI and Quota instruments. The estimates track the baseline. On the full sample, the effect on occupational rank is a precise null under both the baseline and the alternative instruments (-0.013 and -0.014). In the 1910-1930 subsample the point estimates are negative but imprecise, and the Anderson-Rubin confidence interval for the alternative instruments contains zero. Log income scores are positive in every column, insignificant on the full sample, and significant at the 10% level in the 1910-1930 subsample, where the Anderson-Rubin interval excludes zero. The KP F-statistics exceed 10 in every column. The overidentification test does not reject at the 5% level in any specification, with Hansen J p-values of 0.20 and 0.39 for occupational rank and 0.08 and 0.23 for log income scores.

²⁶Neutral countries and Russia are coded as non-Allied countries.

²⁷I use the baseline SSIV for 1900-1910, the WWI IV for 1910-1920, and the Quota IV for 1920-1930, in one pooled 1900-1930 specification.

6.3 Year of immigration

The reported year of immigration (*yrimmig*) enters my analysis in two places. First, it defines the window over which I measure new immigrant inflows. Second, it controls for how long existing immigrants have been in the U.S. Two features of this variable raise concern. There is ambiguity over whether immigrants were reporting the year that they first entered the U.S., or the year of their most recent entry. Furthermore, individuals may have incorrect recall and heap in years ending in 0 or 5. I address each concern in turn.

I first show that heaping is not systematically correlated with literacy. In [Table 18](#), I regress an indicator for reporting a year ending in 0 or 5 on literacy, and find that the coefficient is small and statistically insignificant both with and without census fixed effects. Heaping therefore does not differentially affect less literate immigrants.

Next, I confirm that heaping does not drive the results on either side of the shift-share. On the inflow side, heaping at the round-years could potentially overstate the level of new immigration. I therefore construct an instrument that measures inflows over the narrower window $[t - 7, t]$ rather than $[t - 9, t]$, dropping the two years adjacent to the round-year boundary where reported arrival years heap. Results are reported in [Table 19](#) for aggregate inflows and [Table 20](#) for co-ethnic inflows. Both the first-stage and 2SLS estimates remain similar.

On the existing immigrant side, I control for the year of immigration since immigrants who have been in the U.S. longer are more assimilated ([Abramitzky et al., 2020](#)). To address potential bunching in years ending in 0 or 5, I replace the variable with five-year arrival-cohort indicators centered on the round years.²⁸ The estimates, reported in [Table 21](#), are again similar in magnitude and significance.

Finally, *yrimmig* may capture only an individual’s most recent arrival rather than their first. If this is the case, individuals arriving in the sample window still increase local labor supply and compete with existing immigrants. Thus, the inflow measure still captures the

²⁸For each year t ending in 0 or 5, I assign reported years in $[t - 2, t + 2]$ to a single cohort.

relevant labor-market shock.

7 Conclusion

This paper studied the effect of new immigration on the economic assimilation of existing immigrants between 1900 and 1930. Overall, I document that the arrival of new immigrants is another channel through which the assimilation of prior immigrants can be affected, a mechanism that has received little attention in the literature. New arrivals in the aggregate leave the occupational standing of the first generation unchanged. Instead, I find that the composition rather than the magnitude of inflows is what matters most for labor market outcomes. I document that new co-ethnic immigration has negative effects on the labor market outcomes of existing immigrants, possibly due to labor market competition, and deepening ties to the enclave. As a result, the economic assimilation of immigrants relative to natives is harmed.

To benchmark my results, I conduct a back of the envelope calculation relative to other estimates of assimilation from the literature. [Collins and Zimran \(2023\)](#) find that immigrants in the 1900 to 1930 linked cohort upgraded 2.9 percentile points less than natives who began in the same occupation.²⁹ I find that in the average county and decade, recent co-ethnic arrivals amounted to 0.44% of the county population. Applied to the 2SLS estimate, this implies that co-ethnic inflows reduced the decadal gain in occupational rank of the average existing immigrant by 0.26 percentile points, or 0.79 percentile points cumulated over the three decades from 1900 to 1930. Therefore, continued co-ethnic immigration therefore accounts for roughly a quarter of the gap in immigrants' occupational advancement relative to other natives over this period.³⁰ Several assumptions are important to note. First, my links are across adjacent censuses, so cumulating three decades describes a synthetic cohort rather than the single cohort they follow. Second, the calculation treats rank losses as linear

²⁹I use their conditional estimate which ranks farm family equal to farm laborers for a conservative estimate.

³⁰The penalty is $-0.599 \cdot 0.44 = 0.26$ percentile points per decade, and the gap is $\frac{0.26 \cdot 3}{2.9} \approx 0.27$

and persisting once incurred.

Two features of the setting shape how these estimates should be interpreted. First, because I observe only immigrants who remained in the United States and were successfully linked across censuses, the results describe the effect on those who stayed rather than on the average immigrant, who may have returned home. Second, the Age of Mass Migration combined high inflow volumes with strong occupational segregation by origin, so the co-ethnic competition channel is likely more salient here than in settings where new arrivals disperse across occupations.

