Ethnic Neighborhoods and Assimilation: Evidence from Full Count Historical Census Data

Tianran Dai* Sai Luo† Nathan Schiff ‡

September 2025

PRELIMINARY: do not cite or circulate

Abstract

In this paper we study the effect of ethnic neighborhoods on the assimilation of immigrants using full count historical census data from the United States between 1900 and 1940. A central question in the immigration literature is whether these neighborhoods help or hinder immigrants in adjusting to the economic environment of their new host country. We take advantage of linked census data to examine how an individual's residence in childhood affects their outcomes in early adulthood. Following Chetty and Hendren (2018), we estimate the causal effect of ethnic neighborhoods by comparing siblings who moved across counties at different ages and received differential exposure to ethnic neighborhoods. In our preliminary analysis, we find that a longer childhood exposure to ethnic neighborhoods (counties) leads to higher income in early adulthood. The effect is small on average, but significantly larger for certain ethnic groups. Longer childhood exposure also increases the likelihood of living in an ethnic neighborhood as an adult.

1 Introduction

Ethnic neighborhoods, which are typically defined as areas with unusual concentrations of immigrants from a single source country, are a prominent feature of North American cities and have existed in some form since the mid-19th century (Dai and Schiff 2023, Logan, Zhang, and Alba 2002, Portes and Rumbaut 2014). A central question in the immigration literature is whether these neighborhoods help or hinder immigrants in adjusting to the economic environment of their new host country. The "spatial assimilation" theory in sociology suggests that immigrants—who often have limited resources, little knowledge of the local labor market, face credit constraints, and may have difficulties speaking a new language, among other disadvantages—locate with other immigrants from the same birth country as a type of mutual support system (Alba, Logan, Stults, Marzan, and Zhang 1999, Allen and Turner 1996, Logan, Zhang, and Alba 2002). While this theory implies economic benefits to immigrants living in ethnic neighborhoods, it's also possible that ethnic neighborhoods isolate immigrants, hindering language learning, limiting job opportunities, and harming the

*University of Alberta (email: daitianran@gmail.com; website: http://www.tianrandai.net)
†Shanghai University of Finance and Economics (email: luosaiecon@gmail.com; website: https://sailuo.weebly.com/)
‡University of Alberta (email: nschiff@gmail.com; website: https://nathanschiff.com)

human capital development of children. The empirical evidence shows mixed results with some studies finding positive effects (Cutler, Glaeser, and Vigdor 2008, Damm 2009, Edin, Fredriksson, and Åslund 2003), others negative (Borjas 2000, Danzer, Feuerbaum, Piopiunik, and Woessmann 2022, Eriksson 2019), and a number finding that the effect depends on the average human capital level of the neighborhood (Aslund, Edin, Fredriksson, and Gronqvist 2011, Abramitzky, Boustan, and Connor 2020, Borjas 1995).

In this paper we study the causal effect of ethnic neighborhoods on the assimilation of immigrants using full count historical census data from the United States between 1900 and 1940. We first classify counties as ethnic neighborhoods using the methods described in Dai and Schiff (2023). We then use linked census data to examine how an individual's childhood residence in a given census decade affects adulthood outcomes in the following census decades. We study the effects of ethnic neighborhoods by comparing future outcomes for immigrant children living in neighborhoods with different ethnic population concentrations. The full count data—every observation from each census—gives us tens of millions of people matched across each pair of consecutive census years. This enables comparisons between similar individuals, such as those from different origin locations moving to the same destination, or siblings from the same family. We focus on the outcomes of occupational income and location choice measured in early adulthood around ages 20–35.

To identify the causal effect of childhood exposure to ethnic neighborhoods, we adapt the strategy proposed by Chetty and Hendren (2018a). Specifically, we focus on immigrant children who moved with their families from different origin counties to the same destination county between census years. To address the potential endogenous moving choice of parents, we further compare siblings from the same family but who received differential exposure to the origin and destination neighborhoods due to their age differences. In our preliminary analysis, we find that a longer childhood exposure to ethnic neighborhoods—defined as counties with a high ethnic concentration relative to the national population distribution—has a significant positive effect on occupational income. Additionally, longer exposure to ethnic neighborhoods as a child increases the likelihood of living in an ethnic neighborhood as an adult.

2 Background

Our project is directly related to several notable recent papers studying intergenerational mobility and the economic assimilation of immigrants and their offspring into the U.S. economy. Abramitzky, Boustan, and Eriksson (2014) use linked US census data from 1900, 1910, and 1920 to study the wage growth of immigrants. The previous literature had used cross-sectional data to show that immigrants face a wage penalty upon arrival in the US—they earn less than natives in the same occupation—but that this penalty disappears over time as the immigrant gains experience in the US (immigrant and native wages converge). Abramitzky et al. use panel data—which fixes some biases of cross-sectional analysis—and find that there is no evidence of wage convergence. Additionally, they find that immigrants chose locations with higher nominal wages, compared to native location choices, suggesting an important role for geography.

Abramitzky, Boustan, Jácome, and Pérez (2021) study intergenerational mobility of immigrants in the US by using historical full count census data to match fathers and sons, as well as a smaller panel on immigrant children born in the 1980's. They find that the children of immigrants make greater gains in the wage distribution—have higher mobility—than the children of natives. Echoing an earlier paper (Abramitzky, Boustan, and Eriksson 2014), they argue that one reason for the greater mobility of immigrant children is that they grow up in areas with "better [economic] prospects." While Abramitzky, Boustan, Jácome, and Pérez (2021) suggest that the wage gap between immigrants and natives starts to decline in the second generation, Ward (2020) studies intergenerational mobility across three generations and finds that wage gaps between ethnicities persist into the third generation. To study three generations, he also uses historical full count census data and probabilistic matching to link men in the 1880 US census with their grandchildren

in the 1940 census. When evaluating the mechanism for this wage gap persistence, he suggests that much of the difference in wages can be explained by the neighborhood where the grandfather lived. He finds that when comparing the grandchildren of immigrants from different countries whose grandfathers lived in the same neighborhood (enumeration district or county), the persistence of wage gaps across immigrant groups is much smaller than estimates that do not condition on the grandfather's neighborhood. In contrast to the argument in Abramitzky, Boustan, Jácome, and Pérez (2021), this suggests that the grandfather's location has a negative effect on subsequent generations.[1] Thus all three of these papers suggest that geography is important for the economic assimilation of immigrants, yet whether the effects are positive or negative, and why geography is important is left unclear. Trying to fill the gap left by these important papers is part of the motivation for our project.[2]

Our project also fits into the well-developed literature on neighborhoods and place-based effects. Perhaps most influential are the recent papers by Chetty, Hendren, and Katz (Chetty, Hendren, and Katz 2016, Chetty and Hendren 2018a, Chetty and Hendren 2018b), which use administrative tax data from the United States to study children born in the 1980's over the period 1997-2010. These papers focus on the causal effects on adult outcomes of moving to a better neighborhood at some point in one's childhood. The existing literature had found little evidence for causal effects of neighborhoods, including analyses of the famous Moving to Opportunity experiment, in which families were essentially randomly assigned to new permanent locations. The recent literature shows that in fact a causal effect does exist, but it depends on how long a child was exposed to the better neighborhood. The treatment effect of interest and the sample population of our project is quite different from this recent work on neighborhoods. While this literature studies the effect of moving to high-income neighborhoods using a sample consisting mainly of native-born Americans, we focus on the effect of co-ethnic concentration on immigrant outcomes. Nonetheless, the identification challenges are similar and we will adapt some of the empirical strategies from Chetty and Hendren (2018a) to our context and data (see section 5.1).

The strand of the literature closest to our project studies the causal effects of ethnic neighborhoods on the outcomes of first- and second-generation immigrants. A set of papers have used refugee placement programs, which exogenously assign some immigrants to different locations within a country, to study education and labor market outcomes (Edin, Fredriksson, and Åslund 2003, Damm 2009, Damm 2014, Aslund, Edin, Fredriksson, and Gronqvist 2011, Abramitzky, Boustan, and Connor 2020). These papers rarely have longitudinal data and therefore focus on the outcomes of first-generation immigrants, adults and sometimes immigrant children, as in Aslund, Edin, Fredriksson, and Gronqvist (2011). A seminal paper by Borjas (Borjas 1995) uses longitudinal data from the National Longitudinal Surveys of Youth (NLSY) and a one percent sample from the 1970 census to study the effect of ethnic neighborhoods on the second generation. Borjas finds that neighborhood effects depend on the human capital level of co-ethnics—people who share your ethnic background—consistent with his theory of "ethnic capital," described in (Borjas 1992). These findings were influential but the sample size of the NLSY (about 12,000 individuals across 25 groups) limits the hypotheses that can be tested. Additionally, the main results are descriptive and do not try to rigorously identify causal effects of place.[3] Two recent papers use historical data to study ethnic neighborhoods, as we do. Eriksson (2019) studies the effect of living in Norwegian ethnic neighborhoods in the US in 1920, as

[1] Note that the findings do not directly contradict each other due to differences in the time periods, samples (comparisons across immigrant groups versus comparisons between immigrants and natives, second generation vs third generation), and econometric specifications.

[2] A recent paper by Fouka, Mazumder, and Tabellini (2022) studies how the migration of Black Americans from the South to the North affected non-earnings based assimilation outcomes—including naturalization, intermarriage, and language—for immigrants already resident in the North. While the historical event studied is quite different from our project, we also plan to examine many of these additional (non-earnings) measures of assimilation.

[3] Borjas does address possible measurement error with an instrumental variables strategy.

a child, on adult earnings measured in 1940. This paper is perhaps the closest to our project in that it also used linked historical US census data to study the effect of ethnic neighborhoods on multiple generations. However, Eriksson only studies a single ethnic group and focuses on earnings but not other assimilation outcomes, such as intermarriage, return migration, education, or language acquisition.[4] Abramitzky et al. (2025) examine the effects on first and second generation immigrants of growing up in a neighborhood with a Polish or Italian church, finding that Polish children have lower measures of economic and cultural assimilation as adults (more likely to marry a Polish Catholic, lower English speaking ability, more likely to be employed in manual occupations), but with little effect on Italians. These are interesting findings but for specific groups and for a particular type of ethnic enclave, whereas we seek to provide more general results on the effect of exposure to ethnic neighborhoods across many groups.

