Life after crossing the border:
Assimilation during the first Mexican mass migration

May 2021*

David Escamilla-Guerrero
University of St Andrews

Edward Kosack
Xavier University

Zachary Ward
Baylor University and NBER

Abstract: The first mass migration of Mexicans to the United States occurred in the early twentieth century: from smaller pre-Revolutionary flows in the 1900s, to hundreds of thousands during the violent 1910s, to the boom of the 1920s, and then the bust and deportations/repatriations of the 1930s. Using a new linked sample of males, we find that the average Mexican immigrant held a lower percentile rank, based on imputed earnings, than US-born whites near arrival. Further, Mexicans *fell* behind in the following decade. Mexican assimilation was not uniquely slow since we also find that the average Italian immigrant fell behind at a similar rate. Yet, conditional on geography, human capital, and initial percentile rank, Mexicans had a slower growth rate than both US-born whites and Italians. Mexican assimilation was also remarkably constant throughout various shocks, such as violence in Mexico, migration policy change in the United States, and the Great Depression. We argue that Mexican-specific structural barriers help to explain why Mexican progress was slow and similar across this tumultuous period.

JEL Codes: J15, J61, J62, N31, N32

Keywords: assimilation, immigration, Mexico, mobility

* david.escamilla-guerrero@st-andrews.ac.uk, School of Economics and Finance, University of St Andrews, Castlecliffe, The Scores, St Andrews, KY16 9AZ, Scotland. kosacke@xavier.edu. Department of Economics, Williams College of Business, Xavier University, ML 1212, 3800 Victory Parkway, Cincinnati, OH 45207, USA. Zach.A.Ward@gmail.com. Department of Economics, Hankamer School of Business, Baylor University, One Bear Place 98001, Waco, TX 76798, USA. Part of this paper was previously circulated under "The Uneven Advance of Mexican Americans Prior to World War II". We thank audiences at Xavier University, the 2017 NBER DAE Summer Institute, the 2017 Asian and Australasian Society of Labour Economists Conference, and the 2018 ASSA. Thanks to Lee Alston and Katherine Eriksson who helped Zachary Ward gain access to the full-count census data. Also thanks for Ariell Zimran for sharing data on Italian heights.

“Cuando era yo niño a mi me contaban
que los que pizcaban hacían furor,
compraban expreses, carros y caballos,
gallinas y gallos de un grande valor.
Es pura mentira, a mi no me pasó.”

“When I was a boy, they used to tell me
that those that picked cotton did well.
They used to buy trucks, cars and horses,
hens and roosters of great value.
It’s all a lie, it didn’t happen to me.”

—Mexican immigrant song (ca. 1920s)[1]

Over the course of the 20th century, the largest flow of migrants to the United States, by far, was from Mexico.[2] This paper documents the assimilation of Mexicans during the early stages of this migration between 1900 and 1940. Throughout this tumultuous period, what had been a small flow turned into a mass movement as the American Southwest boomed, deadly conflict in Mexico raged, and immigration restrictions via quotas on Europe increased the demand for Mexican labor. The rapid increase in migration was met with an equally rapid decrease as the Great Depression and a nativist backlash led to the deportation and repatriation of hundreds of thousands of Mexicans and Mexican Americans. Therefore, in addition to documenting early stages of assimilation for the major source country of the 20th century, this paper sheds light on how migrant assimilation is associated with key shocks to immigration, such as sharp changes in policy (e.g., the immigration quotas), violent upheaval in the source country (e.g., the Mexican Revolution), or economic downturns (e.g., the Great Depression).

To understand whether the early twentieth-century Mexican experience was unique, we contrast it with the experience of another important group: Italians. Historical Italians are often compared to modern-day Mexicans because both groups arrived relatively unskilled (Abramitzky

[1] Adapted from song of the Cotton Harvest (Dickey 2006).
[2] Between 1900 and 1997, approximately 5.6 million Mexicans were observed to enter the United States. In contrast, there were 4.5 million Italians, 2 million Germans, 1 million Irish and 1 million Scandinavians (Carter et al. 2006, Ad106-221).

et al. 2019, Huntington 2004, Perlmann 2005). Yet, little is known about the relative economic assimilation of Mexican and Italian immigrant contemporaries when they arrived in the same historical context.[3] This comparison allows us to uncover how Mexican immigrants fared relative to a group that faced some (but not all) of the same challenges and opportunities, and therefore helps to isolate Mexican-specific barriers to progress.

We find that Mexicans arrived in lower ranked jobs than US-born whites and that this gap *widened* with more years of stay. That is, Mexican immigrants experienced negative (or reverse) assimilation.[4] After imputing earnings based on occupation, country of birth, and census division, and using percentile ranks to measure the placement in the distribution, we find that Mexican immigrants fell behind US-born whites by 2.7 percentiles after one decade. This estimate comes from new linked data for 1900-1929 male arrivals, which allow us to accurately estimate migrant assimilation without bias from selective return migration (Borjas 1985, Abramitzky et al. 2014). Accounting for return migration is especially important for Mexicans since temporary migration was common (Gratton and Merchant 2015, Kosack and Ward 2014). While the average Mexican fell behind US-born whites, we find that the average Italian also fell behind by 2.5 percentiles, although the initial gap with US-born whites was smaller for Italians (12 percentiles) than for Mexicans (35 percentiles). The finding that Italians fell behind US-born whites is consistent with Collins and Zimran (2020, Table C.1), who use linked data between 1900 and 1930.

Despite the large shocks to Mexican migration during this period, we find that the assimilation rate was strikingly consistent across Mexican arrival cohorts. This pattern provides

[3] See Smith (2006, Figures 10 and 11) for an intergenerational comparison of the broader groups of Latinos and Europeans. Our results differ by focusing on within-generational outcomes and the narrower groups of Mexicans and Italians. Wildsmith et al. (2003) compare intermarriage rates across historical Italians and Mexicans.
[4] We use the term "assimilation" to refer to the widening of closing of economic gaps with the US-born after more years of stay.

suggestive evidence that barriers to upward mobility limited variation in Mexican outcomes, despite changes in the U.S. business conditions, immigration policy shifts, or changes to selection induced by violence in Mexico. The compression of Mexicans in a limited set of jobs and the lack of upward movement out of these jobs may reflect a ceiling on progress such that Mexicans were unable to break into the broader labor market.

Besides barriers to mobility, other explanations for slow Mexican progress include seclusion in the American Southwest, high levels of residential segregation, and low levels of human capital (Gratton and Merchant 2015). However, observable differences in geography or human capital do not explain the slower growth rate for Mexican immigrants. In fact, we find that the gap between Mexicans and other groups *widens* when controlling for observable characteristics (or the "conditional assimilation rate" (Collins and Zimran 2020)). While Mexicans fell behind US-born whites by three percentiles on average, they fell behind by 17 percentiles when compared with those starting in the same location, with the same literacy rate and English fluency rate, and the same initial rank. Further, conditioning on observables causes a gap to appear between Mexicans and Italians, with Mexicans falling behind Italians by nine percentiles between censuses. The difference in conditional assimilation rates is largely due to the initial rank: while low-ranked Italians and US-born whites were able to improve on their low starting position, Mexicans were unable to do so at the same rate.

Many factors could contribute to the difference in conditional assimilation rates across Mexicans and Italians. Selection may play a role if Mexicans were less productive than Italians based on unobservable characteristics (Borjas 1987). It is unlikely that selection explains the entire gap, and, in fact, there is evidence that Mexicans were *positively* selected from the source whereas Italians were *negatively* selected (Escamilla-Guerrero 2020, Kosack and Ward 2014, Spitzer and

Zimran 2018). Another potential explanation for slower growth is that Mexican migration was more circular due to its proximity to the United States, where Mexicans had less incentive to invest in US-specific human capital. This concern is alleviated by the fact that we focus on permanent migrants who remained at least 10 years; nevertheless, after controlling for English acquisition, we still find a gap between Italians and Mexicans.

We argue that Mexican-specific barriers help to explain the differences in assimilation rates across Mexicans and Italians after controlling for residential segregation, networks, human capital, and other factors. Ethnic animus appears to have been greater against Mexicans than Italians. For instance, Mexican Americans were deported en masse during the Great Depression, while Italians were not; the 1930 Census separated Mexicans into a different race category since they were viewed as "non-white," while Italians never were; Mexican lynchings were common, about three-fourths the lynching rate of Black victims; college student surveys rated Mexicans as less favorable than Italians (Catron 2020). While we cannot show causal evidence that such animus limited Mexican progress, the qualitative evidence suggests that Mexican-specific hostility played a role in their limited upward mobility.