References

- Abramitzky, R., & Boustan, L. (2017). Immigration in American Economic History. *Journal of Economic Literature*, 55(4), 1311–1345.
- Abramitzky, R., Boustan, L., & Connor, D. S. (2024). Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office. *The Journal of Economic History*, 84(2), 352–394.
- Abramitzky, R., Boustan, L., & Eriksson, K. (2020). Do Immigrants Assimilate More Slowly Today than in the Past? *American Economic Review: Insights*, 2(1), 125–141.
- Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., & Pérez, S. (2021). Automated Linking of Historical Data. *Journal of Economic Literature*, 59(3), 865–918.
- Abramitzky, R., Boustan, L. P., & Eriksson, K. (2014). A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *Journal of Political Economy*, 122(3), 467–506.
- Abramitzky, R., Boustan, L. P., & Giuntella, O. (2025). Enclaves and Assimilation in the Age of Mass Migration: Evidence from Ethnic Catholic Churches. *NBER Working Paper*, w33362.
- Achard, P. (2022). Ethnic Enclaves and Cultural Assimilation. *Working Paper*.
- Ager, P., Feigenbaum, J. J., Hansen, C. W., & Tan, H. R. (2024). How the Other Half Died: Immigration and Mortality in U.S. Cities. *Review of Economic Studies*, 91(1), 1–44.
- Albert, C., Glitz, A., & Llull, J. (2026). Labor Market Competition and the Assimilation of Immigrants. *American Economic Review*, 116(5), 1682–1722.
- Alesina, A., Miano, A., & Stantcheva, S. (2023). Immigration and Redistribution. *The Review of Economic Studies*, 90(1), 1–39.
- Alexander, R., & Ward, Z. (2018). Age at Arrival and Assimilation During the Age of Mass Migration. *The Journal of Economic History*, 78(3), 904–937.
- Angrist, J. (2002). How Do Sex Ratios Affect Marriage and Labor Markets? Evidence from America’s Second Generation. *The Quarterly Journal of Economics*, 117(3), 997–1038.
- Bailey, M., Cole, C., & Massey, C. (2020). Simple strategies for improving inference with linked data: A case study of the 1850–1930 IPUMS linked representative historical samples. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 53(2), 80–93.
- Beaman, L. A. (2012). Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S. *The Review of Economic Studies*, 79(1), 128–161.
- Blau, F., Kahn, L., Liu, A., & Papps, K. (2013). The transmission of women’s fertility, human capital, and work orientation across immigrant generations. *Journal of Population Economics*, 26(2), 405–435.
- Bodnar, J. E. (1985). *The transplanted: A history of immigrants in urban America*. Bloomington: Indiana Univ. Press.
- Borjas, G. J. (2003). The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market. *The Quarterly Journal of Economics*, 118(4), 1335–1374.

- Borusyak, K., Hull, P., & Jaravel, X. (2022). Quasi-Experimental Shift-Share Research Designs. *The Review of Economic Studies*, 89(1), 181–213.
- Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. *Journal of Labor Economics*, 19(1), 22–64.
- Card, D. (2009). Immigration and Inequality. *American Economic Review*, 99(2), 1–21.
- Collins, W. J., & Margo, R. A. (2011). Race and Home Ownership from the End of the Civil War to the Present. *The American Economic Review*, 101(3), 355–359.
- Collins, W. J., & Wanamaker, M. H. (2022). African American Intergenerational Economic Mobility since 1880. *American Economic Journal: Applied Economics*, 14(3), 84–117.
- Collins, W. J., & Zimran, A. (2023). Working Their Way Up? US Immigrants’ Changing Labor Market Assimilation in the Age of Mass Migration. *American Economic Journal: Applied Economics*, 15(3), 238–269.
- Dai, T., & Schiff, N. (2023). The structure and growth of ethnic neighborhoods. *Journal of Urban Economics*, 137, 103570.
- Damm, A. P. (2009). Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence. *Journal of Labor Economics*, 27(2), 281–314.
- Eriksson, K. (2020). Ethnic enclaves and immigrant outcomes: Norwegian immigrants during the Age of Mass Migration. *European Review of Economic History*, 24(3), 427–446.
- Eriksson, K., & Ward, Z. (2019). The Residential Segregation of Immigrants in the United States from 1850 to 1940. *The Journal of Economic History*, 79(4), 989–1026.
- Foged, M., & Peri, G. (2016). Immigrants’ Effect on Native Workers: New Analysis on Longitudinal Data. *American Economic Journal: Applied Economics*, 8(2), 1–34.
- Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review*, 110(8), 2586–2624.
- Haines, M. R., & Preston, S. (1991). *Fatal years : Child mortality in late nineteenth-century America*. Princeton, New Jersey: Princeton University Press.
- Jaeger, D. A., Ruist, J., & Stuhler, J. (2024). Shift-Share Instruments and Dynamic Adjustments: The Case of Immigration.
- Kraut, A. M. (1982). *The huddled masses: The immigrant in American society, 1880-1921*. Arlington Heights, Ill: Harlan Davidson.
- Locke, E. (2025). Ethnic-Occupational Niches: Evidence from the Age of Mass Migration. *Working Paper*.
- Munshi, K. (2003). Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market. *The Quarterly Journal of Economics*, 118(2), 549–599.
- Ottaviano, G. I. P., & Peri, G. (2012). Rethinking the Effect of Immigration on Wages. *Journal of the European Economic Association*, 10(1), 152–197.
- Ruggles, S., Nelson, M. A., Sobek, M., Fitch, C. A., Goeken, R., Hacker, J. D., . . . Warren, J. R. (2024). *IPUMS Ancestry Full Count Data: Version 4.0 [dataset]*. Minneapolis, MN: IPUMS, 2024.
- Saavedra, M., & Twinam, T. (2020). A machine learning approach to improving occupational income scores. *Explorations in Economic History*, 75, 101304.
- Sparber, C. (2019). Substitution between groups of highly-educated, foreign-born, H-1B workers. *Labour Economics*, 61, 101756.
- Tabellini, M. (2020). Gifts of the Immigrants, Woes of the Natives: Lessons from the Age

- of Mass Migration. *The Review of Economic Studies*, 87(1), 454–486.
- vom Lehn, C., Wilson, R., & Price, J. (2024). The Winners and Losers of Immigration: Evidence from Linked Historical Data. *Working Paper*.
- Ward, Z. (2025). Women’s Economic Assimilation in the Age of Mass Migration.
- Zimran, A. (2022). US immigrants’ secondary migration and geographic assimilation during the Age of Mass Migration. *Explorations in Economic History*, 85, 101457.