3 Data and Ethnic Neighborhood Definition

We have full count census data—the entire census sample—for all five decadal US censuses from 1900-1940 from the International Public Use Microdata Series (IPUMS). The observation counts are 76 million (1900), 92m (1910), 105m (1920), 123m (1930), and 132m (1940). IPUMS also provides an ID that allows us to link individuals across censuses. This linking is done using a statistical model based on birth dates, birth places, names, gender, and other information,[5] with match rates exceeding 30% and resulting in 30-50 million individuals matched across each census pair. Appendix Figure A1 shows linked record counts across the 1900-1940 censuses.

To examine the effect of location on outcomes, we use a set of individual characteristics collected in each census, including birth year, birthplace (country or US state), the birthplaces of their parents, year of immigration (if immigrants), and residential location when surveyed by the census (county and enumeration district). In Figure 1 we illustrate the geographic variation in immigrant location patterns by plotting county populations for three different immigrant groups in 1920, where country is defined by birthplace. One pattern that immediately stands out from all three panels is that there are many counties with very few (often zero) immigrants from a particular group. This does not result from those counties having small populations (see panel a of Figure 3 for the overall population distribution), but rather reflects that immigrants are highly concentrated, with the regions of concentration differing by group. Panel a) shows that Norwegian immigrants are heavily concentrated in the upper midwest, especially Wisconsin, Minnesota, and North Dakota. Mexican immigrants (panel b) are located along the border with Mexico, while Irish immigrants are mostly found in the Northeast (New York, New Jersey, Connecticut, and Massachusetts).

In Figure 2 we plot the location distribution of Italian immigrants across three different decades as an example of the time series variation in our data. The Italian population increased dramatically over this time period: there were 490,000 Italians in 1900 but over 1.6 million by 1920. However, not all groups increased over this time period. Germans represent a counterexample, with over 2.6 million Germans in 1900 but only 1.6 million by 1920. These examples illustrate that over the time period of our study there is substantial geographic and temporal variation in immigrant locations patterns, which we will use to identify the effects of ethnic neighborhoods.

[4] The paper instruments for the size of Norwegian ethnic neighborhoods in the US using the origin city in Norway: on average, people who emigrate from different cities in Norway end up in ethnic neighborhoods of different sizes. Our identification strategy exploits the size of our dataset to compare similar individuals from different neighborhoods, thus allowing for a more general analysis of neighborhood effects.

[5] A general discussion of the linking process is available here.

3.1 Defining Ethnic Neighborhoods

In Dai and Schiff (2023) we show that a basic model of location choice over the J locations in a city leads to a concentration measure calculated as the city's share of group g (ex: Chinese immigrants) choosing location j divided by the share of the total population choosing location j. This ratio of shares is commonly used in studying the location patterns of firms in specific industries where it is know as the "location quotient" (LQ). Letting N_g be the city level population of group g, n_{gj} the count of g individuals in location j, N the total population in the city (across all groups, natives and immigrants), and n_j the total population in j, the group g concentration of location j is:

$$LQ_{gj} = \frac{\frac{n_{gj}}{N_g}}{\frac{n_j}{N}} \tag{1}$$

When $LQ_{gj} = 1$ it implies that the share of group g locating in j is the same as the overall share of the population. Thus one way to interpret the location quotient is the proportional increase in the share of a group relative to that predicted by the overall population share. For example, if $LQ_{gj} = 5$ it implies that the share of group g locating in j is five times larger than what would be predicted by the overall population share.

In this draft we focus on counties, rather than within city ethnic concentration measures.[6] Counties vary widely in size. There are 2,834 counties in the 1900 data; the median county in 1900 had a population of about 16,000, the 99th percentile county was about 200,000 people, and the very largest county (covering New York City) was over 2 million. While large counties are clearly not "neighborhoods," the literature has shown that counties and even larger areas ("commuting zones") can still have place effects on residents (Chetty and Hendren 2018a, Eriksson 2019).

Figure 3 shows how the location quotient is calculated for Russian immigrants at the county level in 1920. Panel a) shows the distribution of the total population across counties in 1920 while panel b) shows the distribution of the Russian population. In panel c) we show the distribution of Russian county level LQ, which is calculated using the data from panel a) as the denominator and from b) as the numerator. Panel c) shows that there are a number of counties in North and South Dakota with very unusual shares of Russians. Many counties in 1920 have a Russian LQ of zero because there is not a single Russian born resident. Note that because LQ controls for the total population distribution, places with a large group population are not necessarily those with the largest LQ. For example, Chicago (Cook County) was one of the largest cities in 1920 and had a large Russian population, along with large populations of most immigrant groups (see the black county shaped like a seven in the southwest corner of Lake Michigan in panel b). However, the Russian location quotient was only 2.6, meaning Russians were over represented in Chicago, but not nearly as much as in the Dakota counties, where the LQ was often greater than 10, implying there were ten times more Russians than would be expected given the total population.

In Figure 4 we plot the LQs for Germans, Italians, and Polish immigrants in 1920. German LQs are generally highest in the midwest, with quite a few counties having an LQ between 3 and 5. Italians are generally located in fewer counties overall, but a number of counties around New York City and Northern California have LQs in the 3-5 range. The Polish population is unusually concentrated in Wisconsin and in the Northeast.

[6] In Dai and Schiff (2023), we define ethnic locations as those with an LQ greater than the 99th percentile of the location distribution for natives (people born in the US). We then group together contiguous locations that satisfy this threshold and call the grouped clusters "ethnic neighborhoods." We plan to employ a similar strategy in this paper by grouping together enumeration districts—see brief discussion in section 6.

4 Spatial Assimilation: Immigrant Life Cycle Location Changes

Our study focuses on children of immigrants who change locations during childhood and thus it's important to first understand how immigrants move across locations with varying co-ethnic concentrations. As described in our introduction, the "spatial assimilation" theory suggests that immigrants initially locate in ethnic neighborhoods and then move to locations with lower co-ethnic concentration after adjusting to the new country. While this idea is intuitive, it's quite difficult to test since it requires both panel data on immigrants changing locations and a large enough population to measure the ethnic concentration of different locations. In this section we provide the first systematic evidence on spatial assimilation across multiple decades for many immigrant groups.[7]

4.1 Measuring spatial assimilation with panel data

In their study of economic assimilation, Abramitzky, Boustan, and Eriksson (2014) compare the occupation trajectories of immigrants to natives, using a panel derived from the full count census data from 1900-1920. We focus on ethnic concentration, rather than occupation, but the econometric issues are quite similar and so we closely follow their approach. We first assemble a panel using the same filters as Abramitzky, Boustan, and Eriksson (2014), although our dataset is considerably larger.[8] We find all male immigrants in the 1900 data that arrived between 1880 and 1900 and were 18-35 in 1900. We further limit this sample to the 16 immigrant groups used by Abramitzky, Boustan, and Eriksson (2014) so that we can compare our results to theirs.[9] We also follow Abramitzky et al. and assemble a comparative panel of natives.[10] We then match both the immigrant cohort and the native cohort to their observations in the 1910-1930 censuses.

Note that the observation count will be much higher in 1900 than in later years, both because some immigrants moved back to their home country and because some who stayed we are unable to match. In our analysis, we will compare results using all observations to those when restricting to a completely balanced panel of individuals observed in every census from 1900-1930. If there's a difference in results across these two samples then it suggests that there is selection into return migration, or possibly in matching.

We start by showing patterns from the immigrant panel alone. To illustrate the general patterns in the data, we show simple binscatter plots of county LQ versus time in the United States (census year minus year of immigration) for Italians and Germans. Panel a) of Figure 5 shows that in the first six years, the average Italian immigrant lives in a county with 3.5-4 times more Italians than would be predicted by the population (LQ of 3.5-4). However, this declines fairly rapidly as time in the US increases, dropping to about 2.75 for an Italian who has lived in the country for 40 years. Panel b) shows that the German LQ for immigrants from Germany also drops with time in the US, but the scale is quite different. While the LQ drops by 1.5 units for Italians over 40 years, it declines from around 2.2 to 2 for Germans, a drop of only 0.2 points. Both the initial LQ and the rate of decline are quite different across the panels, demonstrating the substantial heterogeneity across immigrant groups. Lastly, for both groups, the LQ after 40 years is still above 2, indicating that

[7] Several papers have conducted case studies of particular groups over a limited time window (less than 10 years), see **?**. Relative to these studies, we use a different measure of ethnic concentration (LQ), study a longer time period (up to four decades), and have much larger sample sizes. This allows us to take a different econometric approach from these case studies whereby we incorporate many fixed effects to capture potential sources of heterogeneity.

[8] Abramitzky, Boustan, and Eriksson (2014) construct a panel using a 5% sample of the 1900 census from IPUMS, and additional data from Ancestry.com. IPUMS later released the full count census data and provided links based on the full count data, which makes our sample considerably larger.

[9] These are immigrants from Austria, England, France, Denmark, Ireland, Italy, Norway, Russia, Sweden, Belgium, Denmark, Finland, Portugal, Scotland, Switzerland, and Wales.

[10] We restrict the initial cohort to men 18-35 in 1900 and exclude Blacks and any men residing in the South, which Abramitzky et al. argue makes the cohort more comparable to the immigrant cohort.

both groups still remain in places with twice as many co-ethnic as would be expected (incomplete spatial assimilation).