A main limitation of the analysis is that economic assimilation can only be observed when an immigrant upgrades his occupation or moves to a higher-paying location, but not when he increases his earnings within occupation. This problem is common to most mobility studies based on historical US census data (e.g., Abramitzky et al. 2014, Collins and Zimran 2019, Minns 2000), and may lead us to mismeasure wage convergence with the US-born (Inwood et al. 2019). The direction of this bias is unclear; while we miss any gains within occupation from increased US-specific human capital, severe downturns (like the Great Depression) may more strongly impact migrant earnings than their occupations (Inwood et al. 2016). Similarly, the limited variation in percentile ranks we observe across Mexican cohorts may mask greater variation in earnings.

This paper adds to our knowledge about the economic mobility of Mexicans in the United States; in particular, our *intra*generational analysis complements the *inter*generational analysis in Kosack and Ward (2020).[5] While both papers find that Mexicans had a large gap with US-born whites, we find that there was divergence in the first generation throughout the lifecycle, as opposed to stagnation across multiple generations. Negative assimilation goes against expectations since the overwhelming majority of the literature finds that immigrants catch up, partially because they acquire United States-specific human capital (Borjas 2015, Chiswick 1978, Chiswick and Miller 2011, Lubotsky 2007). Indeed, we show explicitly that Mexicans fell behind despite large gains in English fluency and residential integration with the US-born. Moreover, our decomposition of first-generation progress sheds light on common arguments for why Mexicans did not move up (Gratton and Merchant 2015); for instance, we find that geography (i.e., location in the Southwest or residential segregation) does not strongly influence the Mexico/US gap, but the initial occupation does (Collins and Zimran 2020). Besides focusing on intragenerational assimilation, our paper differs from Kosack and Ward (2020) since our data covers all early 20th century cohorts, while Kosack and Ward (2020) only examine pre-Revolutionary cohorts. Therefore, our data allows us to document how Mexican assimilation changes during the major shocks of the early 20th century (e.g., Mexican Revolution, US policy changes, Great Depression).

Our main contribution is to document initial assimilation patterns of what would become the most important migration flow of the 20th century (Borjas 2007), but we also contribute to the

[5] Kosack and Ward (2020) examine intergenerational mobility with 1880, 1910 and 1940 cross-sectional data and linked data between 1910 and 1940. We examine intragenerational mobility with linked data between 1910-1920, 1920-1930 and 1930-1940.

broader literature on assimilation. A long literature argues that selection into migration can help to explain earnings gaps across US-born and immigrants (Chiswick 1978, Borjas 1987). First, we find that Mexican assimilation was not faster than Italian assimilation despite more positive selection for Mexicans. We also find that Mexican assimilation was constant throughout the period and so had little association with early 20th century shocks. For example, arrival records suggest that there was a rapid increase in "cohort quality" of Mexican arrivals between 1914 and 1915, yet we are unable to detect the same increase in the census data. This pattern contrasts with the standard finding that refugees start with larger gaps with US-born but have faster gains after arrival (e.g., Brell et al. 2020, Cortes 2004).[6] If barriers to upward mobility from labor market discrimination or ethnic animus are strong enough, then migrant outcomes will be compressed despite changes in selection. Barriers to upward mobility are consistent with other evidence that Mexicans and US-born whites operated in different labor markets in the early 20th century, such that Mexican outflows did not positively affect low-skilled whites (Lee et al. 2019).

I. A brief overview of the rise of Mexican mass migration in the early 20th century

Mexican immigration changed dramatically in the early 20th century due to a mix of push and pull factors (see Figure A1 for changes in the Mexican-born population over time). Employment opportunities in agriculture, mining, and railroads increased as the American Southwest boomed (Cardoso 1980, Clark 1908, Gratton and Merchant 2015, Kuntz 1995, Taylor 1929). As more, mainly unskilled, laborers from the north and central plateau of Mexico arrived, migration networks developed, which further increased migration (Henderson 2011, López-Alonso 2007, Oñate 1991, Rosenzweig 1965, Escamilla-Guerrero 2019, Morales 2016). Some

[6] Collins and Zimran (2019) find that refugee Irish during Great Famine held lower-wealth occupations and pre-famine Irish arrivals. Also, see Braun and Dwenger (2020) on the assimilation of displaced migrants in Germany after World War II.

migrants were recruited by American employers as part of the *enganche* system, where wages were offered in advance and transportation costs were covered (Brass 1990, Durand 2016, Escamilla-Guerrero 2020). However, since most immigrants were thought to be temporary, US officials did not pay much attention to Mexican arrivals.[7]

Mexican immigration increased rapidly during the 1910s due to one of the deadliest conflicts in world history: the Mexican Revolution. During the Revolution, about 1.4 million died and another 350,000 fled across the border (McCaa 2003, Table 1). The United States government expected that refugees would return soon after arrival, yet many stayed in search of employment (1913 Annual Report of Commissioner General of Immigration, p. 241). In parallel, World War I led to a surge in agricultural prices and a shortage of labor. As a response, the US Secretary of Labor issued orders to exempt Mexicans from the literacy test and head tax first implemented in 1917. This became the basis of the first Bracero Program, under which over 80,000 Mexicans arrived in the United States to work mainly in the fields (Scruggs 1960, Woodruff and Zenteno 2007). This exemption lasted until 1921 and the open-door policy effectively ended.

Mexican immigration continued to boom in the 1920s as unskilled labor from Europe was halted by the Immigration Acts of 1921 and 1924. No quotas were imposed on Mexico and so thousands moved northward to cities like Chicago to replace excluded Europeans (Abramitzky et al. 2019, Innis-Jimenez 2013). The growing number of Mexicans and their entry into new places outside of the Southwest led to a nativist backlash (Hoffman 1976). Unauthorized entry was criminalized, the Border Patrol was established to control areas between entry stations, and

[7] For example, the Dillingham Commission 1907-1910 has relatively scarce information on the Mexican immigrant (Benton-Cohen 2011).

officials debated whether they should place a quota on Mexican arrivals (Reisler 1976).[8] Yet, immigration further intensified with the Cristero War (1926-29), an anti-Catholic armed conflict that led to another exodus of Mexican refugees.

Following the boom of the 1920s came the bust of the 1930s, and with it the extensive repatriation and deportation of Mexicans. Deportations were pursued by local and state governments to improve labor market conditions for natives (Lee et al. 2019). Overlooking their immigration status, about 350 to 400 thousand Mexicans repatriated or were deported (Gratton and Merchant 2013, Verduzco 1995).[9] In the aftermath, the Mexican-born population fell by 40 percent, and those who managed to stay may have seen their assimilation process limited or interrupted.

Discrimination and racial prejudice against Mexicans

Throughout the booms and busts of the Mexican migration flow, Mexicans faced increasing hostility in the United States that affected different aspects of their lives. Animosity against Mexicans manifested in mob violence (lynchings) that persisted until the Great Depression.[10] Tensions about labor, land ownership, and market control were usually catalysts for Mexican lynchings. Native-born whites and Europeans used mob violence to intimidate and expel Mexicans from mining camps in Arizona and California, as they resented competing with pre-settled and more experienced Mexican miners (Morefield 1956). In Texas and New Mexico,

[8] The Border Patrol also worked in areas up to a hundred miles away from the border, sometimes arresting immigrants without a warrant (Ngai 2003).
[9] Among these repatriates were second-generation immigrants, i.e. US citizens. Balderrama and Rodríguez (2006) estimate the return flow at more than one million.
[10] Carrigan and Webb (2013) estimate that about 158 Mexicans were lynched from 1900-1928 in the American Southwest. Figure A2 depicts the spatial distribution of Mexican lynchings. However, the number of victims of racial violence soars when considering other acts of aggression. Muñoz Martinez (2018) estimates that the number of dead could be as many as several thousand in Texas only.

Anglos used lynchings to expropriate lands of Mexicans and to control commercial routes traditionally dominated by Mexican families (Carrigan and Webb 2013). In addition to mob violence, vigilante groups and state police used extralegal violent procedures to persecute and repress Mexicans. For example, in 1918, a group of Texas Rangers massacred all Mexican American men and boys of Porvenir—a border town in Presidio County—for the unconfirmed crime of sympathizing with bandits (Muñoz Martinez 2018).

Policies were also used to segregate Mexicans from Anglos and maintain hierarchies of race and class. In 1917, a quarantine policy mandating the bathing of Mexicans crossing the border was enacted to combat an outbreak of typhus in the El Paso region (Pierce 1917). Although there were no new cases of typhus after four months, the detention and forced-bathing of Mexicans, Mexican Americans, and other ethnic minorities continued until the 1950s (Mckiernan-González 2012). The quarantine campaign was intertwined with a eugenic discourse that portrayed Mexicans as biologically inferior peasants who were carriers of deadly diseases (Stern 1999, p.73). Similar eugenic language was used in education to justify the segregation of Mexican students. During the 1920s, local school boards characterized Mexican students as slow learners who needed special instruction in separate schools, as they would impede the academic progress of Anglo children (Menchaca and Valencia 1990). Ethno-racial hierarchy was also present in higher education. Information from college surveys (1925-1959) shows that Mexicans and African Americans were the furthest in perceived social distance from native-born whites, while Western European students (English, Irish, Germans) were closer (Catron 2020).