8 Tables and figures

Table 7: Summary statistics

	Mean	SD	Min	Max	N
<i>First generation</i>					
Age	34.79	8.68	18	50	1,150,610
Year of immigration	1894.54	12.68	1815	1920	1,150,610
Change in occupational rank	0.02	0.27	-1	1	1,148,280
Starting occupational rank	0.51	0.25	0	1	1,149,389
Ending occupational rank	0.53	0.25	0	1	1,149,331
Starting income	1408.68	1313.55	42	7359	1,106,705
Ending income	1789.81	1796.84	40	9227	1,107,290
Change in log income	0.15	0.91	-3	3	1,076,821
Moved county	0.37	0.48	0	1	1,150,610
Homeowner (start)	0.40	0.49	0	1	721,551
Homeowner (end)	0.58	0.49	0	1	721,551
Married by end of decade	0.64	0.48	0	1	291,741
Spouse US born	0.28	0.45	0	1	291,741
Spouse same origin	0.38	0.49	0	1	291,741
Enclave RI rank (start)	0.67	0.24	0	1	1,150,610
Enclave RI rank (end)	0.65	0.25	0	1	1,140,986
Change in enclave RI rank	-0.02	0.26	-1	1	1,140,986
<i>Unique individuals</i>	1150610				
<i>County characteristics</i>					
County population (thousands)	33.62	97.54	0.27	2067.88	4,583
Recent inflows (%)	2.39	3.33	0.00	31.40	4,583
Recent co-ethnic inflows (%)	0.44	1.02	0.00	19.28	4,583
<i>Counties</i>	1,547				
<i>Birthplace characteristics</i>					
Recent co-ethnic inflows (%)	0.20	0.40	0.00	2.61	96
<i>Birthplaces</i>	32				
<i>Occupational dissimilarity</i>					
Co-ethnic occ. dissimilarity (D)	0.28	0.08	0.07	0.66	1,442
<i>County $\times$ origin $\times$ year cells</i>	1,442				

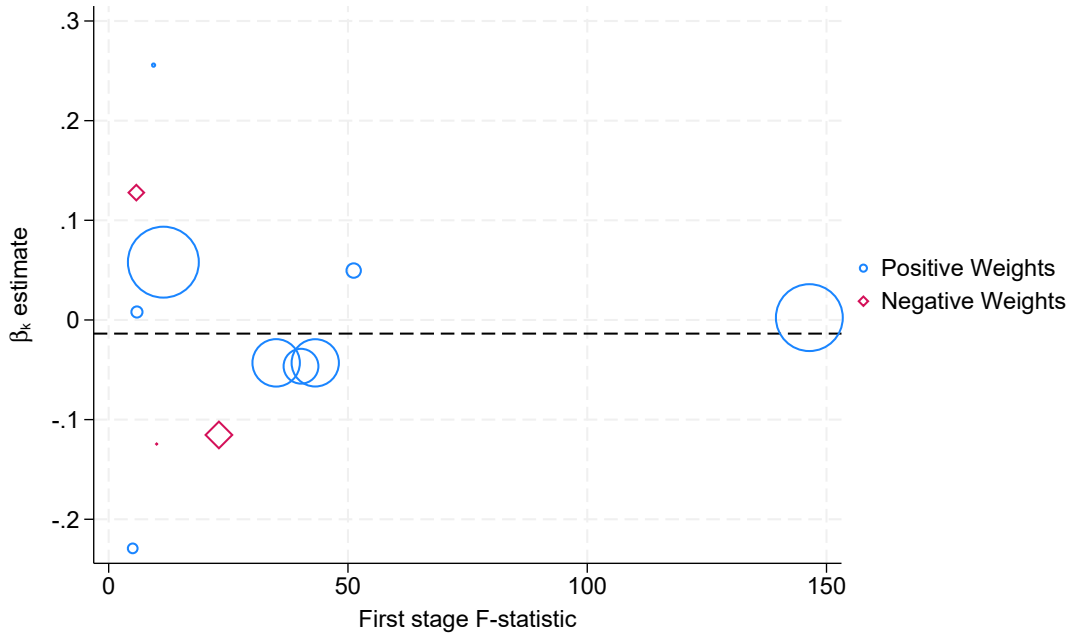
Notes: This table shows the summary statistics of the main variables used in this paper.

Source: Authors' calculations using [Ruggles et al. \(2024\)](#).

Table 8: Shift-share first-stage results

	All inflows	Co-ethnic inflows	
	(1)	(2)	(3)
Share of recent immigrants	0.379*** (0.056)	0.279*** (0.055)	0.336*** (0.065)
KP F-Stat	46.500	25.956	26.417
Adjusted R^2	0.933	0.686	0.756
Observations	1150005	1150005	1149940
Controls	Yes	Yes	Yes
County FE	Yes	Yes	No
State by year FE	Yes	Yes	No
County by year FE	No	No	Yes

Notes: Each column reports the coefficient from the first stage regression of the actual share of recent immigrants on the predicted share. All specifications are inverse probability weighted and include birthplace fixed effects. Standard errors in parentheses are clustered at the county level in column (1) and two-way clustered by county and birthplace in columns (2) and (3).