We now estimate general spatial assimilation patterns by pooling all sixteen immigrant groups. To address the heterogeneity across groups, as well as additional heterogeneity by age, year of arrival, or over time, we use a fixed effect specification with indicators for group (birth country) g, arrival year a, census year t, as well as a cubic in age. Similar to Abramitzky et al., we break up duration in the US into a set of discrete bins $j \in J$. Let US duration indicator $D_{iat}(j)$ be equal to one if the time since arriving $t-a$, as measured in census year t for immigrant i, is contained within duration bin j (ex: 0-1 years). We then estimate:

$$y_{igt} = \sum_{j=1}^{20} \gamma_j D_{iat}(j) + age_{it} + age_{it}^2 + age_{it}^3 + \mu_a + \theta_t + \lambda_g + \varepsilon_{igt} \tag{2}$$

We use the county LQ measured in year t for immigrant i birth country g as our dependent variable y_{igt}. Then the γ_j coefficients describe how the county LQ changes with i's duration in the US. We use twenty duration bins incrementing every two years, but with the final bin capturing forty or more years. We normalize the coefficient on the first bin, years 0-1, to zero, and thus γ_2 to γ_{20} capture changes in LQ relative to the first two years.

In panel a) of Figure 6 we show results from the cross-section (unbalanced) panel. The first coefficient shows that the county LQ two to three years after arrival is about 1.5 less than the LQ within the first two years. The LQ continues to drop until reaching a relatively flat point at year 12. When restricting to the completely balanced sample, which reduces the observation count from 4.6 million to 471 thousand, the overall trend is similar but the initial drop is much less. This suggests that immigrants who remain in the US longer choose different initial locations, and indeed, we find that the initial LQ is about 0.82 units higher for an immigrant observed in all four decades compared to one observed only in 1900. Nonetheless, both panels show clear evidence of spatial assimilation with about a three unit drop in LQ—quite large—after 40 years.

One issue with the analysis shown in Figure 6 is that we cannot distinguish between spatial assimilation effects and general age effects. For example, individuals may choose to work in large cities when they are young and then move to smaller cities when they start to form families. Since large cities tend to have more immigrants, this general life cycle effect could result in a similar pattern of LQ changes as predicted by spatial assimilation theory. To distinguish age effects from spatial assimilation, we now re-run Equation (2) separately by immigrant group, but also include natives in the sample. For example, we will compare the Italian county LQ of Italian immigrants to the Italian county LQ of natives of equal age. We set all $D_{iat}(j)$ terms to zero for natives, making them the reference group, and thus the coefficient on the first bin ($j=1$) shows how the county LQ for a group within the first two years of arrival compares to natives of the same age. We use the fully balanced panel sample so that both immigrants and natives have observations over all four decades (1900-1930).

In Figure 7 we plot the results for Italians (panel a) and Germans (b). The results are broadly in line with the simple bin scatter plots in Figure 5, which excluded natives, suggesting that most of the change in LQ since year of immigration does not result from factors affecting natives, such as life cycle location patterns. Italians initially locate in a county with an Italian LQ roughly 3.5 units higher than natives. This then drops with duration in the US until after 40 years the difference is about 1.5 units. Germans initially locate in areas with a German LQ 0.9 to one unit higher than natives—a much lower difference than Italians. Over time this drops until after 40 years the difference is about 0.8 units.

These results are considerably different than the results on economic assimilation (occupation score) in Abramitzky, Boustan, and Eriksson (2014). In that paper the authors find that there is essentially no convergence in occupations, writing that "*regardless of starting point, immigrants experienced occupational upgrading similar to that of natives, thereby preserving the initial gaps between immigrants natives over*

time" (p469). Taking our results into account, this suggests that while immigrants move into areas with lower concentrations of co-ethnics over time, this does not affect any occupational differences with natives.

5 The Causal Effect of Growing Up in an Ethnic Neighborhood

5.1 Identification

The biggest empirical challenge to identifying the effect of neighborhoods is distinguishing between selection—individuals or households who choose a specific location may share (often unobservable) characteristics—and a true place effect. As an example, an important question is how living in an ethnic neighborhood affects English language acquisition. If we compare the children of Chinese immigrants living in a Chinese neighborhood with those of Chinese immigrants living in a non-Chinese neighborhood, we may find that the children who grew up in a Chinese neighborhood speak English at a lower level of fluency as adults. However, it would be difficult to argue that this is the causal effect of growing up in a Chinese neighborhood because it's likely that the characteristics of their parents differ. Chinese immigrants who choose to live outside of Chinese neighborhoods may already speak English at a higher level of proficiency and this would have an effect on the English language fluency of their children, no matter which neighborhood they grew up in.

To address this empirical challenge, Chetty and Hendren (2018a) study children who moved neighborhoods during childhood. While both the origin neighborhood and destination neighborhood are likely correlated with family characteristics, the key identification assumption is that the exact age when a child moved is exogenous. Under this assumption, children from the same birth cohort who moved from origin neighborhoods with equal ethnic concentration to the same destination neighborhood, but in a different move year, will have exogenously different exposure to each location. We describe this strategy more formally below, following the framework of Chetty and Hendren (2018a), but with modifications specific to our context.

We model an outcome y_i for child i in birth cohort s who moved in year m as a function of exposure to the ethnic concentration (LQ) of the origin neighborhood o, exposure to the ethnic concentration of the destination neighborhood d, birth cohort by moving year fixed effects α_{sm}, and origin interacted with destination fixed effects, α_{od}:

$$y_i = \alpha_{sm} + \alpha_{od} + b \times Exposure_i^o \times LQ^o + b \times Exposure_i^d \times LQ^d + \theta_i \tag{3}$$

The birth cohort by moving year fixed effects capture time-varying shocks that could affect adult outcomes, including wars and other historical events from this time period. The origin by destination fixed effects ensure that we compare children who moved across the same set of locations; if these locations are correlated with parental characteristics then this term helps to control for sorting. The error term θ_i is specific to child i and can include unobserved parental investment. Note that we explicitly assume the marginal effect of exposure to the destination and origin neighborhoods are equal (b).[11]

Following Chetty and Hendren (2018a), we further assume that exposure to a neighborhood affects adult outcomes up to a given age k. Then, given that children move once before turning age k, exposure time in each location must sum to k: $Exposure_i^d = k - Exposure_i^o$. Substituting back into the above equation

[11] Suppose the effect of exposure to the origin (b_o) and destination (b_d) neighborhoods are different, we rewrite the equation as $y_i = \alpha_{sm} + \alpha_{od} + b_o \times Exposure_i^o \times LQ^o + b_d \times Exposure_i^d \times LQ^d + \theta_i$. With the assumption that effect ends at age k, we can get $y_i = \alpha_{sm} + \alpha_{od} + b_d \times k \times LQ^d + b_o \times Exposure_i^o \times LQ^0 - b_d \times Exposure_i^o \times LQ^d + \theta_i$. We can estimate b_o and b_d separately using the same controls as in equation (4).

yields[12]:

$$y_i = \alpha_{sm} + \alpha_{od} + b \times Exposure_i^o \times (LQ^o - LQ^d) + \theta_i \tag{4}$$

Estimating Equation (4) will not recover b if the neighborhoods' exposure term, $Exposure_i^o \times (LQ^o - LQ^d)$, is correlated with θ_i. The selection effect can be written as $\delta_i = \frac{cov(\theta_i, Exposure_i^o \times (LQ^o - LQ^d))}{var(Exposure_i^o \times (LQ^o - LQ^d))}$. Chetty and Hendren (2018a) show that if the selection effect does not vary with the age when a child moves, then estimating different coefficients for age at move and taking the difference between consecutive ages yields the exposure effect of a single year.[13] In notation, our identifying assuming for estimating the unbiased effect of b is that $Corr(\theta_i, Exposure_i^o \times (LQ^o - LQ^d) | (\alpha_{sm}, \alpha_{od})) = 0$.

The difficulty in our context is that we only observe a child's location in each census decade, thus while we can observe moves, we do not know the intercensal year of the move. This prevents us from using Chetty and Hendren's exact identification strategy and also introduces measurement error in our exposure variables since we do not know exactly how long a child was in the origin and destination neighborhoods.

Our solution to both the selection effect and the unknown move year is to compare children within the same family. Specifically, we add a family fixed effect to our specification, which allows us to measure differential exposure using age differences and simultaneously differences out the selection effect, under the assumption this is constant within a family.

To illustrate the family fixed effect strategy, let $M_{f(i)}$ be the year child i born in birth year $BrthYr_i$ moved with their family f from o to d. Then we can define exposure to each location as:

$$Exposure_i^o = M_{f(i)} - BrthYr_i; \quad Exposure_i^d = k + BrthYr_i - M_{f(i)} \tag{5}$$

Consider a family f with two children, i and j, that moves from o to d in year M_f. The difference in exposure to the origin is $Exposure_i^o - Exposure_j^o = BrthYr_j - BrthYr_i$, thus the difference in birth years captures the differential exposure regardless of the unknown intercensal move year. Then the difference in outcomes for i and j is equal to:

$$y_i - y_j = (\alpha_{sm(i)} - \alpha_{sm(j)}) + b \times (BrthYr_j - BrthYr_i) \times (LQ^o - LQ^d) + (\theta_i - \theta_j) \tag{6}$$

The error term θ_i can be written as the sum of the selection effect δ_i and an idiosyncratic term ε_i: $\theta_i = \delta_i + \varepsilon_i$. If we assume that this selection effect is constant within a family, $\delta_k = \delta_f, \forall k \in f$, then differencing i and j above will difference out the selection effect: $(\theta_i - \theta_j) = (\delta_f + \varepsilon_i) - (\delta_f + \varepsilon_j) = \varepsilon_i - \varepsilon_j$. More generally, by adding family fixed effects to Equation (4) we can estimate the causal effect of differential exposure to the origin neighborhood without knowing when a family moved.[14]

[12] The $b \times k \times LQ^d$ term is captured by the origin-destination fixed effects α_{od}.