Ethnic prejudice against Mexicans was also present in the labor markets. Mexican immigrants were paid lower wages than US-born whites and Europeans of the same class (Clark 1908, Gamio 1930, Guerin-Gonzalez 1994). From the nativist's perspective, wage differentials and occupational segregation were justified since Mexicans had limited potentialities that were racially determined. In addition to economic discrimination, Mexican-American unions were not always recognized, and many Mexicans were excluded from other unions. Moreover, few Mexican Americans acted as union delegates (Garcia 1982). The absence of labor protection mechanisms facilitated mass deportations of Mexicans driven by radical unions and political leaders arguing for the protection of "white men's jobs" (Cardoso 1980, Reisler 1976). Despite the lack of effective union building, Mexicans organized strikes demanding a closing of pay differentials between Anglo and Mexican workers. Yet, like in the Bisbee Deportation of 1917, strikes were often dissolved using extralegal violence supported by local governments (O'Neill 1993).

Overall, during the early twentieth century, mob violence, violent policing, social segregation, and economic discrimination likely influenced the economic progress of Mexicans in the United States. However, the extent to which these structural barriers affected Mexican progress relative to US-born whites and other immigrants remains unclear. In this paper, we are the first to empirically document how Mexicans improved with more years of stay with a new dataset that tracks thousands from census to census.

II. Data

Linking Mexican immigrants.

To estimate Mexican immigrants' rate of economic assimilation, we build a linked dataset of males from the 1910, 1920, 1930, and 1940 complete-count United States Censuses, accessed at the National Bureau of Economic Research (Ruggles et al. 2020).[11] Note that females are not

[11] The censuses are the best available data during this period, but there are limitations. Primarily, there are issues of under enumeration, the extent of which for Mexican immigrants is unknown (Cardoso 1980, Hacker 2013). Therefore, our results only apply to the population of observed migrants. While selection into the census is unknown, those

included since they may change surname between censuses. Linked data are needed to accurately estimate economic assimilation since repeated cross-sections conflate selective return migration with assimilation (Abramiztky et al. 2014). To link data, we follow a machine-learning approach (Feigenbaum 2016). First, we hand linked a random sample of 2,000 Mexican immigrants across the 1910-1920 censuses. After keeping potential links with sufficient closeness in name, year of birth, and year of arrival, two researchers separately chose the best link. The matches were reviewed and agreed upon. We ran a probit to predict the most likely match out of a set of candidate matches and then used the predicted probabilities to choose the best matches for the entire census. The best match depends on variables that should be constant across censuses such as first name, last name, country of birth, year of birth, and year of arrival.[12] One concern is that the first name may change between censuses if immigrants Anglicized their names (e.g., Jose to Joe) (Biavaschi et al. 2017). To account for this, we Anglicize all names before linking, according to information from behindthename.com, which was collected by Alexander and Ward (2018).[13] We describe the full linking method in detail in Appendix B. Note that we also estimate negative assimilation for Mexicans when using automated linking methods from the Census Linking Project (Abramitzky et al. 2020), which does not Anglicize names; however, the conservative ABE method links about a third of the Mexicans we link in our preferred method.[14]

We do not link all Mexican immigrants between censuses but are interested in the rate of occupational upgrading for newly arrived immigrants. Therefore, we first draw males who arrived

---

enumerated are likely higher-skilled than the unobserved population, which reinforces our point that the Mexican/US-born white gap was large and widening over time. See Escamilla-Guerrero et al. (2021) for the data.

[12] We do not use the year of arrival for predicting the 1930-1940 link.

[13] Pérez (2017) does a similar progress for linking immigrants in Argentina.

[14] The conservative ABE method keeps only those with unique first name, last name, country of birth and year of birth combinations within a two-year birth window.

within the last ten years at first observation and who are between 21 and 40 years old.[15] We then link migrants to the next census; for example, we link 1900-1909 arrivals between 1910 and 1920. We do not wish to further link migrants a second time because linking rates are low (mostly <10 percent) and we do not want to lose observations. After linking, we keep only those who report an occupation in both years.[16] This leaves us with a sample of 3,874 Mexicans linked between 1910 and 1920, 10,771 between 1920 and 1930, and 8,766 between 1930 and 1940.

We compare Mexican occupational upgrading to that of Italians and US-born whites. The Italian dataset comes from Ward (2020) and the US-born white datasets are from Ward (2019, forthcoming).[17] All datasets were linked using the same machine-learning approach. Importantly, we set the positive predictive value ($PPV = \frac{\#truelinks}{\#matched}$) to be the same across all groups (0.90); the positive predictive value determines the number of false positives in the dataset, and setting it at 0.90 implies that about 10 percent of our matched sample are false positives (Feigenbaum 2016). Since the linked datasets for Mexicans, Italians, and US-born whites have the same PPV, the difference in occupational mobility across groups should not be due to a difference in false positives (Bailey et al. 2020). Yet, the trade-off for setting the PPV for all groups at 0.90 is that the true positive rate ($TPR = \frac{\#truelinks}{\#observations}$) differs across groups due to differences in name commonness, enumerator error, age heaping, death, and out-migration. We successfully link 9-11 percent of Mexicans forward to the next census, 4-6 percent of Italians, and 32-36 percent of US-

[15] We do not link those who arrived in the same year as the census since we do not observe the full arrival cohort.
[16] Some occupations in the full-count data are not-yet-classified (occ1950=979). We clean these occupation codes by merging the occupation string with the most common occ1950 code per string in the 1940 Census. We also clean obvious spelling errors, such as "clekr" for "clerk" and "engiener" for "engineer". After cleaning the 979 occupation strings, the bulk of the missing occupation codes are "N/A (blank)" (occ1950=999) (88 percent).
[17] Ward (forthcoming) builds the 1920-1930 and 1930-1940 data to estimate the internal migration premium. Ward (2019) builds the 1910-1920 data to estimate intergenerational mobility. Ward (2020) does not include immigrants from 1930 to 1940, so we follow the same method as described in Appendix B to link this group.

born whites. Despite low linking rates for migrants, the full-count data still contain 51,947 Italians and 11,013,388 US-born whites.

A major concern with the linked samples is that they are unrepresentative. To account for this issue, we weight our linked sample using inverse probability weights (Bailey et al. 2020).[18] Importantly, this weighting is done with respect to the second census (e.g., 1920 for the 1910-1920 link), because the first census includes both permanent and temporary migrants. We weight by each 5-year arrival cohort in case selection into the linked sample varies across cohorts. Of course, this method may not account for unobservable selection into the sample. To the extent that unobservable selection into the sample is similar for each group (US-born whites, Mexicans, and Italians), then the bias would be differenced out in the estimation.

Earnings score proxy.

Due to data limitations in censuses before 1940, we measure assimilation based on imputed earnings rather than with actual wages or income. Following Collins and Wanamaker (2017), we mostly impute earnings with wage income for wage workers in the 1940 Census. Appendix C describes the imputation in greater detail, but there are a few key points to note. First, since self-employed income is not in the 1940 Census, we impute it using information from the 1960 census, where we assume that total income over wage income by occupation is the same in 1940 as it was in 1960. We also add perquisites for farmers and farm laborers using information from the United States Department of Agriculture (see Appendix C), which increases their position in the earnings distribution. We then regress adjusted earnings (that includes estimated self-employment income

[18] We weight based on age fixed effects (5-year bins); broad occupation categories (the first digit of 3-digit occ1950 code); the ability to read and write (or have more than 8 years of education in the 1940 census), and region of residence (see Appendix B for more detail).

and perquisites) on occupation (3-digit code), census division, a quadratic in age, country of birth, country of birth interacted with division, occupation (1-digit code) interacted with country of birth, and occupation (1-digit code) interacted with region.[19] This process is similar to Abramitzky et al. (2021) and Saavedra and Twinam (2020). The mean predicted income by 3-digit occupation, census division and country of birth is the earnings proxy. Therefore, for example, every Mexican who was a laborer in the West South Central division is assigned the same earnings score.

Rather than using earning scores directly, our main results are for percentile ranks. Percentile ranks are calculated by birth cohort and year, which implicitly accounts for age and period effects. Thus, a change in rank across censuses occurs due to either a change in the earnings score or a change in the distribution of earnings scores. The main reason for using percentile ranks is that imputations from the 1940 census may not be accurate when applied to the 1910, 1920, and 1930 censuses (Collins and Zimran 2020). On the other hand, the mean income by occupation/division/country of birth may be more stable in the earnings distribution over time. We will later show that negative assimilation is robust to using a linear combination of our 1940 earnings score and the 1890 score from Sobek (1996) for the 1910, 1920, and 1930 censuses, which accounts for changes in an occupation's position in the income distribution over time.