Figure 1: Heterogeneity of $\hat{\beta}_k$ 

Notes: This figure plots the relationship between each instrument's coefficient β_k , first stage F-statistic, and Rotemberg weights. The dashed line is plotted at the value of the overall $\hat{\beta}$. Only instruments with first-stage F-statistics above 5 are included.

Table 9: Summary of Rotemberg Weights

Panel A: Negative and positive weights					
	Sum	Mean	Share		
Negative	-0.078	-0.006	0.067		
Positive	1.078	0.049	0.933		
Panel B: Correlations					
	α_k	g_k	$\hat{\beta}_k$	F_k	$\text{Var}(z_k)$
α_k	1				
g_k	0.675	1			
β_k	0.082	0.168	1		
F_k	0.676	0.381	0.085	1	
$\text{Var}(z_k)$	-0.175	-0.330	-0.301	-0.148	1
Panel C: Top 5 Rotemberg weight countries					
	$\hat{\alpha}_k$	g_k	$\hat{\beta}_k$	95 % CI	
Russia	0.337	1.51e+06	0.058	(-0.040,0.138)	
Austria	0.300	9.22e+05	0.002	(-0.067,0.073)	
Hungary	0.151	5.11e+05	-0.043	(-0.160,0.053)	
Italy	0.150	1.63e+06	-0.043	(-0.159,0.038)	
Finland	0.081	1.24e+05	-0.046	(-0.120,0.002)	
Panel D: Estimates of β_k for positive and negative weights					
	α -weighted Sum	Share of overall β	Mean		
Negative	-0.011	0.781	-2.003		
Positive	-0.003	0.219	-3.090		

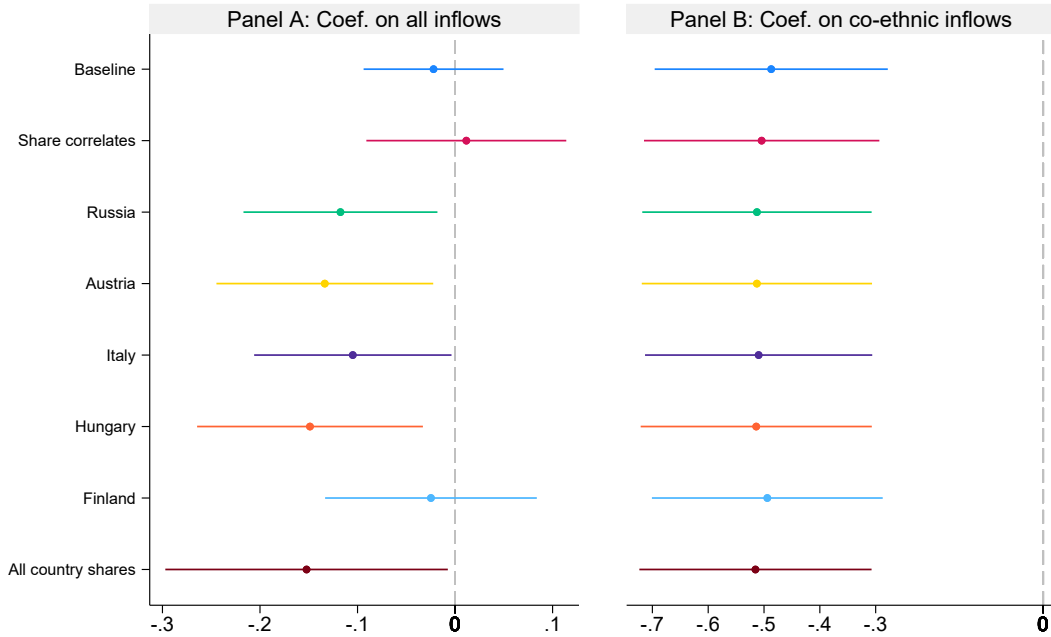
Notes: This table reports statistics regarding Rotemberg weights (Goldsmith-Pinkham et al., 2020). Panel A reports the total, mean, and share of positive and negative Rotemberg weights. Panel B reports correlations of Rotemberg weights ($\hat{\alpha}_k$) with each country's total immigrant inflow from 1901 to 1930 (g_k), the just-identified IV coefficients ($\hat{\beta}_k$) and its corresponding first stage F-stat (F_k), and the variance in country shares across counties ($\text{Var}(z_k)$). Panel C reports the countries associated with the five largest Rotemberg weights.

Table 10: Falsification check: pre-trends for aggregate inflows

	(1)	(2)
	Occ rank	Log income
Panel A: OLS		
Share of recent immigrants	-0.116* (0.063)	0.100 (0.217)
Adjusted R^2	0.540	0.409
Panel B: 2SLS		
Share of recent immigrants	-0.066 (0.134)	0.860 (0.810)
Mean (Dep var)	0.078	-0.167
State FE	Yes	Yes
KP F-Stat	15.37	14.79
Observations	1,509	1,505

Notes: The sample is collapsed to counties over the 1880–1900 pre-period. Panel A reports OLS and Panel B reports 2SLS. Robust standard errors, clustered at the state level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2: Effect of new inflows on economic outcomes of existing immigrants, additional controls



Notes: This figure plots the coefficient and corresponding 95% confidence intervals from a regression of Equation 3 with the additional control listed on the vertical axis.