[13] To see this in our context, let β_m represent the true effect of m years of exposure to an origin neighborhood with LQ^o and $k-m$ years of exposure to a destination with LQ^d. Then rewrite Equation (4) to estimate a different coefficient for each origin exposure duration m, including destination by move year fixed effects α_{dm}: $y_i = \alpha_{sm} + \alpha_{dm} + \sum_{m=1}^{k} b_m \times I(Exposure_i^o = m) \times (LQ^o - LQ^d) + \theta_i$. The bias or selection effect for m exposure years can be written as $\delta_m = \frac{cov(\theta_i, I(Exposure_i^o = m) \times (LQ^o - LQ^d))}{Var(I(Exposure_i^o = m) \times (LQ^o - LQ^d))}$. Then the estimated coefficient b_m is $b_m = \beta_m + \delta_m$. Chetty and Hendren's identifying assumption in our context is that the selection effect does not vary with the number of years of origin exposure: $\delta_m = \delta, \forall m$. Then differencing b_m between two consecutive exposure durations yields $b_m - b_{m-1} = (\beta_m + \delta) - (\beta_{m-1} + \delta) = \beta_m - \beta_{m-1}$. This difference represents the effect of one additional year of exposure at $m-1$ years of exposure. However, the empirical results of Chetty and Hendren (2018a) show that these exposure effects are constant, $\beta_m - \beta_{m-1} = \gamma, \forall m$, and thus differencing any two coefficient estimates b_m and b_{m-1} yields the marginal effect of a year of exposure.

[14] Note that we cannot fully capture α_{sm} due to unobserved move year. However, as explained in the next section, we proxy for this using an interaction between birth year and moving cohort.

5.2 Implementation

In our estimation, we will estimate the effect of childhood exposure to ethnic neighborhoods among children who moved between year t and year $t+10$ on outcomes measured in $t+20$.[15] We pool together three cohorts of child movers, subscripted m: those who move between 1900 and 1910, 1910 and 1920, and 1920 and 1930. For each mover cohort, we restrict the sample to children aged 15 or less in the initial year, $t(m)$. As an example, for the 1900–1910 cohort we study children born between 1885 and 1900, thus they are 20 years old or less if their family moves in 1905 and they are aged 20–35 when the outcomes are measured. We measure the origin ethnic concentration in year $t(m)$ and the destination concentration in year $t(m)+10$.

This leads to the following specification:

$$y_{im,time\geq t(m)+20} = \alpha_{sm} + \alpha_{od} + b \times \left(M_{f(i)} - Byr_i\right) \times (LQ^o_{t(m)} - LQ^d_{t(m)+10}) + \varepsilon_i \tag{7}$$

Our sample covers children born in different cohorts over a 40-year period of history in which there were many momentous changes. Therefore it is worth emphasizing that the birth year by mover fixed effect α_{sm} implies that we are always comparing effects within the same birth cohort. Thus significant changes over time to our outcome variables, such as occupation or intra-marriage, are differenced out since we estimate exposure effects within a cohort. This α_{sm} term also fully captures any age effect on the outcomes. As discussed earlier, we include origin county by destination county fixed effects, α_{od}, so that our variation comes from comparing children within the same origin-destination pair, but who had differential exposure due to age differences. In specifications with family fixed effects, the variation comes from differential exposure within a family (across siblings) to the same origin-destination pair.

We run our specification on a sample of children born in the US with at least one parent born in a foreign country. By focusing on second-generation children we exclude any effects arising from the year of immigration. We also require that these children moved once with their families, which we define as living in a different county in census year $t(m)$ and census year $t(m)+10$. Our specification implicitly assumes that children observed in year $t(m)$ have been there since birth. This is likely often true because only about 10% of any group move between two consecutive census years. However, we also limit the sample to children whose resident state in year $t(m)$ is the same as their birth state (we do not observe birth county).

We define exposure to the origin county by assuming that all movers in a cohort moved in $t(m)+5$: $Exposure^o_{im} = (t(m)+5) - BrthYr_i$. This assumption can be relaxed when we include family fixed effects and compare children within the same family. As long as children from the same family moved together, the cross-sibling difference in exposure to the origin neighborhood will be the age difference between siblings.

We calculate the change in LQ as the origin ($t(m)$) county LQ minus the destination ($t(m)+10$) county LQ, where LQ is specific to the immigrant group. As we have shown in section 4, immigrant families tend to move from ethnic neighborhoods to non-ethnic neighborhoods, which provides the variation in the change in LQ (ΔLQ).[16] To get a sense of how the change in LQ correlates with other neighborhood characteristics, in Table 1, we use our main sample of movers (pooling different ethnic groups) and regress the origin-destination change in county LQ on the origin-destination change in a set of county covariates, including average occupation score, share speaking English, literacy rate, average age, and working-age population (20-65). In the first six columns we include the county covariates separately, and in the last column we include all county covariates in one regression. The correlations suggest that, when immigrant families move from ethnic to non-ethnic counties, the destination counties are generally richer (higher occupation

[15] To examine outcomes measured in later years (such as $t+30$), we need to match individuals across more census years, which comes with a loss of sample size.

[16] In our data, we observe moves in both directions (ethnic to non-ethnic, non-ethnic to ethnic), and there are more ethnic to non-ethnic moves on average.

score), have more people speaking English, have a higher literacy rate, have more natives, and have a larger working-age population.

In this draft, we examine two dependent variables: the occupation income score of the individual in $t(m+20)$ and the LQ of the county in which they lived in $t(m+20)$. Until 1950, the US Census did not record income, but they did record occupation. The occupation income score in year t is defined as the median income associated with each occupation in 1950, assigned to the same occupation in t. This measure is denominated in hundreds of 1950 dollars; if the occupation score is one unit higher for individual A than individual B, it implies that people with the occupation of A earned on average an additional $100 in annual income in 1950 compared to those with B's occupation.

5.3 Validation

Chetty and Hendren (2018a) conducted several tests to validate their empirical design. Their identification strategy and main results have also been verified and replicated in contexts outside the U.S. (Deutscher 2020, Britto, Fonseca, Pinotti, Sampaio, and Warwar 2022). While we largely adapt Chetty and Hendren's identification strategy, two differences in the data and context are worth discussing. First, as mentioned in section 5.1, to address the measurement errors in the moving year using the census data (and to further address the potential selection in moving time), our main analysis relies on including family fixed effects and comparing between siblings. While Chetty and Hendren (2018a) showed their main results are robust with and without family fixed effects, their main findings do not include family fixed effects. Second, individuals in our sample are eighty years to a decade older than individuals analyzed by Chetty and Hendren (2018a), and we focus on children with at least one immigrant parent. The sample and contextual differences mean that their findings may not necessarily apply to us.

Before estimating the causal effect of ethnic neighborhoods from Equation (7), we first try to replicate the main result of Chetty and Hendren (2018a) using the historical censuses 1900–1940. Specifically, we replace $LQ^o - LQ^d$ with $\overline{OccScore^o} - \overline{OccScore^d}$—the origin-destination difference in county average occupation score—in our main specification, Equation (7). We then estimate the effect of childhood exposure to high-income neighborhoods (as measured by the average occupation score of adults) by comparing children who moved from different origin neighborhoods to the same destination. The outcome of interest is a child's occupation score in early adulthood (at ages 25–30).

Table 2 summarizes the result replicating Chetty and Hendren (2018a) using the historical census data. To draw a more direct comparison to Chetty and Hendren (2018a), the first two columns of Table 2 include both children from immigrant families (which makes our main sample of analysis) and children from native families in the same cohort. The next two columns include only children from immigrant families. We find qualitatively consistent results with Chetty and Hendren (2018a). A longer childhood exposure to a high-income neighborhood (county) leads to a higher income (occupation score) as an adult. This significant positive effect is observed in the full sample and the immigrant sample, and with and without family fixed effects, while the magnitude of the effect varies across samples and specifications. While there are differences in data and context between our work and that of Chetty and Hendren (2018a), we take this replication result as a validation of our application of Chetty and Hendren's identification strategy to the historical census data and to our setting.

5.4 Results (*Preliminary*)

In this draft, we show results from estimating Equation (7) using a sample pooling ethnic groups accounting for 95% of the immigrant population in 1910.[17] To do so, we replace the birth year by mover cohort fixed effects in Equation (7) with the ethnic group by birth year by mover cohort fixed effects ($\alpha_{g,sm}$). We also replace the origin-destination pair fixed effects with this term interacted with ethnicity: $\alpha_{g,od}$. Note that the family fixed effects are already ethnic group-specific.[18] Some of the smaller groups can have very high LQ values for a few counties since a relatively small county population can lead to a large share when the national population of a group is small. We therefore exclude outliers in $\Delta LQ \times Exposure^o$ larger than the 99th percentile (positive outliers) or smaller than the 1st percentile (negative outliers).

The first column of Table 3 shows the results for occupation score when including ethnic group by origin-destination county fixed effects, but not family fixed effects. The coefficient is positive but small at 0.0086. In Column (2) we add family fixed effects, thus estimating the full specification shown in Equation (7), and the coefficient doubles in magnitude to 0.0161. For the sample in Column (2), the mean exposure to the origin county is 10 years and the mean change in LQ for movers to a lower LQ destination county is 1.8. For these values, the implied effect is $0.016 \times 1.8 \times 10 = 0.29$. The average occupation score is 19.72, thus the effect is about 1.5% of the average occupation score. This is a modest effect, but economically significant if it persists through one's lifetime.

As noted earlier, the distribution of LQ varies significantly across groups and the ratio itself can become quite large for small groups, which may make it difficult to interpret the coefficient in Columns (1) and (2). Due to these issues, we create two alternative measures of the change in ethnic neighborhood concentration. The first is simply the change in LQ decile, $\Delta\widetilde{LQ}$, calculated as the decile of the origin neighborhood LQ minus the decile of the destination neighborhood.[19] In the second measure we define an ethnic county ($Nbhd = 1$) as a place with an $LQ > 1.16$, which is the 99th percentile of the LQ distribution for natives across counties. We then define a new variable $\Delta Nbhd$ which takes a value of 1 if a family moved from an ethnic county to a non-ethnic county, and a value of -1 for moves from a non-ethnic county to an ethnic county. If the origin county type is equal to the destination county type, then $\Delta Nbhd = 0$.[20] We then multiply $Nbhd$ times the exposure to the origin and use this as the regressor. This variable essentially truncates the change in LQ to make it more comparable across groups. For specifications using these alternative measures, we don't drop outliers in LQ.