Assimilation based on earnings score may fail to capture wage assimilation. However, because we allow scores to vary by country of birth rather than just occupation, we can more accurately estimate gaps across groups (Inwood et al. 2019). We will also show, however, that Mexicans did not upgrade occupational categories at the same rate as US-born whites (i.e., white

[19] Including a quadratic in age follows Abramitzky et al. (2021) and Saavedra and Twinam (2020) but does not matter for our results since we do not merge scores based on age. Regressing the score (collapsed to the occupation/division/country of birth level) without the quadratic on the score with the quadratic yields a 99.4 percent R-squared.

collar, farmer, semi-skilled and unskilled), which reaffirms our argument that Mexicans fell behind US-born whites in the early 20th century without any earnings imputations.

III. The economic assimilation of Mexican immigrants

The Mexicans in our sample are first observed far behind US-born whites and Italians in both percentile rank and earnings score (see Table 1). The gaps were large: Mexicans were 34 percentiles lower than US-born whites and 22 percentiles lower than Italians at first observation (i.e., the first census after arrival for immigrants).[20] Alternatively, Mexicans had earnings scores that were 54 log points (42 percent) lower than US-born whites and 43 log points lower than Italians.

The main reason why Mexicans started behind both groups is that most Mexicans worked in unskilled jobs (83 percent, see bottom of Table 1). In contrast, 33 percent of US-born whites and 66 percent of Italians worked in unskilled jobs. Besides unskilled work, 4 percent of Mexicans were farmers (owners or tenants), 7 percent were semi-skilled, and 6 percent were white collar. We present a detailed breakdown of Mexican occupations in Table 2, which shows that 39 percent of Mexicans in our sample were general laborers, 24 percent were farm laborers and 5 percent were miners.[21] This concentration in a handful of occupations could reflect that higher-paying jobs were unattainable for Mexicans. Table 2 also shows that Mexicans had lower percentile ranks *within* occupation. For example, the Mexican laborer was at the 14th percentile, while the US-born white laborer was at the 19th and the Italian laborer was at the 27th. This pattern is consistent with

[20] Figure A3 shows the distribution of Mexican percentile ranks, where Mexicans rarely made it above the 60th percentile and only 7 percent were above the 40th percentile.
[21] See Figure A4 for the cumulative density of the top 50 occupations by group. It shows that Mexicans were more concentrated in a few occupations than Italians or US-born whites.

field observations in the early 20th century that Mexicans were paid lower wages within occupation (Clark 1908).

It is slightly surprising that Italians held higher-skilled jobs than Mexicans since arrival records suggest both groups were similarly skilled in the early 20th century. The 1910 *Annual Report of the Commissioner General of Immigration* documents that the share of both Mexican and Italian immigrants with unskilled pre-migration occupations was very similar: 86 and 89 percent, respectively (see Figure 1A). The 1910 Census records, however, suggest that both groups downgraded occupations in the first year of arrival as 95 percent of the Mexican immigrants and 92 percent of the Italian immigrants were unskilled (see Table A1). However, this comparison is not perfect since the *Annual Report* and the census data do not perfectly match.[22] A more consistent comparison of Italian and Mexican human capital can be made based on the heights recorded in arrival records. In Figure 1B we plot the height distributions of Italians and Mexicans from samples of arrival records and show that Mexican immigrants were five centimeters taller than Italian immigrants.[23] To the extent that height is a proxy for human capital, one would expect that Mexicans ended up in better-paying jobs in the United States and that they might improve more rapidly with more years of stay.

However, Mexicans had *zero* growth in earnings scores over one decade from the first to the second observation; in fact, they fell behind by 1 log point (see Table 1). No gain is surprising,

[22] First, the *Annual Report* groups arrivals based on the fiscal year (July 1909-June 1910) while the Census records include those who arrived in 1909 and the start of 1910 (Jan 1909-April 1910). However, this issue applies to both Italians and Mexicans, so it may not influence the difference in unskilled across groups. Second, the *Annual Report* includes immigrants whose entry is observed, while the Census should include both observed and unobserved entrants. Unobserved entry was likely more common among Mexicans than Italians, which could bias the comparison across records. A future study could link arrival records to the Census to address these issues.

[23] The underlying data is a sample of 22-65 year-old males from 1907, 1908 and 1920 arrival records. These records were originally collected by Escamilla-Guerrero (2020), Kosack and Ward (2014), and Spitzer and Zimran (2018). We thank Ariell Zimran for generously sharing the height data with us.

especially since immigrants are expected to catch up due to human capital acquisition in areas like English fluency or knowledge about US job markets. This result suggests that opportunities for Mexican upward mobility were limited. Due to this lack of progress, Mexicans fell behind US-born whites by 10 log points or 3 percentiles. Occupational categories paint a more optimistic picture about Mexican progress: Mexicans transitioned out of unskilled occupations between censuses (from 83 to 76 percent) and entered white-collar, farming, or skilled work. However, these gains were small and did not translate into improvements in percentile ranks. Mexicans were not the only group with limited gains: Italians also dropped 3 percentiles between censuses. Therefore, Mexicans and Italians in our sample had similar growth rates, though Italians started in higher-skilled occupations.

A more detailed occupational category analysis confirms that Mexicans were less likely than US-born whites and Italians to escape unskilled jobs. For instance, Panel A of Figure 2 shows that for the group of white-collar workers at first observation, 72 percent of US-born whites, 45 percent of Italians, and 37 percent of Mexicans ended up in a white-collar job ten years later. Not only were Mexicans less likely to climb the occupational ladder, but they were also more likely to fall off: 63 percent white-collar Mexicans moved down to a less-skilled occupation 10 years later. Panels B, C, and D show where the white-collar Mexican workers ended up: 8 percent in semi-skilled jobs, 51 percent in unskilled jobs and 4 percent in farming; therefore, the white-collar to unskilled transition was the most common for white-collar Mexicans. The key occupation category is in panel C: no matter in which job category a Mexican started, the majority ended up in an unskilled job ten years later, at a higher rate than Italians or US-born whites.

Table 1 pools all arrival cohorts together, but a key feature of early 20th century Mexican immigration is that shocks to immigration from violence in Mexico, policy changes in the United

States, and the Great Depression led to big swings in the composition of migrant cohorts. Indeed, arrival records suggest that the skill composition of Mexicans varied substantially across year of arrival (see Figure 1A). The most dramatic shift occurred between 1912 and 1915, when the percent of unskilled dropped from 89 percent to 45 percent in only three years, before returning to 80 percent by the end of the decade. These years are associated with some of the most intense fightings during the Mexican Revolution and likely reflect high-skilled refugees. Figure 1A also shows that the composition became less unskilled at the end of the 1920s, which could be due to the Mexican Cristero War or the start of the Great Depression collapsing demand for unskilled work.

Yet this variation of skill in the arrival records does not show up in the census data. Instead, the census data report a remarkable consistency of outcomes across cohorts. Panel A of Table 3 reports percentile rank gaps (relative to US-born whites) for Mexican arrival cohorts split into five-year groups by census year. First, the percentile rank gap at first observation does not vary strongly across arrival cohorts. The 1910-1914 arrival cohort, which should include some of the high-skilled refugees observed in the arrival records, had a similar gap at first observation as other cohorts (36 percentiles – we will later show this result holds for the narrower cohort of 1914-1915). Furthermore, this cohort had a similar growth rate across censuses (negative 3.6 percentiles), which contrasts with modern-day research that suggests refugees have faster growth rates due to a stronger incentive to invest in US-specific human capital (Cortes 2004). On the other hand, it is possible that most refugees returned home by the 1920 census and thus do not appear in our census data. To test for selection into return migration, we follow Abramitzky et al. (2014) and compare growth rates in panel data (that include permanent migrants) to growth rates in repeated cross sections (that include temporary and permanent migrants). Our estimates suggest a *negative* selection into return migration for every arrival cohort, which is inconsistent with the story that most returnees who fled the Revolution were high-skilled refugees.[24] Once again, however, note that positively selected refugees may have already returned before the first observation in 1920.

Mexicans were not unique in their negative assimilation as pre-quota Italians also fell behind US-born whites. Panel B of Table 3 shows that the average pre-quota Italian fell behind US-born whites by 5 to 7 percentiles for 1900-1919 arrivals. Interestingly, the Italian cohorts who entered during the 1920s (and mostly after the 1921 Emergency Quota Act) no longer fell behind US-born whites but maintained their relative position. This difference could be because opportunities for upward mobility were more plentiful due to a decrease in competition with other migrants (Ward 2017). However, the initial gap widened for Italians and so the overall gap with US-born held at approximately 15 percentiles. We provide a graphical representation of the previous results in Figure 4.