Table 11: Falsification check: pre-trends for co-ethnic inflows

	(1)	(2)	(3)	(4)	(5)
	Occ rank	Log income	Moved county	US-born spouse	Same ethnicity
Panel A: OLS					
Share of recent co-ethnic immigrants	-0.171 (0.131)	-0.562 (0.668)	0.895*** (0.255)	-1.369*** (0.352)	2.107*** (0.314)
Adjusted R^2	0.066	0.137	0.313	0.185	0.222
Panel B: 2SLS					
Share of recent co-ethnic immigrants	-0.307 (0.222)	0.022 (0.903)	0.269 (0.416)	-1.451*** (0.432)	2.754*** (0.742)
Mean (Dep var)	0.075	-0.088	0.596	0.355	0.353
County FE	Yes	Yes	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes	Yes
KP F-Stat	34.42	35.01	34.42	46.73	46.73
Observations	11,986	10,706	11,988	7,307	7,307

Notes: The sample is collapsed to county by birthplace cells over the 1880–1900 pre-period. The treatment is the co-ethnic inflow, instrumented by the corresponding shift-share term. Columns 4–5 are restricted to men never married at the start of the period. Panel A reports OLS and Panel B reports 2SLS. All specifications include county and birthplace fixed effects. Robust standard errors, two-way clustered by county and birthplace, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Relationship between immigrant shares and county characteristics

	(1)	(2)	(3)	(4)	(5)
	Russia	Austria	Italy	Hungary	Finland
Log population	0.100*** (0.024)	0.091*** (0.018)	0.085*** (0.017)	0.094*** (0.017)	0.015 (0.014)
Manufacturing share	-0.008 (0.010)	-0.011 (0.007)	-0.002 (0.007)	-0.004 (0.007)	-0.007 (0.006)
Agriculture share	0.002 (0.005)	-0.005 (0.003)	-0.001 (0.003)	-0.004 (0.003)	-0.016*** (0.003)
Literate share	-0.007** (0.003)	-0.007*** (0.002)	-0.006** (0.002)	-0.008*** (0.002)	-0.004** (0.002)
Urban share	0.007*** (0.001)	0.005*** (0.001)	0.006*** (0.001)	0.004*** (0.001)	0.001 (0.001)
R^2	0.06	0.08	0.09	0.09	0.05
Joint nullity p-value	0.013	0.000	0.000	0.000	0.000
All $\beta = 0$ p-value	0.000				
Observations	1540				

Notes: Each column reports the result of a seemingly unrelated regression of 1900 immigrant shares of the top five Rotemberg weight countries on county characteristics in 1900. Outcome and all covariates that are shares are scaled by 100 to be percentage points. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: Robustness to controlling for own-origin 1900 county shares by year, co-ethnic inflows, 1900–1930

	Co-ethnic inflows	
	(1) Occ rank	(2) Log income
Panel A: OLS		
Share of recent co-ethnic immigrants	-0.462*** (0.117)	-1.005 (0.652)
Adjusted R^2	0.300	0.068
Panel B: 2SLS		
Share of recent co-ethnic immigrants	-0.644*** (0.189)	-1.968** (0.891)
Decadal mean (Dep var)	0.02	0.15
County by year FE	Yes	Yes
Birthplace FE	Yes	Yes
KP F-Stat	25.15	25.47
Std. beta	-0.03	-0.02
1 SD effect relative to mean	-26.77	-11.74
Observations	1,148,211	1,075,814

Notes: Both columns control for the 1900 county share of the individual's own birthplace interacted with year indicators. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted and include county by year and birthplace fixed effects. Controls: age, age-squared, year of immigration, and starting occupational rank. Robust standard errors, two-way clustered by county and birthplace, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: Effect of new immigration on occupation transition probabilities, 1900-1930

	All inflows				Co-ethnic inflows			
	(1) Entered WC	(2) Left unskilled farming	(3) Left farming	(4) Same occ	(5) Entered WC	(6) Left unskilled farming	(7) Left farming	(8) Same occ
Panel A: OLS								
Share of recent immigrants	0.152*** (0.028)	0.066 (0.067)	0.454*** (0.140)	-0.299*** (0.040)				
Share of recent co-ethnic immigrants					-0.229 (0.219)	-0.195 (0.316)	-1.273*** (0.202)	0.159 (0.260)
Adjusted R^2	0.034	0.053	0.090	0.061	0.034	0.053	0.095	0.062
Panel B: 2SLS								
Share of recent immigrants	0.152*** (0.049)	0.177 (0.135)	0.844*** (0.303)	-0.312*** (0.069)				
Share of recent co-ethnic immigrants					-0.743** (0.296)	-0.121 (0.480)	-2.389*** (0.515)	0.419 (0.411)
Decadal mean (Dep var)	0.126 Yes	0.536 Yes	0.264 Yes	0.420 Yes	0.126 No	0.536 No	0.264 No	0.420 No
County FE	Yes	Yes	Yes	Yes	No	No	No	No
State by year FE	Yes	Yes	Yes	Yes	No	No	No	No
County by year FE	No	No	No	No	Yes	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
KP F-Stat	43.18	59.20	25.66	46.52	24.28	34.16	50.51	26.36
Std. beta	0.02	0.02	0.06	-0.03	-0.03	-0.00	-0.06	0.01
1 SD effect (pp)	0.51	0.59	2.81	-1.04	-0.76	-0.12	-2.44	0.43
1 SD effect (% of mean)	4.04	1.10	10.62	-2.47	-6.04	-0.23	-9.22	1.02
Observations	920,485	247,726	136,844	1,149,385	920,414	247,416	136,622	1,149,320