In Columns (3) and (4) of Table 3, we see a similar pattern of coefficients increasing in magnitude once family fixed effects are added. Most families move across neighborhoods that vary by at most one decile, such that the 25th percentile of $\Delta\widetilde{LQ}$ is a decrease of one decile and the 75th percentile is an increase of one decile. For a child whose destination LQ is one decile lower than the origin LQ and who was exposed to the origin for 10 years, the implied effect from column (4) is a $0.0095 \times 10 = 0.095$ increase in occupation score as an adult. In columns (5) and (6) we show the results using the neighborhood change measure. For a family moving to a destination county with a lower LQ, a child who was exposed to ten more years in the origin county than their younger sibling would have an occupation score 0.588 points higher, or 3.0% of the mean occupation score. Across all three measures of ethnic neighborhood concentration in Table 3, we see an increase in the magnitude of the coefficient when including family fixed effects. There are two possible

[17] These are Germany, Ireland, Canada, England, Sweden, Italy, Russia, Poland, Norway, Austria, Scotland, Czechoslovakia, Denmark, Hungary, Switzerland, Mexico, France, and the Netherlands.

[18] There are some cases where a child has two immigrant parents from different countries. Dropping these observations has little effect on our estimates.

[19] We calculate the decile based upon the distribution of LQ values across all ethnic groups in a given year. We have also tried an alternative that calculates the decile within an ethnic group, which leads to similar coefficients in our estimation results.

[20] We plan to examine alternative cut-off values, following Dai and Schiff (2023).

explanations for this difference. First, there could be a notable selection effect by family in making moving decisions. By making within-family comparisons between siblings, we recover the causal place effects. Second, there could be a mismeasurement of moving year in the census data, and adding family fixed effects helps address the problem.

The occupation score is equal to zero if an individual is unemployed and thus early ethnic neighborhood exposure could increase occupation score by raising the likelihood of being employed (extensive margin), increasing the income for individuals who would always be employed (intensive margin), or some combination of the two effects. In Table A1 we show results for employment alone (extensive margin) by defining the dependent variable as an indicator for $occscore > 0$ and multiplying this by 100 so that coefficient is measured in percentage points. We also run separate specifications where the dependent variable is occscore but the sample is limited to individuals with $occsore > 0$ (intensive margin). We find positive effects on both margins, but larger magnitudes (and smaller standard errors) for the employment indicator. This suggests that ethnic neighborhood exposure increases occupation score mostly by raising the likelihood of employment.

In Table 4 we examine how exposure to ethnic neighborhoods affects the county of residence twenty years later. In columns 1 and 2 the outcome of interest is the LQ of the county the children live in as adults, ages 20–35. Using the average LQ change of 1.8 as before, the coefficient in column 2 implies that a ten year exposure leads to a $0.0018 \times 1.8 \times 10 = 0.03$ point increase in LQ, a very small effect. Generally, there is little variation in the adult locations of siblings: roughly 90% of siblings live in the same county as adults (as measured in the third consecutive census). Additionally, this estimate could be affected by extreme values in the LQ of the adult location. Thus in columns 3-6 we use our other measures of LQ, both as independent variables and to measure the ethnic concentration of the adult location. The largest coefficient, that of column 5, suggests that ten additional years of exposure to an ethnic neighborhood as a child increases the probability the child lives in an ethnic neighborhood as an adult by 3.3%.

In Appendix Table A2 we try an alternative measure of ethnic concentration not related to LQ: the change in the ethnic fraction of the county, where ethnic fraction is county ethnic population divided by total county population. We find a similar pattern of coefficients (positive effects that increase significantly in magnitude when including family FE), with somewhat smaller magnitudes.[21] However, if we include both the LQ measure and the ethnic fraction measure in the same regression with family fixed effects, then the coefficient on ethnic fraction becomes small and insignificant while the coefficient on LQ exposure is quite similar to the estimates in column 2 of Tables 3 and 4.[22]

Lastly, in Tables 5 we show results of estimating Equation (7) for individual groups. In particular, we choose three of the largest sending countries: Germany (Columns 1-2), Italy (Columns 3-4), and Russia (Columns 5-6). All columns include family fixed effects. Not surprisingly, there is significant heterogeneity across groups. We find a substantially larger positive effect on occupation score for Italians and Russians than in the pooled sample, but a smaller and statistically insignificant effect for Germans. For county LQ as an adult, we find a larger effect for Germans than in the pooled sample, but small and insignificant results for Italians and Russians. We plan to explore reasons for this heterogeneity in future versions of this paper.

[21] The coefficient on ethnic fraction by exposure is 0.878 in column 2. The average change in ethnic fraction for cases where the origin ethnic fraction is larger than the destination is 0.023, thus assuming 10 years of exposure, the effect on occupation score is $0.878 * 0.023 * 10 = 0.202$. The corresponding effect using LQ was 0.29.

[22] We also tried a simple measure using the change in the ethnic population of the county, but nearly all results were tiny in magnitude and statistically insignificant (available upon request).

6 Conclusion

We investigate how childhood exposure to ethnic neighborhoods affects immigrant children's economic and spatial assimilation. Using full count historical census 1900–1940 and linking individuals across censuses, we estimate the causal effect of childhood exposure to ethnic neighborhoods on income and location choices in early adulthood, following Chetty and Hendren (2018a)'s empirical design. We find that increased exposure to ethnic neighborhoods tends to increase adult income (measured by occupation score) and the likelihood of living in an ethnic neighborhood as an adult, but by modest magnitudes. Looking across groups, we find significant heterogeneity with larger neighborhood effects on occupation score for Italians and Russians and only an effect on adult location for Germans.

The results shown in this draft are illustrative, but still preliminary. One important next step is to conduct the analysis using more precise measures of ethnic neighborhoods. To our best knowledge, none of the existing literature on historical immigration has tried to precisely identify neighborhoods using clustering methods. One likely reason is that although every individual is assigned an enumeration district in the census, spatial (GIS) data on these historic enumeration districts has not been available, making it impossible to know which enumeration districts are contiguous. However, the Urban Transition Historical GIS Project of Brown University has recently released spatial data that we can use for our project.[23] For the 1930 and 1940 censuses we have geocoded enumeration districts, and even geocoded streets, for 40 major cities. Additionally, Shertzer, Walsh, and Logan (2016) have made available enumeration district spatial data for 10 of the largest US cities, for every census from 1900–1930. These cities accounted for a significant portion of the national immigrant population. With these datasets we may be able to define neighborhoods as clusters of enumeration districts.

In addition, we plan to extend our analysis to include a richer set of outcomes, such as education, employment, language ability, and marriage. These outcomes provide additional measures to understand how immigrants and their children assimilated into the new country.[24] We also plan to investigate the childhood exposure effect of ethnic neighborhoods further into one's adulthood and into the next generation. This requires matching individuals across more census years and across generations.

[23] The Urban Transition Historical GIS Project can be accessed here.

[24] The availability of different outcome variables varies by census year. For example, the census did not record a detailed measure of education until 1940.

References

ABRAMITZKY, R., L. BOUSTAN, E. JÁCOME, AND S. PÉREZ (2021): "Intergenerational mobility of immigrants in the United States over two centuries," *American Economic Review*, 111(2), 580–608.

ABRAMITZKY, R., L. P. BOUSTAN, AND D. CONNOR (2020): "Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office," *NBER Working Paper 27372*.

ABRAMITZKY, R., L. P. BOUSTAN, AND K. ERIKSSON (2014): "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration," *Journal of Political Economy*.

ABRAMITZKY, R., L. P. BOUSTAN, AND O. GIUNTELLA (2025): "Enclaves and Assimilation in the Age of Mass Migration: Evidence from Ethnic Catholic Churches," *NBER Working Paper*.

ALBA, R. D., J. R. LOGAN, B. J. STULTS, G. MARZAN, AND W. ZHANG (1999): "Immigrant groups in the suburbs: A reexamination of suburbanization and spatial assimilation," *American Sociological Review*, pp. 446–460.

ALLEN, J. P., AND E. TURNER (1996): "Spatial patterns of immigrant assimilation," *The Professional Geographer*, 48(2), 140–155.

ASLUND, O., P.-A. EDIN, P. FREDRIKSSON, AND H. GRONQVIST (2011): "Peers, Neighborhoods, and Immigrant Student Achievement: Evidence from a Placement Policy," *American Economic Journal: Applied Economics*, 3(2), 67–95.

BORJAS, G. (1992): "Ethnic Capital and Intergenerational Mobility," *Quarterly Journal of Economics*.

——— (1995): "Ethnicity, Neighborhoods, and Human-Capital Externalities," *The American Economic Review*, pp. 365–390.

——— (2000): "Ethnic enclaves and assimilation," *Swedish Economic policy review*, 7(2), 89–122.

BRITTO, D., A. D. A. FONSECA, P. PINOTTI, B. SAMPAIO, AND L. WARWAR (2022): "Intergenerational mobility in the land of inequality," .

CHETTY, R., AND N. HENDREN (2018a): "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *The Quarterly Journal of Economics*, 133(3), 1107–1162.

——— (2018b): "The impacts of neighborhoods on intergenerational mobility II: County-level estimates," *The Quarterly Journal of Economics*, 133(3), 1163–1228.

CHETTY, R., N. HENDREN, AND L. F. KATZ (2016): "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *American Economic Review*, 106(4), 855–902.

CUTLER, D., E. GLAESER, AND J. VIGDOR (2008): "When are ghettos bad? Lessons from immigrant segregation in the United States," *Journal of Urban Economics*, 63(3), 759–774.

DAI, T., AND N. SCHIFF (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.

——— (2014): "Neighborhood quality and labor market outcomes: Evidence from quasi-random neighborhood assignment of immigrants," *Journal of Urban Economics*, 79, 139–166.