The results so far show simple differences in means across censuses, but it is possible to predict how far Mexicans fell behind US-born whites with each additional year of stay via regression (Chiswick 1978).[25] Figures 3A and 3B show the assimilation profiles for occupational categories and percentile ranks, respectively. There was some convergence in occupational categories after 20 years of stay: for farmers, the predicted gap dropped from 20 to 13 percentage points; for unskilled, the gap converged from 53 to 48; and for semi-skilled, the gap reduced from

---

[24] Abramitzky et al. (2014, Figure 5) note that the repeated cross-sectional gap is a weighted average of the permanent migrant gap and temporary migrant gap. Back-of-the-envelope calculations suggest that return migrants were 1.8 to 4.4 percentiles lower than permanent migrants in the first census (see Table A2).
[25] That is, we estimate $y_{ict} = g(YearsInUS_{ict}) + \text{f}(\text{Age}_{ict}) + \gamma_c + \theta_t + \varepsilon_{ict}$. In this regression, the percentile rank is regressed on a quadratic in years in the United States, a quadratic in age, cohort of arrival fixed effects and year fixed effects. US-born whites are the excluded group for both the arrival cohort and years in the United States variables. A quadratic was chosen because it appeared to fit the binscatter relationship between percentile rank and years in the United States. Negative assimilation is not sensitive to using a quadratic or a quartic in years in the United States (See Figure A5).

13 to 11. Thus, about 60-95 percent of the initial occupational category gaps remained. For percentile ranks, Mexicans fell behind US-born whites by 7 percentiles (from -33 to -40) and Italians fell behind by 5 percentiles (from -10 to -15).

Finally, our finding of negative Mexican assimilation is robust to using different earning scores. For instance, instead of imputing earnings based on occupation/division/country of birth, we could impute based on (1) solely occupation, (2) occupation/division, or (3) occupation/country of birth (Inwood et al. 2019). Figure C1 shows that no matter which score one uses, Mexicans fell behind US-born whites. However, the gaps with US-born whites do vary by earnings score. Most importantly, the percentile-rank gap drops by about 40 percent when ignoring country of birth, which suggests that lower pay within occupation is a key reason for the wide gap with US-born whites. Another issue with earnings scores is that earnings imputed from the 1940 census may fail to capture changes in the relative status of occupations in 1910, 1920 or 1930.[26] However, negative assimilation still holds when using a linear interpolation of the 1890 occupational score from Sobek (1996) and our 1940 occupational score, but the ranked gaps are about 5 percentiles smaller (Figure A6). Finally, we also find negative assimilation when using the publicly available links from the Census Linking Project (Abramitzky et al. 2020) (Figure A7).

IV. Explaining the differences in levels and the differences in growth rates

We have shown that Mexicans fell behind US-born whites on average, but this pattern is the *unconditional* assimilation rate without controls for observable characteristics. It may be that differences in human capital, location, networks, or residential segregation explain the gap. To explore the economic gaps at first observation, we estimate:

[26] For example, clerical workers gained in relative standing over this period, which may bias inference since Mexicans, US-born whites, and Italians have different propensities of holding clerical jobs.

$$y_i = \gamma_0 + \gamma_1 Mexico_i + \boldsymbol{\gamma}_2 Italy_i + \Pi' X_i + \varepsilon_i \tag{2}$$

where we regress the first census's percentile rank on an indicator variable for Mexico and Italy. US-born white is the excluded group. We control for human capital (literacy, ability to speak English), location (urban, log of county population, county fixed effects), networks (log number of Mexicans in the county for Mexicans, log number of Italians for Italians, with US-born whites at zero), residential segregation (segregation from US-born), and census year.[27] Using data from Carrigan and Webb (2013), we also control for whether a lynching occurred in the county in the previous decade.[28]

Human capital differences explain some of the initial gap with US-born whites (Figure 5A). Controlling for literacy and the ability to speak English decreases the US-Mexico gap by one-fourth, from 35 to 26 percentiles. The remaining gap suggests that observable human capital differences do not entirely explain Mexicans' lower position in the earnings distribution. Of course, the human capital controls are limited; the ability to speak English was a judgment by the enumerator and literacy is the ability to read and write in any language. Better human capital controls may narrow the gap further, which we will discuss later.

While some may expect that location, such as living in the American Southwest or residential segregation, was a key reason for Mexicans holding low-paying jobs (e.g., Gratton and Merchant 2015), conditioning on location does not narrow the gap between Mexicans and US-born whites. In fact, the gap slightly widens from 26 to 29 percentiles. However, differences in

[27] Note that county fixed effects do not capture time-varying county characteristics, such as the log population of the county.

[28] These lynching data are taken from Carrigan and Webb (2013) and are recorded mainly from newspapers, government documents, and photographic records. However, the data mostly cover periods before 1910, so only about 5 percent of Mexicans in our linked dataset are in a county with a lynching. The lynching variable is zero for Italians and US-born whites.

urban status and residential segregation do appear to be important for the initial rank (see Table A3). Living in an urban area is associated with a 19-percentile rank increase, which reflects that urban areas had higher wages than rural ones (Boustan et al. 2014). Since Mexicans were more rural than Italians (52 percent v 84 percent), controlling for urban status narrows the gap between them. The estimate for residential segregation suggests that a standard deviation increase in residential segregation (0.40) is associated with a 3-unit decrease in rank; however, those who are not segregated are predicted to still be behind US-born whites. We also control for whether there was a lynching in the county, using data from Carrigan and Webb (2013), but do not find that Mexicans held lower percentile ranks relative to US-born whites in these counties. Of course, these associations for urban status, residential segregation, and lynching should not be interpreted as causal since location may be correlated with unobservable characteristics. Rather, we show that conditioning on covariates does not eliminate the Mexico/US-born white gap.

As opposed to the gap in the *initially observed rank* shrinking after conditioning on observables, the gap in the *growth rate* widens. Figure 5B shows estimates from regressing the decadal growth rate on mostly the same set of control variables:[29]

$$(y_{i,t} - y_{i,t-1}) = \gamma_0 + \gamma_1 Mexico_i + \gamma_2 Italy_i + \Pi' X_{i,t-1} + \varepsilon_i \qquad (3)$$

However, we do add a key new variable: the initial percentile rank. Collins and Zimran (2020) show that differences in where immigrants and US-born whites start in the distribution have a large impact on their growth rate in the following decade, since immigrants tended to start in high-upward mobility occupations like laborer or operative.

[29] The control variables for human capital and geography are based on the initial observable outcome in the first period. See Table A4 for underlying regression and controls.

Figure 5B shows that Mexicans fell behind US-born whites by 3 percentiles with no covariates. A main reason why the average Mexican fell behind by only 3 percentile ranks is that they were in a job with plenty of room for upward mobility. However, after comparing Mexicans to US-born whites starting off in the same position $(y_{i,t-1})$, Mexicans fell behind by 17 percentiles. Therefore, while Mexicans were in jobs with room for upward mobility, they did not actually move upward, unlike comparable US-born whites. This result is consistent with Collins and Zimran (2020), who show that historical European migrants were less likely to "work their way up" when compared to natives starting out with the same rank. However, controlling for initial rank is more impactful on Mexicans than Italians; while the Mexican gap widened from 3 to 17 percentiles, the Italian gap widened only from 3 to 8 percentiles.

Controlling for human capital and location do not explain why Mexicans fell behind US-born whites (see Figure 5B). Differences in human capital do not appear to matter much for the gap in the growth rate with US-born whites, suggesting that a lower literacy rate for Mexicans was not entirely to blame for falling behind. The unimportance of location is once again surprising since Mexican concentration in the American Southwest and residential segregation are cited as reasons for barriers to higher wages (Gratton and Merchant 2015). Yet, when compared to natives who resided in the same counties and after controlling for segregation, Mexicans are predicted to fall behind natives at 17 percentiles. Our main result that Mexicans fell behind US-born whites, conditional on observables, contrasts with Feliciano (2001, Table 1), who finds that Mexicans in 1910 *completely* closed the gap with US-born whites after five years of stay. When we mimic Feliciano's data structure and specification with the full-count 1910 census (which is representative of the population), we do not find convergence in percentile ranks. This difference does not appear to be due to the earnings score measure: we also do not find convergence of unskilled occupations (see Table A5). While convergence in wages may have occurred for the Dillingham Commission sample, it is unlikely that it occurred in the entire country.[30]

Gaps by arrival cohort

In this section, we turn to estimate whether conditional assimilation varied across Mexican arrival cohorts. Figure 6A shows that the conditional gap at first observation with US-born whites did vary.[31] After controlling for observable characteristics, the Mexico-US gap increased in magnitude across arrival cohorts, from negative 18 percentiles for the 1900-1904 cohort, to negative 22 percentiles for the 1910-1914 cohort, to negative 26 percentiles for the 1920-1924 cohort. It is unclear why the conditional gap grew across arrival cohorts. It may be that as opportunities increased in the early 20th century American Southwest, they were not available for new Mexican arrivals who started with similar skills. Another possibility is that the start of the Great Depression led to worse relative outcomes for 1920-1929 arrivals as observed in the 1930 Census. Either way, the results do not indicate that the refugee cohorts during the 1910s had a unique economic experience in the United States.