Notes: The first three outcomes are estimated on the men who could make each move. Columns 1–4 report the effect of all recent inflows and columns 5–8 the effect of recent co-ethnic inflows; the fixed effect rows report which specification each column uses. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, starting occupational rank, and literacy. Robust standard errors in parentheses. Columns 1–4 cluster at the county level, columns 5–8 are two-way clustered by county and birthplace. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 15: Alternative instrument: 1890 country shares, economic outcomes, 1900-1930

	All inflows		Co-ethnic inflows	
	(1) Occ rank	(2) Log income	(3) Occ rank	(4) Log income
Panel A: OLS				
1890 shares	0.006 (0.019)	0.072 (0.091)		
1890 shares, co-ethnic			-0.424*** (0.126)	-0.879 (0.744)
Adjusted R^2	0.301	0.065	0.302	0.069
Panel B: 2SLS				
1890 shares	0.009 (0.038)	0.207 (0.349)		
1890 shares, co-ethnic			-0.898*** (0.297)	-2.215* (1.168)
Decadal mean (Dep var)	0.02	0.15	0.02	0.15
State by year FE	Yes	Yes	No	No
County FE	Yes	Yes	No	No
County by year FE	No	No	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes
KP F-Stat	11.42	11.38	7.97	7.93
Std. beta	0.00	0.01	-0.04	-0.03
Anderson-Rubin CI	[-0.092, 0.080]	[-0.449, 1.208]	[-2.026, -0.406]	[-7.109, -0.507]
Observations	1,134,685	1,063,224	1,134,625	1,063,159

Notes: The shift-share instrument is constructed from county by birthplace shares in the 1890 census. Norway and Denmark enter as a single origin, since several states report only a combined count in 1890. Columns 1–2 instrument the aggregate inflow; columns 3–4 instrument a separately built 1890-share co-ethnic inflow with its own shift-share term. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, starting occupational rank, and literacy. Robust standard errors in parentheses. Columns 1–2 cluster at the county level, columns 3–4 are two-way clustered by county and birthplace. Anderson-Rubin intervals are reported since the 1890-share instrument has a weaker first stage. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 16: Alternative instruments: WWI and Quota shocks, occupational rank, 1900–1930

	1900–1930		1910–1930	
	(1)	(2)	(3)	(4)
	Baseline	Lvl. hyb.	Baseline	Lvl. IV
Panel A: OLS				
Share of recent immigrants	0.011 (0.019)	0.011 (0.019)	0.115** (0.054)	0.115** (0.054)
Adjusted R^2	0.301	0.301	0.290	0.290
Panel B: 2SLS				
Share of recent immigrants	-0.013 (0.037)	-0.014 (0.031)	-0.450 (0.343)	-0.130 (0.162)
Sample	1900-30	1900-30	1910-30	1910-30
Decadal mean (Dep var)	0.02	0.02	0.02	0.02
State by year FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes
KP F-Stat	46.48	134.43	12.53	17.59
Hansen J p-value	.	0.20	.	0.39
Anderson-Rubin 95% CI				[-.582756, .197177]
Observations	1,148,276	1,148,276	865,857	865,857

Notes: The outcome is the change in occupational rank, first generation. Columns 1–2 use the full 1900–1930 sample and columns 3–4 are limited to 1910–1930. Columns (1) and (3) use the baseline shift-share instrument. Column (2) uses the baseline instrument for 1900–1910 and the WWI and Quota instruments for 1910–1920 and 1920–1930. Column (4) uses the WWI and Quota instruments alone. For column (4), where the first stage is weaker, I report weak-instrument-robust confidence intervals. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted. Robust standard errors, clustered at the county level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 17: Alternative instruments: WWI and Quota shocks, log income score, 1900–1930

	1900–1930		1910–1930	
	(1)	(2)	(3)	(4)
	Baseline	Lvl. hyb.	Baseline	Lvl. IV
Panel A: OLS				
Share of recent immigrants	0.085 (0.092)	0.085 (0.092)	0.154 (0.242)	0.154 (0.242)
Adjusted R^2	0.064	0.064	0.077	0.077
Panel B: 2SLS				
Share of recent immigrants	0.207 (0.161)	0.140 (0.124)	2.524* (1.313)	1.347* (0.753)
Sample	1900-30	1900-30	1910-30	1910-30
Decadal mean (Dep var)	0.15	0.15	0.15	0.15
State by year FE	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes
KP F-Stat	47.48	134.05	12.39	17.49
Hansen J p-value	.	0.08	.	0.23
Anderson-Rubin 95% CI				[.396033, 3.46576]
Observations	1,075,883	1,075,883	811,986	811,986

Notes: The outcome is the change in the log 1940 predicted income score, first generation. Columns 1–2 use the full 1900–1930 sample and columns 3–4 are limited to 1910–1930. Columns (1) and (3) use the baseline shift-share instrument. Column (2) uses the baseline instrument for 1900–1910 and the WWI and Quota instruments for 1910–1920 and 1920–1930. Column (4) uses the WWI and Quota instruments alone. For column (4), where the first stage is weaker, I report weak-instrument-robust confidence intervals. Panel A reports OLS and Panel B reports 2SLS. All specifications are inverse probability weighted. Robust standard errors, clustered at the county level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 18: Round-year heaping in reported year of immigration and literacy

	(1) Heaps (0/5)	(2) Heaps (0/5)
Literate	-0.001 (0.002)	-0.001 (0.002)
Observations	1151230	1151230
Start-census FE		Yes

Notes: The outcome is an indicator for a reported year of immigration ending in 0 or 5. The sample is linked first-generation immigrants, 1900–1930. Column (2) adds start-census fixed effects. Robust standard errors, clustered at the county level, in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 19: Robustness to a narrowed aggregate inflow window