DANZER, A. M., C. FEUERBAUM, M. PIOPIUNIK, AND L. WOESSMANN (2022): "Growing up in Ethnic Enclaves: Language Proficiency and Educational Attainment of Immigrant Children," *Journal of Population Economics*, 35(3), 1297–1344.

DEUTSCHER, N. (2020): "Place, Peers, and the Teenage Years: Long-Run Neighborhood Effects in Australia," *American Economic Journal: Applied Economics*, 12(2), 220–249.

EDIN, P., P. FREDRIKSSON, AND O. ÅSLUND (2003): "Ethnic enclaves and the economic success of immigrants—Evidence from a natural experiment," *The Quarterly Journal of Economics*, 118(1), 329–357.

ERIKSSON, K. (2019): "Ethnic Enclaves and immigrant outcomes: Norwegian immigrants during the Age of Mass Migration," *European Review of Economic History*.

FOUKA, V., S. MAZUMDER, AND M. TABELLINI (2022): "From immigrants to Americans: Race and assimilation during the Great Migration," *The Review of Economic Studies*, 89(2), 811–842.

LERSCH, P. M. (2013): "Place stratification or spatial assimilation? Neighbourhood quality changes after residential mobility for migrants in Germany," *Urban Studies*, 50(5), 1011–1029.

LOGAN, J., W. ZHANG, AND R. ALBA (2002): "Immigrant enclaves and ethnic communities in New York and Los Angeles," *American Sociological Review*, pp. 299–322.

PORTES, A., AND R. G. RUMBAUT (2014): *Immigrant America, 4th edition*. University of California Press.

SHERTZER, A., R. P. WALSH, AND J. R. LOGAN (2016): "Segregation and neighborhood change in northern cities: New historical GIS data from 1900-1930," *Historical Methods*.

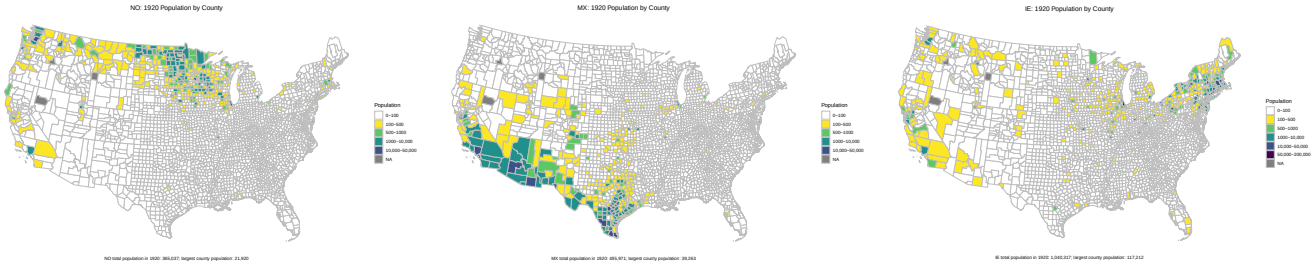
VOGIAZIDES, L., AND G. K. CHIHAYA (2020): "Migrants' long-term residential trajectories in Sweden: persistent neighbourhood deprivation or spatial assimilation?," *Housing Studies*, 35(5), 875–902.

WARD, Z. (2020): "The Not-So-Hot Melting Pot: The Persistence of Outcomes for Descendants of the Age of Mass Migration," *American Economic Journal: Applied Economics*.

WESSEL, T., R. ANDERSSON, T. KAUPPINEN, AND H. S. ANDERSEN (2017): "Spatial integration of immigrants in Nordic cities: The relevance of spatial assimilation theory in a welfare state context," *Urban Affairs Review*, 53(5), 812–842.

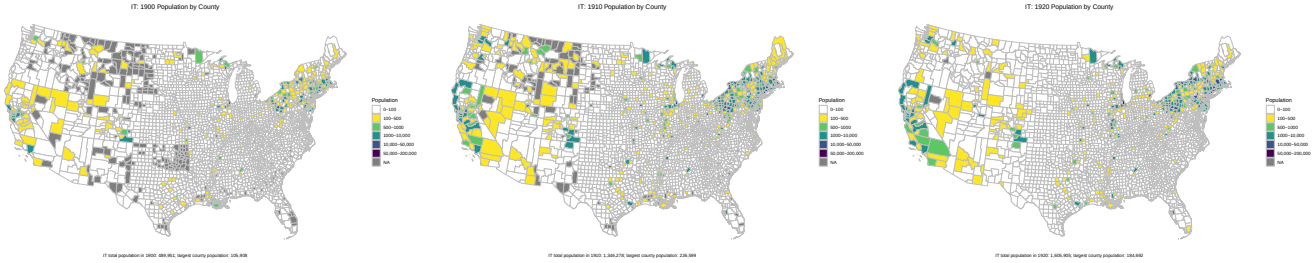
Figures and Tables

Figure 1: Norwegian, Mexican, and Irish County Populations in 1920



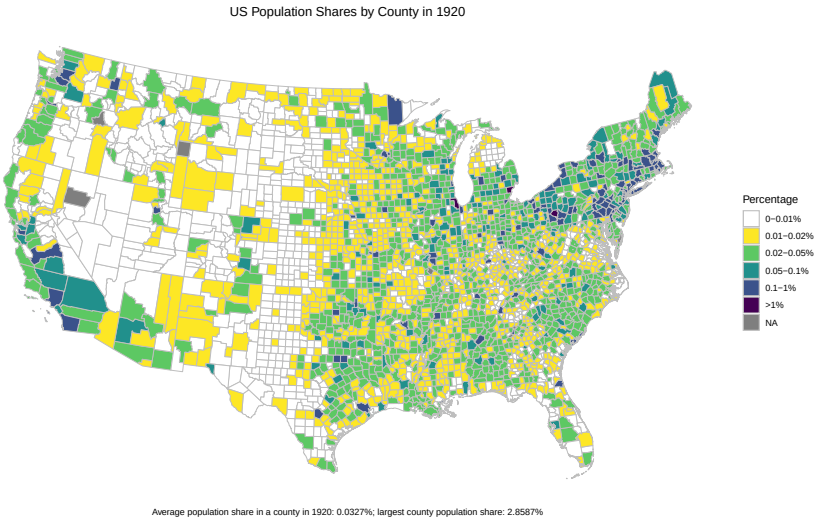
(a) Norwegians (b) Mexicans (c) Irish

Figure 2: Italian County Populations: 1900-1920

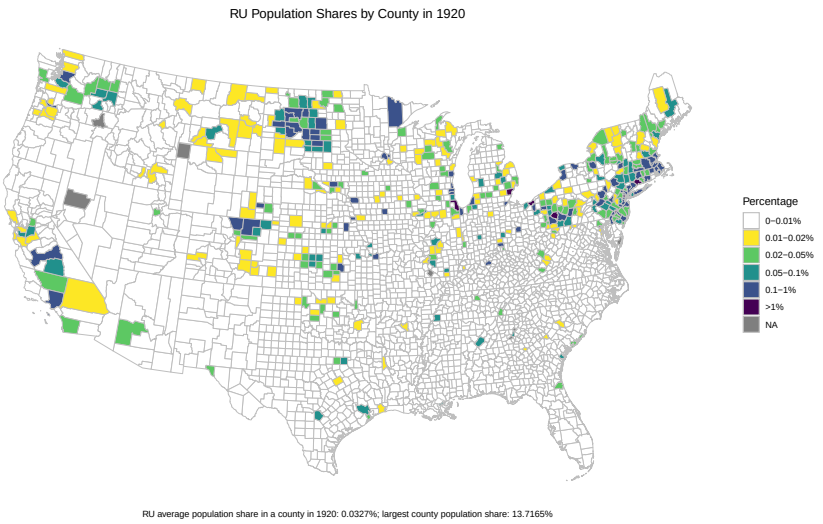


(a) 1900 (b) 1910 (c) 1920

Figure 3: Calculation of County-level LQs: Russians, 1920



(a) Total population (denominator)



(b) Russian population (numerator)