Finally, in Figure 6B, we show that the conditional assimilation of Mexican immigrants was similar for all cohorts. All cohorts fell behind by about 12-16 percentiles, which suggests that the experience in the United States was broadly similar across time. Once again, this result is surprising given the shifts in cohort quality, at least as observed in the *1910 Annual Report of the*

[30] A major difference between our approaches is data. Feliciano uses aggregated industry/ethnic data from a single cross-section (the 1911 Dillingham Commission), while we use individual-level panel data between 1910 and 1940. Our data also include all states and industries while the Dillingham Commission covers specific states and industries that hired Europeans. For example, Texas is not included in the Dillingham data, and neither are agriculture, construction, or services. Moreover, the Commission targeted large firms in their survey, leaving the representativeness of the data unclear (McGouldrick and Tannen 1977). Yet, a key advantage of the Commission data is that it records daily wages, which improves on earnings scores.
[31] We omit Italians from the regression.

Commissioner General of Immigration. In Table A7, we explicitly test whether narrower arrival cohorts had faster growth rates than others but find no evidence of different assimilation rates. For example, the higher-skilled 1914-1915 arrival group also fell behind by 14 percentiles, like most of the other 1900-1929 arrivals.[32]

V. Discussion of reasons for negative Mexican assimilation

One reason for a persistent gap in earnings growth score is that the human capital controls we use (literacy and English fluency) do not fully capture differences across Mexicans, US-born whites, and Italians. While not available in all of our data, a better human capital control exists in 1940: educational attainment. Educational gaps across groups were large; Mexicans in our sample had 4.0 years of education on average, while Italians had 5.7 years and US-born whites had 9.2 years. Controlling for years of education narrows the assimilation deficit from 17 percentiles to 13 percentiles (see Figure 7A), which suggests that imperfect human capital controls do explain some of the residual gap between Mexicans and US-born whites. However, a large gap remains and so the inability to perfectly measure human capital does not appear to explain the persistent results that Mexicans fell behind.

On a related note, Mexicans lagging behind US-born whites could be because of limited investment into US-specific human capital after arrival. Human capital acquisition is often viewed as the main reason why immigrants catch up to natives (Chiswick 1978); for instance, Borjas (2015) highlights the slowing rate of English acquisition as a key reason for why there is a slowing rate of economic assimilation for post-1980 arrival cohorts. This explanation is not satisfactory in the early 20th century since Mexicans did acquire English fluency, from 16 percent after one year

[32] We do not find differences for those who arrived during the most intense years of the Cristero War.

of stay to 51 percent after 20 years (see Figure A7).[33] However, differences in English acquisition may explain the Italian and Mexican gap since Mexicans acquired English fluency at a slower rate. In our regression, we controlled for the baseline level of fluency, but not for the acquisition across the decade, partially because fluency is not observed in 1940. Yet controlling for English acquisition in the 1910-1920 and 1920-1930 panels does not narrow the gap between Italians and Mexicans (see Figure 7B), which is consistent with a weak association between acquiring English fluency and improving earnings scores in the early 20th century (Ward 2020).

The fact that Mexicans fell behind observationally equivalent US-born whites, even after accounting for human capital differences, is consistent with unobserved Mexican-specific barriers to progress. Barriers to progress may also explain why arrival cohorts had similar rates of assimilation despite shifts in composition; for instance, if Mexican immigrants operated in a different labor market that was difficult to escape. While Mexicans certainly faced discrimination in the early 20th century, the contribution of such barriers to lagging progress and limited variation in assimilation across cohorts is difficult to precisely quantify. We can, however, lean on some empirical results that are consistent with this explanation. We show with our data that there is no upward movement in the raw earnings score (see Table 1), that there was an overwhelming concentration of Mexican immigrants in unskilled occupations (see Tables 1 and 2), and that almost no Mexican immigrants made it above the 60th percentile of earnings scores (see Figure A2). Moreover, we know that the large outflow of Mexicans during the repatriations of the 1930s did not improve the labor market outcomes for low-skilled natives in the same counties, suggesting

[33] Figure A8 also show that Mexicans spatially assimilated at a slower rate when using the county-level residential segregation measure from Eriksson and Ward (2019). Differences in human capital or location do not explain the difference in English acquisition or change in spatial assimilation (see Tables A8 and A9).

that even migrants and natives with the same skill level did not operate in the same labor market (Lee et al. 2019).

Another way to test for the importance of barriers, and especially those resulting from ethnic animus, is to estimate whether lynching was associated with slower growth for Mexicans, which we did in the regression reported in the previous section (Table A5). However, the point estimate will not be causal if lynching was endogenous. Indeed, lynchings often occurred due to competition over land and resources, perhaps because Mexicans were gaining status relative to US-born whites (Carrigan and Webb 2013). Consistent with this story, we estimate that the rate of assimilation in the previous decade was positively correlated with a lynching in the county (see Table A10, Column II). Moreover, US-born white percentile rank growth was slower in the decade prior to lynching. Given that Mexican ranks were trending up while US-born white ranks were trending down prior to lynching, the association between lynching and Mexican assimilation may be positively biased. Indeed, we estimate that lynching was positively associated with Mexican economic assimilation. It is highly unlikely that mob violence caused Mexican outcomes to improve. Rather, it may be that lynching was aimed to suppress Mexicans who were gaining status.[34]

[34] Table A10 reports the association between lynching and the decadal change in percentile rank. Besides endogeneity of lynching, another difficultly with identifying the impact of lynching with decadal census data is that most lynching occurred between censuses, and Mexicans were highly mobile between censuses. This pattern makes it unclear whether the Mexicans in our sample were still in the county when the lynching took place. Table A10, Column III splits the Mexican sample into those who remained in a lynching county at first and second observation, those who moved away from one and those who moved into one. The results show that there is a positive association with those moving away, no association for those staying, and a negative association for those moving in. This result is consistent with lynching driving away Mexicans who were moving up the earnings distribution, but it is difficult to pinpoint this explanation since we do not observe year of internal migration.

While we cannot pinpoint a causal effect of lynching, the fact that extrajudicial violence existed is evidence of severe ethnic animus against Mexicans. It is also notable that Mexicans fell behind US-born whites even in counties with no observable lynching. As we discussed in the historical background, Mexicans were also excluded from unions, persecuted by the Texas Rangers, faced prejudicial views from natives, singled out as a separate race in the 1930 census, and deported in large numbers during the Great Depression. Based on this qualitative evidence, it would be unsurprising that barriers to progress in the labor market caused Mexicans to lag US-born whites.

VI. Conclusions

We show that during the first wave of Mexican mass migration, Mexicans started off far behind US-born whites, and fell further behind in the decade after. Observable characteristics do not fully explain these gaps; in fact, controlling for characteristics *widens* the gap in the upgrading rate from census to census. We also find that Mexican arrival cohorts between 1900 and 1929, which varied in skill composition due to violence in Mexico and policy in the United States, all experienced similar outcomes in the United States. The results that assimilation rates were similar across arrival cohorts and that Mexicans fell behind observably similar Italians and US-born whites are consistent with Mexican-specific structural barriers limiting upward mobility.

Many compare modern-day Mexican immigration to historical Italian immigration and argue that both groups would experience similar outcomes if they were placed in the same context. When we do a systematic comparison of contemporaneous Italian and Mexican immigrants, we show that there is some evidence to support this claim, but also some evidence against it. Both groups had similar growth rates relative to US-born whites; however, conditional on geography, human capital, or starting jobs, important differences emerge across Mexicans and Italians. The evidence suggests that one should not be surprised if differences in economic assimilation appear across groups, especially if barriers to upward mobility vary across groups.

Bibliography

Abramitzky, Ran, Leah Boustan, and Dylan Connor. *Leaving the Enclave: Historical Evidence on Immigrant Mobility from the Industrial Removal Office*. No. w27372. National Bureau of Economic Research, 2020.

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration." *Journal of Political Economy* 122.3 (2014): 467-506.

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. "Do Immigrants Assimilate More Slowly today than in the past?." *American Economic Review: Insights* 2.1 (2020): 125-41.

Abramitzky, Ran, Leah Boustan, Elisa Jácome, and Santiago Pérez. *Intergenerational Mobility of Immigrants in the US over Two Centuries*. No. w26408. National Bureau of Economic Research, 2019.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James J. Feigenbaum, and Santiago Pérez. Automated linking of historical data. *The Journal of Economic Literature*, forthcoming

Abramitzky, Ran, Philipp Ager, Leah Boustan, Elior Cohen, and Casper W. Hansen. *The Effects of Immigration on the Economy: Lessons from the 1920s Border Closure*. No. w26536. National Bureau of Economic Research, 2019.