	Occ rank		Log income	
	(1) Main	(2) Donut	(3) Main	(4) Donut
Panel A: OLS				
Share of recent immigrants	0.011 (0.019)	0.010 (0.021)	0.085 (0.092)	0.100 (0.099)
Adjusted R^2	0.301	0.301	0.064	0.064
Panel B: 2SLS				
Share of recent immigrants	-0.013 (0.037)	-0.005 (0.037)	0.207 (0.161)	0.214 (0.169)
Decadal mean (Dep var)	0.02	0.02	0.15	0.15
County FE	Yes	Yes	Yes	Yes
State by year FE	Yes	Yes	Yes	Yes
County by year FE	No	No	No	No
Birthplace FE	Yes	Yes	Yes	Yes
KP F-Stat	46.48	53.22	47.48	54.58
Std. beta	-0.00	-0.00	0.01	0.01
1 SD effect relative to mean	-2.09	-0.76	4.62	4.79
Observations	1,148,276	1,148,276	1,075,883	1,075,883

Notes: The “Main” columns use the baseline inflow window $[t - 9, t]$; the “Donut” columns narrow it to $[t - 7, t]$, dropping the two years adjacent to the round-year boundary where reported arrival years heap. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, average occupational rank, and literacy. Robust standard errors, clustered at the county level, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 20: Robustness to a narrowed inflow window, co-ethnic inflows

	Occ rank		Log income	
	(1) Main	(2) Donut	(3) Main	(4) Donut
Panel A: OLS				
Share of recent co-ethnic immigrants	-0.407*** (0.114)	-0.420*** (0.125)	-0.925 (0.693)	-0.909 (0.788)
Adjusted R^2	0.302	0.302	0.068	0.068
Panel B: 2SLS				
Share of recent co-ethnic immigrants	-0.599*** (0.190)	-0.642*** (0.226)	-1.856* (0.925)	-1.942* (1.072)
Decadal mean (Dep var)	0.02	0.02	0.15	0.15
County FE	No	No	No	No
State by year FE	No	No	No	No
County by year FE	Yes	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes
KP F-Stat	26.34	22.07	26.69	22.40
Std. beta	-0.03	-0.02	-0.02	-0.02
1 SD effect relative to mean	-28.54	-30.57	-12.70	-13.29
Observations	1,148,211	1,148,211	1,075,814	1,075,814

Notes: The co-ethnic counterpart of Table 19. The “Main” columns use the baseline inflow window $[t - 9, t]$; the “Donut” columns narrow it to $[t - 7, t]$, dropping the two years adjacent to the round-year boundary where reported arrival years heap. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, average occupational rank, and literacy. Robust standard errors, two-way clustered by county and birthplace, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

9 Appendix

9.1 List of sending countries

Australia	Germany	Portugal
Austria	Greece	Romania
Belgium	Hungary	Russia
Bulgaria	Ireland	Scotland
Canada	Italy	Spain
Cuba	Lithuania	Sweden
Czechoslovakia	Mexico	Switzerland
Denmark	Netherlands	Turkey
England	Norway	Wales
Finland	Philippines	Yugoslavia
France	Poland	

9.2 Classification of birthplaces

For consistency, I apply several standard birthplace classifications following the literature and as defined by [Ruggles et al. \(2024\)](#). While Poland was not a country for a portion of the period studied, enumerators in 1900 were instructed to distinguish between Austrian, German, and Russian Poland. Respondents were allowed to report their birthplaces as any of the three or just as Poland. I define Poland based on *bpl* code 45000, which includes all of the above birthplaces. I treat Austria and Hungary as separate entities following [Tabellini \(2020\)](#).

9.3 Economic assimilation: other outcomes

9.3.1 Horserace between co-ethnic inflows and all other inflows

Table 22 presents results from a horserace between recent co-ethnic inflows and recent inflows from all other origins. The effect of co-ethnic inflows on occupational rank remains negative and is significant at the 10% level. A one standard deviation increase in the share of recent co-ethnic immigrants lowers occupational rank by 0.41 percentile points, or roughly 20% relative to the decadal mean. In contrast, the effect of other-origin inflows is positive and significant at the 10% level for both occupational rank and log income scores. The one standard deviation effects of the two inflows are statistically different at the 5% level for occupational rank ($p = 0.014$) and at the 10% level for log income scores ($p = 0.088$). This provides more evidence that what matters for existing immigrants is the origin composition of new arrivals. Co-ethnic inflows lower their occupational standing, while inflows from other origins, if anything, raise it.

9.3.2 Heterogeneity by tenure and ethnic concentration

Given that the negative effects of co-ethnic inflows are moderated by occupational dissimilarity between existing immigrants and new arrivals, a natural question is whether that effect is distributed evenly across existing immigrants (Beaman, 2012). Within existing immigrants, more recent and longer-tenured immigrants may occupy different positions in the labor market. This suggests that there may be heterogeneous effects across groups. To examine this, I estimate:

$$y_{icbmt} = \alpha + \sum_k \beta_k (Inflow_{cbt} \times \mathbb{I}_{Tenure=k}) + \sum_{k \neq 0} \eta_k \mathbb{I}_{Tenure=k} + \delta D_{icbmt} + \phi_{ct} + \gamma_b + \psi_m + \varepsilon_{icbmt} \quad (12)$$

where y is the change in occupational rank or change in log income scores. I interact the share of recent co-ethnic immigration $Inflow$ with indicators for the three bins of years since migration indexed by k , and m indexes year of immigration. The sample is split into recent

immigrants (0-3 years), medium-term (4-6 years), and established immigrants (7 or more years), the last of which is the omitted category $k = 0$. All interactions are instrumented with the corresponding co-ethnic shift-share instrument. I include county by year, birthplace, and year of immigration fixed effects.