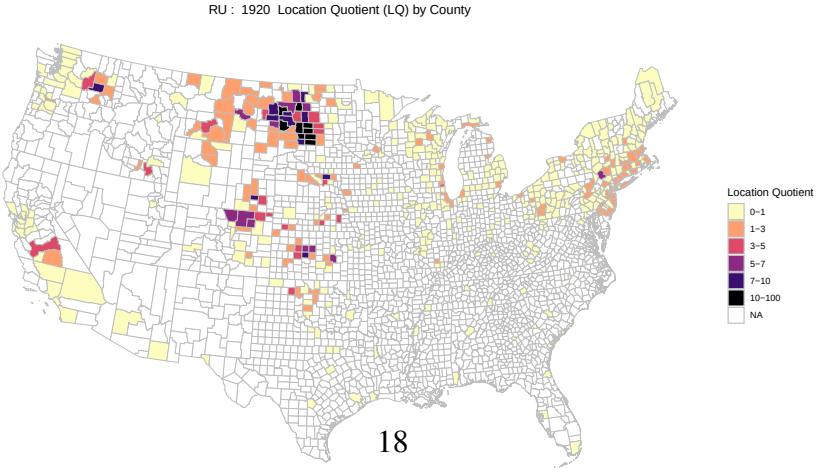
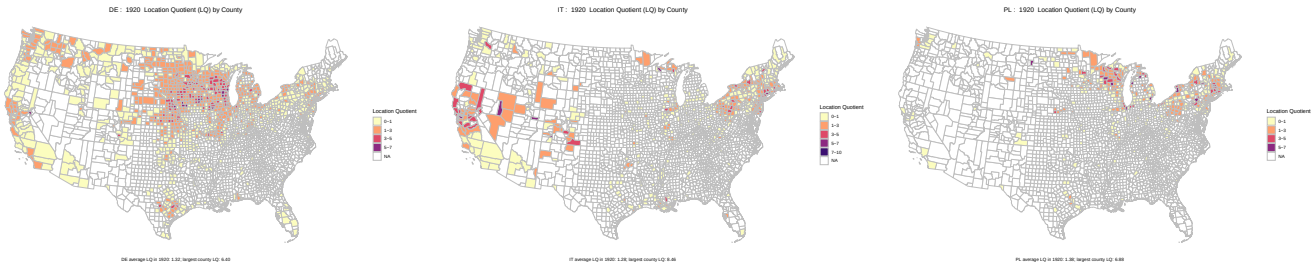
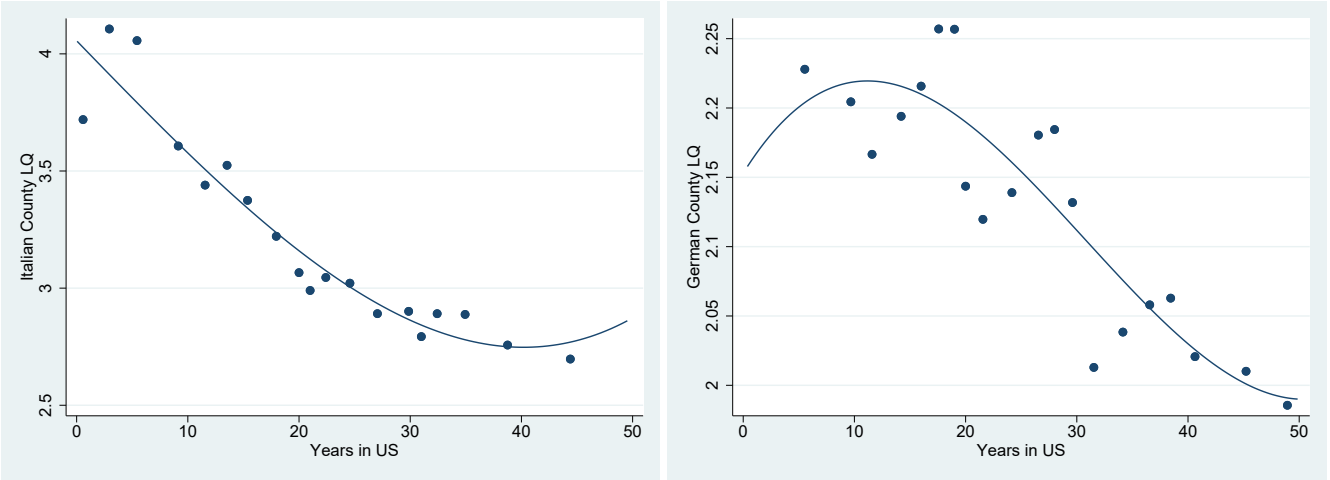


Figure 4: German, Italian, Polish County LQs in 1920



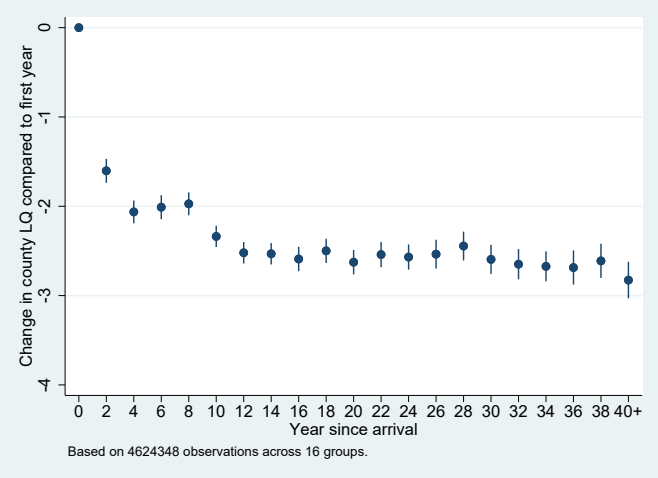
(a) Germans (b) Italians (c) Polish

Figure 5: Location Quotient (LQ) vs US Duration: Bin-scatter plots

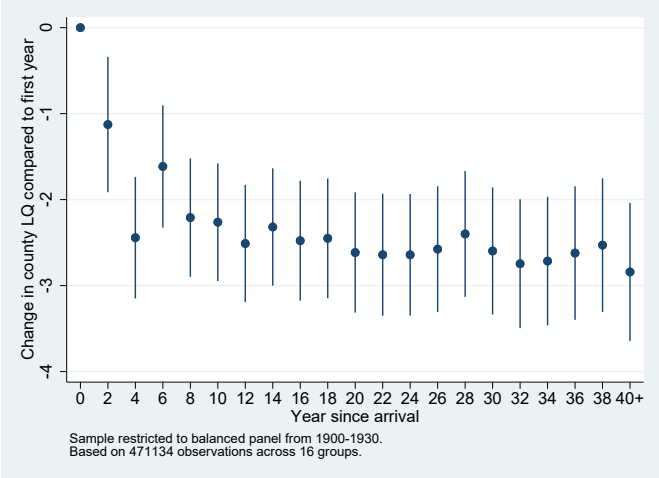


(a) Italians (b) Germans

Figure 6: County Location Quotient (LQ) vs US Duration; Pooled across Groups

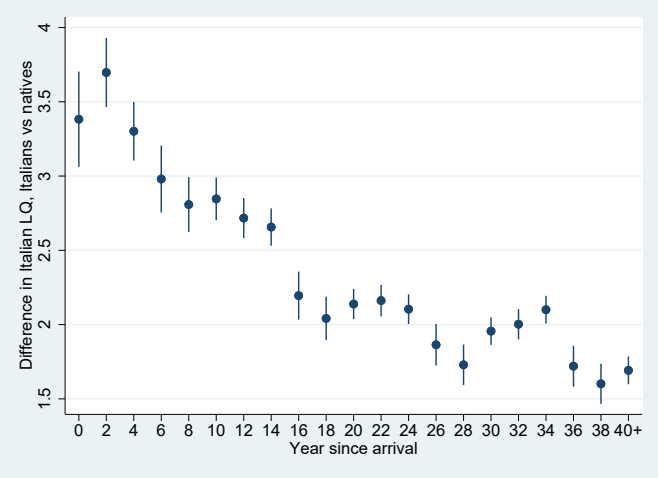


(a) Cross-section

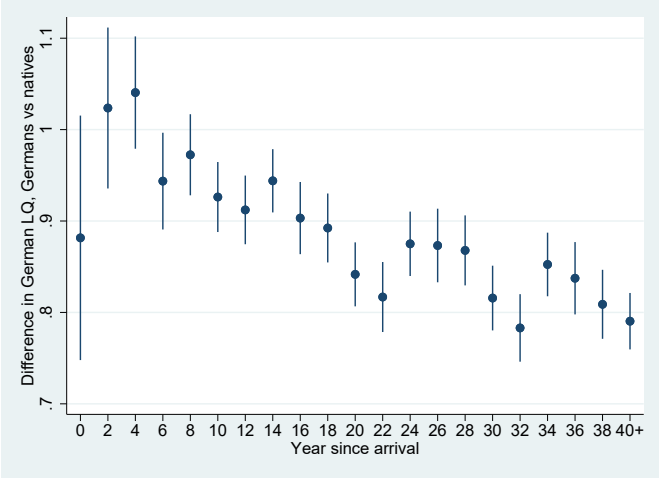


(b) Panel

Figure 7: County LQ, immigrants vs natives



(a) Italians



(b) Germans

Table 1: The Correlation between Changes in LQ and County Covariates

	(1) ΔLQ	(2) ΔLQ	(3) ΔLQ	(4) ΔLQ	(5) ΔLQ	(6) ΔLQ	(7) ΔLQ
$\Delta Occscore$	-0.012						-0.084***
	(0.015)						(0.031)
$\Delta SpeakEnglish$		-28.485***					-16.557***
		(3.440)					(3.144)
$\Delta Literacy$			-25.172***				-3.176
			(4.328)				(3.003)
ΔAge				-0.420***			0.242***
				(0.039)			(0.065)
$\Delta ForeignShare$					12.793***		11.109***
					(0.785)		(0.506)
$\Delta Population$						-0.002***	-0.010***
						(0.001)	(0.001)
Observations	1312495	1312495	1312495	1312495	1312495	1312495	1312495
Adj. R-sq	0.280	0.309	0.290	0.283	0.311	0.280	0.324

Notes: The sample pools different ethnic groups. The dependent variable ΔLQ is the origin-destination difference in county LQ. The independent variable is the origin-desitnation difference in a set of county covariates, including average occupation score, share of people speaking English, literacy rate, average age, share of foreign born, and working-age population (20-65). All specifications include ethnic group-birth year-move cohort fixed effects. Standard errors are clustered by ethnic group-birth year-move cohort.

Table 2: Replicating Chetty & Hendren using Historical Census Data

	(1) Occscore	(2) Occscore	(3) Occscore	(4) Occscore
$\Delta\overline{OccScore} \times Exp^o$	0.0122***	0.0048***	0.0085***	0.0113***
	(0.0003)	(0.0011)	(0.0006)	(0.0023)
Observations	2,884,701	1,536,982	752,084	470,569
R-sq	0.09	0.56	0.10	0.54
Clusters	2,006,560	658,841	475,477	193,962
FE-level	Dest. County	Family	Dest. County	Family
Sample	Full	Full	Immigrant Family	Immigrant Family

Notes: The outcome is occupation score (measured in 1950 occupation median income). $\Delta\overline{OccScore}$ is the origin-destination difference in county average occupation score. Exp^o is the years of exposure to the origin county. Columns (1) and (2) use the full sample of native and immigrant families, and Columns (3) and (4) use only immigrant families. Columns (1) and (3) include destination county fixed effects, and Columns (2) and (4) include family fixed effects. All specifications include birth year-move cohort fixed effects. Standard errors are clustered at the family level.