Alanis, Fernando Saúl. “No cuenten conmigo: La pólitica de repatriación del gobierno mexicano y sus nacionales en Estados Unidos, 1910—1928.” *Mexican Studies* 19, no. 2 (2003): 401-431.

Alesina, Alberto, and Eliana La Ferrara. "Who trusts others?" *Journal of Public Economics* 85, no. 2 (2002): 207-234.

Alexander, Rohan, and Zachary Ward. "Age at arrival and assimilation during the age of mass migration." *The Journal of Economic History* 78.3 (2018): 904-937.

Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. How Well Do Automated Methods Perform in Historical Samples? Lessons from US Historical Data. *The Journal of Economic Literature,* forthcoming.

Balderrama, Francisco E., and Raymond Rodríguez. *Decade of betrayal: Mexican repatriation in the 1930s*. UNM Press, 2006.

Benton-Cohen, Katherine. "Other Immigrants: Mexicans and the Dillingham Commission of 1907-1911." *Journal of American Ethnic History* 30, no. 2 (2011): 33-57.

Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique. "The economic payoff of name Americanization." *Journal of Labor Economics* 35.4 (2017): 1089-1116.

Borjas, George J. "Assimilation, changes in cohort quality, and the earnings of immigrants." *Journal of Labor Economics* 3.4 (1985): 463-489.

Borjas, George J. "Self-Selection and the Earnings of Immigrants." *The American Economic Review* (1987): 531-553.

Borjas, George J. "The slowdown in the economic assimilation of immigrants: Aging and cohort effects revisited again." *Journal of Human Capital* 9, no. 4 (2015): 483-517.

Brass, Tom. "The Latin American enganche system: Some revisionist reinterpretations revisited." *Slavery and Abolition* 11, no. 1 (1990): 74-103.

Braun, Sebastian T., and Nadja Dwenger. "Settlement location shapes the integration of forced migrants: Evidence from post-war Germany." *Explorations in Economic History* 77 (2020): 101330.

Brell, Courtney, Christian Dustmann, and Ian Preston. "The labor market integration of refugee migrants in high-income countries." *Journal of Economic Perspectives* 34.1 (2020): 94-121.

Cardoso, Lawrence A. 1980. *Mexican Emigration to the United States, 1897–1931*. University of Arizona Press.

Cardoso, Lawrence A. *Mexican Emigration to the United States 1897-1931: Socio-Economic Patterns*. University of Arizona Press (Tucson, AZ), 1980.

Carrigan, William D., and Clive Webb. "The lynching of persons of Mexican origin or descent in the United States, 1848 to 1928." *Journal of social history* (2003): 411-438.

Carrigan, William D., and Clive Webb. *Forgotten dead: mob violence against Mexicans in the United States, 1848-1928*. Oxford University Press, 2013.

Catron, Peter. 2021. "The Alien Citizen: Social Distance and the Economic Returns to Naturalization in the Southwest." *Social Problems* (forthcoming).

Chiswick, Barry R. "The effect of Americanization on the earnings of foreign-born men." *Journal of political Economy* 86.5 (1978): 897-921.

Clark, Victor S. "Mexican Labor in the United States." *U.S. Department of Labor Bulletin*. 78 (1908): 466-522.

Collins, William J., and Ariell Zimran. *Immigrants' Changing Labor Market Assimilation in the United States during the Age of Mass Migration*. No. w26414. National Bureau of Economic Research, 2020.

Collins, William J., and Ariell Zimran. "The economic assimilation of Irish Famine migrants to the United States." *Explorations in Economic History* 74 (2019): 101302.

Collins, William J., and Marianne H. Wanamaker. *Up from slavery? African American intergenerational economic mobility since 1880*. No. w23395. National Bureau of Economic Research, 2017.

Cortes, Kalena E. "Are refugees different from economic immigrants? Some empirical evidence on the heterogeneity of immigrant groups in the United States." *Review of Economics and Statistics* 86.2 (2004): 465-480.

Dickey, Dan. "Corridos y Canciones de las Pizcas: Ballads and Songs of the 1920s Cotton Harvests." *Western Folklore* 65, no. 1/2 (2006): 99-136.

Durand, Jorge. *Historia mínima de la migración México-Estados Unidos*. El Colegio de Mexico, 2016.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. "Ethnic enclaves and the economic success of immigrants—Evidence from a natural experiment." *The quarterly journal of economics* 118, no. 1 (2003): 329-357.

Eriksson, Katherine, and Zachary Ward. "Immigrants and Cities during the Age of Mass Migration" *Regional Science and Urban Economics*, forthcoming.

Eriksson, Katherine, and Zachary Ward. "The Residential Segregation of Immigrants in the United States from 1850 to 1940." *The Journal of Economic History* 79.4 (2019): 989-1026.

Eriksson, Katherine. "Ethnic enclaves and immigrant outcomes: Norwegian immigrants during the Age of Mass Migration." *European Review of Economic History* 24, no. 3 (2020): 427-446.

Escamilla-Guerrero, David, Kosack, Edward, and Ward, Zachary. Replication data for: Life after Crossing the Border: Assimilation during the First Mexican Mass Migration. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-05-14. https://doi.org/10.3886/E140543V1

Escamilla-Guerrero, David, and Moramay Lopez-Alonso. *Self-selection of Mexican migrants in the presence of random shocks: Evidence from the Panic of 1907*. Oxford Economic and Social History Working Paper No. 179, 2020.

Escamilla-Guerrero, David. "Revisiting Mexican migration in the Age of Mass Migration: New evidence from individual border crossings." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* (2020): 1-19.

Escamilla-Guerrero, David. *Initial Determinants of Mexican mass migration.* Mimeo, 2019.

Feigenbaum, James J. "Automated census record linking: A machine learning approach." (2016). *Mimeo*

Feliciano, Zadia M. "The skill and economic performance of Mexican immigrants from 1910 to 1990." *Explorations in Economic History* 38.3 (2001): 386-409
.

Ferrara, Andreas, and Price V. Fishback. *Discrimination, Migration, and Economic Outcomes: Evidence from World War I*. No. w26936. National Bureau of Economic Research, 2020.

French, Will J., Anna L. Saylor, George H. Hecke. "Mexicans in California. Report of Governor CC Young's Mexican Fact-Finding Committee " *Sacramento, CA: State of California Department of Industrial Relations* (1930).

Gamio, Manuel. *Mexican Immigration to the United States*. Arno Press, 1930.

García, Mario T. 1982. *Desert Immigrants: The Mexicans of El Paso, 1880-1920*. Yale University Press.

Gratton, Brian, and Emily Klancher Merchant. "An Immigrant's Tale: The Mexican American Southwest 1850 to 1950." *Social Science History* 39, no. 4 (2015): 521-550.

Gratton, Brian, and Emily Merchant. "Immigration, repatriation, and deportation: The Mexican-origin population in the United States, 1920–1950." *International Migration Review* 47.4 (2013): 944-975.

Hacker, J. David. "New estimates of census coverage in the United States, 1850–1930." *Social Science History* 37.1 (2013): 71-101.

Hall, Linda B. "El Refugio: Migración Mexicana a los Estados Unidos: 1910-1920." *Boletín del Instituto de Investigaciones Históricas* 8 (1982): p. 23-44.

Hatton, Timothy J. "The immigrant assimilation puzzle in late nineteenth-century America." *Journal of Economic History* (1997): 34-62.

Hatton, Timothy J., and Zachary Ward. "International Migration in the Atlantic Economic 1850-1940." In: Diebolt, Claude, Haupert Michael (eds) Handbook of Cliometrics. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40458-0_41-1

Huntington, Samuel P. *Who are we?: The challenges to America's national identity*. Simon and Schuster, 2004.

Innis-Jiménez, Michael. *Steel barrio: The great Mexican migration to south Chicago, 1915-1940*. Vol. 10. NYU Press, 2013.

Inwood, Kris, Chris Minns, and Fraser Summerfield. "Occupational Income Scores and Immigrant Assimilation. Evidence from the Canadian Census." *Explorations in Economic History* 72 (2019): 114-122.

Inwood, Kris, Chris Minns, and Fraser Summerfield. "Reverse assimilation? Immigrants in the Canadian labour market during the Great Depression." *European Review of Economic History* 20, no. 3 (2016): 299-321.

Kosack, Edward, and Zachary Ward. "El sueño americano? The generational progress of Mexican Americans prior to World War II." *The Journal of Economic History* (2020): forthcoming.

Kosack, Edward, and Zachary Ward. "Who crossed the border? Self-selection of Mexican migrants in the early twentieth century." *The Journal of Economic History* (2014): 1015-1044.

Lee, Jongkwan, Giovanni Peri, and Vasil Yasenov. *The Employment Effects of Mexican Repatriations: Evidence from the 1930's*. No. w23885. National Bureau of Economic Research, 2019.