Table 23 presents results from Equation 12. Column (1) implies that the penalty from co-ethnic immigration is similar across arrival cohorts. A one standard deviation increase in the share of recent co-ethnic immigration lowers occupational rank by 0.62 percentile points for the most recent arrivals, 0.67 percentile points for the medium-term cohort, and 0.66 percentile points for established immigrants. Each is roughly 30% relative to the decadal mean, and all three are significant at the 1% level. Column (2) shows that the effect on log income scores has the same sign for every cohort. The same increase lowers income scores by about 2.0% for the most recent arrivals and 1.8% for established immigrants, both significant at the 10% level, while the estimate for the medium-term cohort is insignificant. The first stage is weak in both columns, with Kleibergen-Paap F-statistics just below 5, so these estimates should be read with caution. Taken together, the results suggest that co-ethnic competition lowered the occupational standing of existing immigrants regardless of how long they had been in the United States.

9.3.3 New arrivals in the household

Table 24 examines whether existing immigrants took new arrivals into their households. The outcome is the number of immigrants who arrived during the decade and live in the man's household at the end of the decade. I restrict the sample to men who were household heads in both censuses. Column (1) shows that a one standard deviation increase in the share of recent co-ethnic immigration raises the number of new arrivals in the household by 0.08, or 46% relative to the mean. Most of this increase comes from family members. Column (2) implies an increase of 0.06 new arrivals who are relatives of the household head, compared with 0.02 boarders, lodgers, and other non-relatives in column (3). Column (4) shows that

nearly all of these new arrivals share the birthplace of the household head. These results are consistent with existing immigrants bringing family members over from the country of origin. Larger households of newly arrived relatives may also help explain why co-ethnic inflows raised homeownership ([Table 3](#)).

9.4 Appendix tables and figures

Table 22: Horserace between recent co-ethnic and other-origin inflows, 1900-1930

	(1) Occ rank	(2) Log income
Panel A: OLS		
Share of recent co-ethnic immigrants	-0.291** (0.114)	-0.559 (0.676)
Share of recent other-origin immigrants	0.063*** (0.022)	0.197* (0.115)
Adjusted R^2	0.301	0.064
Panel B: 2SLS		
Share of recent co-ethnic immigrants	-0.404* (0.201)	-0.755 (1.012)
Share of recent other-origin immigrants	0.075* (0.044)	0.431* (0.240)
Decadal mean (Dep var)	0.021	0.149
County FE	Yes	Yes
State by year FE	Yes	Yes
County by year FE	No	No
Birthplace FE	Yes	Yes
KP F-Stat	14.35	14.07
Std. beta, co-ethnic	-0.02	-0.01
Std. beta, other-origin	0.01	0.02
1 SD effect, co-ethnic (pctile pts / %)	-0.41	-0.77
p-value: equal		
1 SD effects	0.014	0.088
Observations	1,148,276	1,075,883

Notes: All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, starting occupational rank, and literacy. Robust standard errors, two-way clustered by county and birthplace, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 23: Heterogeneous effects of co-ethnic immigrant inflows on existing first-generation immigrants by duration of residence, 1900-1930

	(1) Occ Rank	(2) Log income
Recent (0-3 years) arrivals * inflow	-0.611*** (0.207)	-1.987* (1.092)
Recent (3-6 years) arrivals * inflow	-0.661*** (0.231)	-1.667 (1.210)
Earlier arrivals * inflow	-0.651*** (0.195)	-1.766* (1.011)
Mean (Dep var)	0.02	0.15
KP F-Stat	4.800	4.949
County by year FE	Yes	Yes
Birthplace FE	Yes	Yes
Immigration cohort FE	Yes	Yes
Observations	1,148,210	1,075,813

Notes: All specifications are inverse probability weighted and include county by year, birthplace, and immigration cohort fixed effects. Controls: age, age-squared, starting occupational rank, and literacy. Robust standard errors, two-way clustered by county and birthplace, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 24: Effect of new co-ethnic immigration on new arrivals in the household, 1900-1930

	(1)	(2)	(3)	(4)
	Any new immigrant	Family	Non-family	Co-ethnic
Panel A: OLS				
Share of recent co-ethnic immigrants	7.929*** (1.241)	5.572*** (1.068)	2.356*** (0.464)	6.937*** (1.139)
Adjusted R^2	0.056	0.041	0.027	0.047
Panel B: 2SLS				
Share of recent co-ethnic immigrants	7.605*** (1.331)	5.778*** (1.090)	1.827*** (0.592)	6.951*** (1.224)
Mean (Dep var)	0.170	0.125	0.045	0.120
County by year FE	Yes	Yes	Yes	Yes
Birthplace FE	Yes	Yes	Yes	Yes
KP F-Stat	33.59	33.59	33.59	33.59
Std. beta	0.13	0.11	0.07	0.13
1 SD effect (% of mean)	45.52	47.07	41.23	58.92
Observations	721,440	721,440	721,440	721,440

Notes: The outcome is a count of new immigrant arrivals living in the man's household at the end of the decade, split by their relationship to him. The sample is restricted to men who were household heads in both censuses. All specifications are inverse probability weighted. Controls: age, age-squared, year of immigration, starting occupational rank, literacy, homeownership in starting year. Robust standard errors, two-way clustered by county and birthplace, in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.