Table 3: Pooled Group Results: Occupation Score

	(1) Occscore	(2) Occscore	(3) Occscore	(4) Occscore	(5) Occscore	(6) Occscore
$\Delta LQ \times Exp^o$	0.0086***	0.0161***				
	(0.0020)	(0.0036)				
$\Delta\widetilde{LQ} \times Exp^o$			0.0048***	0.0095***		
			(0.0015)	(0.0027)		
$\Delta Nbhd \times Exp^o$					0.0048	0.0588***
					(0.0095)	(0.0160)
Observations	511860	346496	540988	369144	540988	369144
R-sq	0.25	0.54	0.26	0.55	0.26	0.55
Clusters	309262	144397	324642	153331	324642	153331
Family FE		YES		YES		YES

Notes: The sample pools different ethnic groups. The outcome is occupation score (measured in 1950 occupation median income). ΔLQ is the origin-destination difference in county LQ; $\Delta\widetilde{LQ}$ is the change in the LQ decile (decile of origin minus decile of destination). $\Delta Nbhd = 1$ if a family moved from an ethnic county ($LQ > 1.16$) to a non-ethnic county while $\Delta Nbhd = -1$ if a family moved from a non-ethnic county to an ethnic county; otherwise, $\Delta Nbhd = 0$. Exp^o is the years of exposure to the origin county. Columns (1), (3), and (5) include ethnic group by origin county by destination county fixed effects; that is, ethnic group, origin county, destination county tuples. In columns (2), (4), and (6) we include family fixed effects, which absorb all variation in ethnicity and origin-destination pairs. All specifications also include ethnic group by birth year by move cohort fixed effects. Standard errors are clustered at the family level. Extreme outliers in $\Delta LQ \times Exp^o$ are dropped from columns (1) and (2), but kept in all other columns.

Table 4: Pooled Group Results: County LQ 20 Years Later

	(1) LQ t+20	(2) LQ t+20	(3) $\widetilde{LQ}$ t+20	(4) $\widetilde{LQ}$ t+20	(5) ethNbd t+20	(6) ethNbhd t+20
$\Delta LQ \times Exp^o$	0.0012***	0.0018***				
	(0.0003)	(0.0004)				
$\Delta\widetilde{LQ} \times Exp^o$			0.0005***	0.0008***		
			(0.0001)	(0.0001)		
$\Delta Nbhd \times Exp^o$					0.0033***	0.0022***
					(0.0003)	(0.0002)
Observations	511860	346502	540988	369144	540988	369144
R-sq	0.92	0.98	0.90	0.98	0.87	0.97
Clusters	309262	144400	324642	153331	324642	153331
Family FE		YES		YES		YES

Notes: The sample pools different ethnic groups. The outcome in columns (1) and (2) is the LQ of the county where an immigrant child lived 20 years after the initial census year (adult location). ΔLQ is the origin-destination difference in county LQ; $\Delta\widetilde{LQ}$ is the change in the LQ decile (decile of origin minus decile of destination). $\Delta Nbhd = 1$ if a family moved from an ethnic county ($LQ > 1.16$) to a non-ethnic county while $\Delta Nbhd = -1$ if a family moved from a non-ethnic county to an ethnic county; otherwise, $\Delta Nbhd = 0$. Exp^o is the years of exposure to the origin county. In columns (3) and (4) the dependent variable is the LQ decile of the adult location while in columns (5) and (6) it's an indicator for whether the adult location was an ethnic county. Columns (1), (3), and (5) include ethnic group by origin county by destination county fixed effects; that is, ethnic group, origin county, destination county tuples. In columns (2), (4), and (6) we include family fixed effects, which absorb all variation in ethnicity and origin-destination pairs. All specifications include ethnic group by birth year by move cohort fixed effects. Extreme outliers in $\Delta LQ \times Exp^o$ are dropped from columns (1) and (2), but kept in all other columns.

Table 5: Individual Group Results: Germans, Italians, Russians

	(1)	(2)	(3)	(4)	(5)	(6)
	Occscore	LQ 20+	Occscore	LQ 20+	Occscore	LQ 20+
$\Delta LQ \times Exp^o$	0.0110	0.0025***	0.0287*	-0.0004	0.0253***	0.0007
	(0.0115)	(0.0004)	(0.0163)	(0.0004)	(0.0064)	(0.0005)
Observations	73630	73630	37025	37025	41998	41998
R-sq	0.58	0.97	0.47	1.00	0.53	0.99
Clusters	30484	30484	14943	14943	17833	17833
Group	DE	DE	IT	IT	RU	RU

Notes: The table presents results for three of the largest ethnic groups: Germans (Columns 1-2), Italians (Columns 3-4), and Russians (Columns 5-6). The outcome is the occupation score or the LQ of the county where an immigrant child lived 20 years after the initial census year. ΔLQ is the origin-destination difference in county LQ. Exp^o is the years of exposure to the origin county. All specifications include family fixed effects, and ethnic group-birth year-move cohort fixed effects. Standard errors are clustered at the family level. Extreme outliers in $\Delta LQ \times Exp^o$ are dropped.

A Appendix Exhibits

Figure A1: 1900-1940 Censuses: Linked Observations

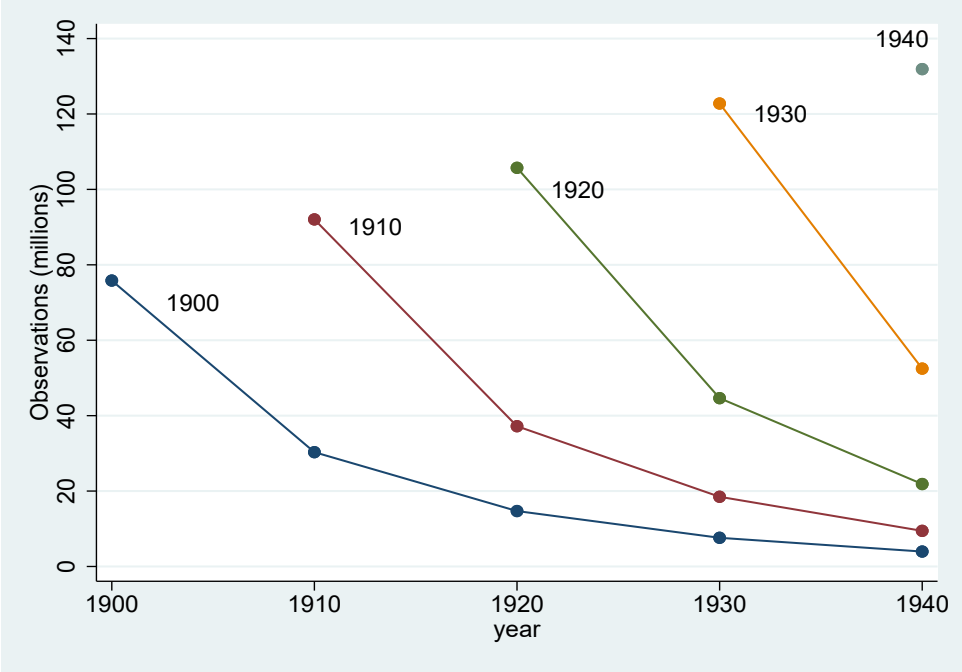


Table A1: Occupation Score and Employment

	(1) Occscore	(2) Employed	(3) Employed Occscore	(4) Occscore	(5) Employed	(6) Employed Occscore
$\Delta LQ \times Exp^o$	0.0086***	0.0354***	0.0010	0.0161***	0.0645***	0.0060**
	(0.0020)	(0.0058)	(0.0017)	(0.0036)	(0.0118)	(0.0028)
Observations	511860	511860	411407	346502	346502	257312
R-sq	0.25	0.19	0.35	0.54	0.49	0.64
Clusters	309262	309262	264021	144400	144400	110296
Family FE				YES	YES	YES

Notes: The sample pools different ethnic groups. The outcome in columns (1) and (4) is the occupation score, same as in Table 3. In columns (2) and (5) we the dependent variable is a binary indicator for whether the occscore>0, which indicates employment, multiplied by 100 ("employed" is measured in percentage points). In columns (3) and (6) the dependent variable is the occupation score given that occscore>0, or the occscore conditional on being employed. ΔLQ is the origin-destination difference in county LQ while Exp^o is the years of exposure to the origin county. Columns (1)-(3) include ethnic group by origin county by destination county fixed effects; that is, ethnic group, origin county, destination county tuples. In columns (4)-(6) we include family fixed effects, which absorb all variation in ethnicity and origin-destination pairs. All specifications also include ethnic group by birth year by move cohort fixed effects. Standard errors are clustered at the family level. Extreme outliers in $\Delta LQ \times Exp^o$ are dropped.

Table A2: Alternative Measure of Ethnic Concentration: Ethnic Fraction

	(1)	(2)	(3)	(4)	(5)	(6)
	Occscore	Occscore	Occscore	LQ t+20	LQ t+20	LQ t+20
$\Delta Eth\% \times Exp^o$	0.3716**	0.8784***	-0.3401	0.1768***	0.1127***	-0.0048
	(0.1605)	(0.2766)	(0.4403)	(0.0307)	(0.0155)	(0.0519)
$\Delta LQ \times Exp^o$			0.0195***	-0.0002		0.0019**
			(0.0057)	(0.0005)		(0.0009)
Observations	511860	346502	346502	511860	346502	346502
R-sq	0.25	0.54	0.54	0.92	0.98	0.98
Clusters	309262	144400	144400	309262	144400	144400
Family FE		YES	YES		YES	YES

Notes: The sample pools different ethnic groups. The outcome is occupation score (measured in 1950 occupation median income) or the LQ of the county where an immigrant child lived 20 years after the initial census year. ΔLQ is the origin-destination difference in county LQ; $\Delta Eth\%$ is the ethnic fraction in the origin county (ethnic pop/total pop) minus the ethnic fraction in the destination county. Columns (1) and (4) include ethnic group by destination county fixed effects, columns (2), (3), (5), and (6) include family fixed effects. All specifications include ethnic group-birth year-move cohort fixed effects. Standard errors are clustered at the family level. Extreme outliers in $\Delta LQ \times Exp^o$ are dropped.