Lubotsky, Darren. "Chutes or ladders? A longitudinal analysis of immigrant earnings." *Journal of Political Economy* 115, no. 5 (2007): 820-867.

McCaa, Robert. "Missing millions: The demographic costs of the Mexican Revolution." *Mexican Studies* 19.2 (2003): 367-400.

Mckiernan-González, John. 2012. *Fevered Measures: Public Health and Race at the Texas-Mexico Border, 1848-1924*. Duke University Press.

Menchaca, Martha, and Richard R. Valencia. 1990. “Anglo-Saxon Ideologies in the 1920s-1930s: Their Impact on the Segregation of Mexican Students in California.” *Anthropology & Education Quarterly* 21 (3): 222–49.

Minns, Chris. "Income, cohort effects, and occupational mobility: a new look at immigration to the United States at the turn of the 20th century." *Explorations in economic history* 37.4 (2000): 326-350.

Morales, Daniel. *The Making of Mexican America: Transnational Networks in the Rise of Mass Migration 1900-1940*. Diss. Columbia University, 2016.

Morefield, Richard Henry. "Mexicans in the California Mines, 1848-53." *California Historical Society Quarterly* 35.1 (1956): 37-46.

Muñoz Martinez, Monica. 2018. *The Injustice Never Leaves You*. Harvard University Press.

Ngai, Mae M. "The Strange Career of the Illegal Alien: Immigration Restriction and Deportation Policy in the United States, 1921–1965." *Law and History Review* 21, no. 1 (2003): 69-108.

O'Neill, Colleen. 1993. "Domesticity Deployed: Gender, Race and the Construction of Class Struggle in the Bisbee Deportation." *Labor History* 34 (2–3): 256–73.

Pérez, Santiago. "The (South) American dream: Mobility and economic outcomes of first-and second-generation immigrants in nineteenth-century Argentina." *The Journal of Economic History* 77, no. 4 (2017): 971-1006.

Perlmann, Joel. *Italians Then, Mexicans Now: Immigrant Origins and the Second-Generation Progress, 1890-2000*. Russell Sage Foundation, 2005.

Pierce, C. C. 1917. "Combating Typhus Fever on the Mexican Border." *Public Health Reports (1896-1970)* 32 (12): 426–29.

Reisler, Mark. 1976. "Always the Laborer, Never the Citizen: Anglo Perceptions of the Mexican Immigrant during the 1920s." *Pacific Historical Review* 45 (2): 231–54.

Rigby, David, and Charles Seguin. "The Racial Position of European Immigrants 1883–1941: Evidence from Lynching in the Midwest." *Social Currents* 5, no. 5 (2018): 438-457.

Ruggles Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek. IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D010.V10.0

Saavedra, Martin, and Tate Twinam. "A machine learning approach to improving occupational income scores." *Explorations in Economic History* 75 (2020): 101304.

Scruggs, Otey M. "The first Mexican farm labor program." *Arizona and the West* 2, no. 4 (1960): 319-326.

Smith, James P. "Immigrants and the labor market." *Journal of Labor Economics* 24, no. 2 (2006): 203-233.

Stern, Alexandra Minna. 1999. “Buildings, Boundaries, and Blood: Medicalization and Nation-Building on the U.S.-Mexico Border, 1910-1930.” *Hispanic American Historical Review* 79 (1): 41–81.

Taylor, Paul S. *Mexican Labor in the United States: Migration Statistics I.* University of California Press, 1929.

Tolnay, Stewart Emory, and Elwood M. Beck. *A festival of violence: An analysis of southern lynchings, 1882-1930*. University of Illinois Press, 1995.

Verduzco, Gustavo. "La migración mexicana a Estados Unidos: recuento de un proceso histórico." *Estudios Sociológicos* 13, no. 39 (1995): 573-594.

Ward, Zachary. "The low return to English fluency during the Age of Mass Migration." *European Review of Economic History* 24.2 (2020): 219-242.

Ward, Zachary. “Internal Migration, Education and Intergenerational Mobility: Evidence from American History.” *The Journal of Human Resources*, forthcoming.

Ward, Zachary. *Intergenerational Mobility in American History: Accounting for Race and Measurement Error*. No. 10. Centre for Economic History, Research School of Economics, Australian National University, 2019.

Wildsmith, Elizabeth, Myron P. Gutmann, and Brian Gratton. "Assimilation and intermarriage for US immigrant groups, 1880–1990." *The History of the Family* 8, no. 4 (2003): 563-584.

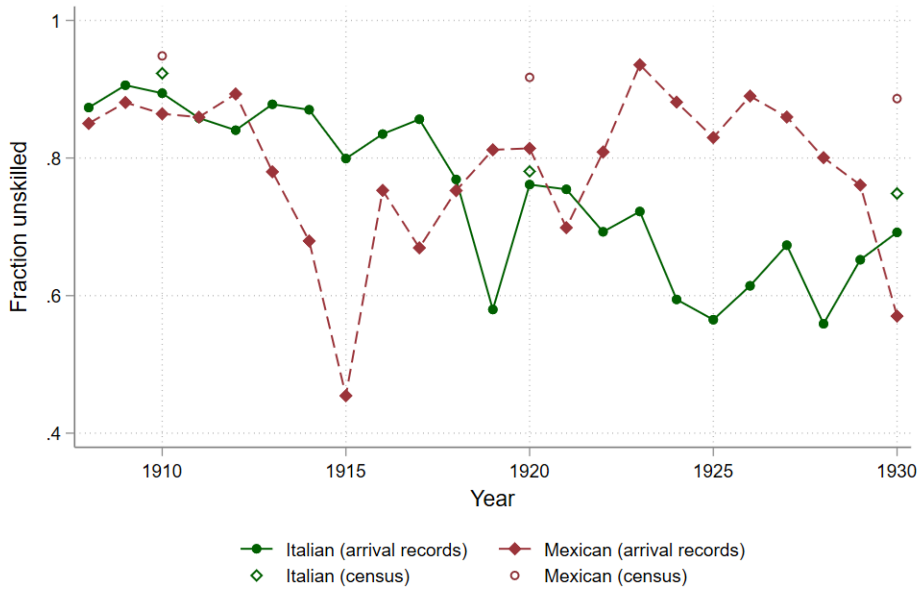
Woodruff, Christopher, and Rene Zenteno. "Migration networks and microenterprises in Mexico." *Journal of Development Economics* 82, no. 2 (2007): 509-528.

Young, Julia G. “Cristero Diaspora: Mexican Immigrants, the U.S. Catholic Church, and Mexico’s Cristero War, 1926-29.” The Catholic Historical Review 98.2 (2012): 271-300.

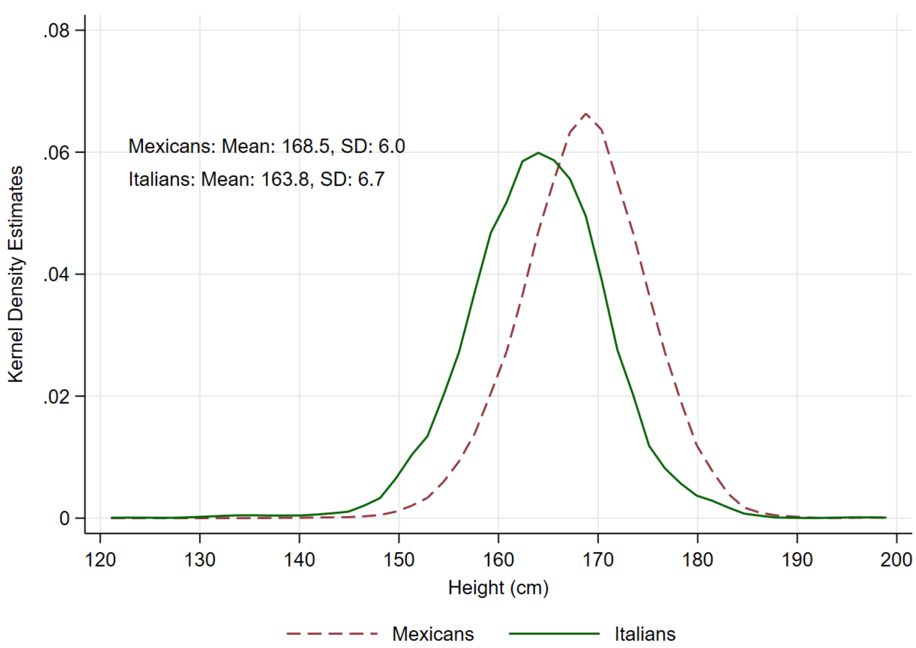
Young, Julia G. *Mexican exodus: Emigrants, exiles, and refugees of the cristero war*. Oxford University Press, 2015.

Figure 1. Mexican and Italian outcomes observed in arrival records

Panel A. Fraction unskilled in arrival records and the census

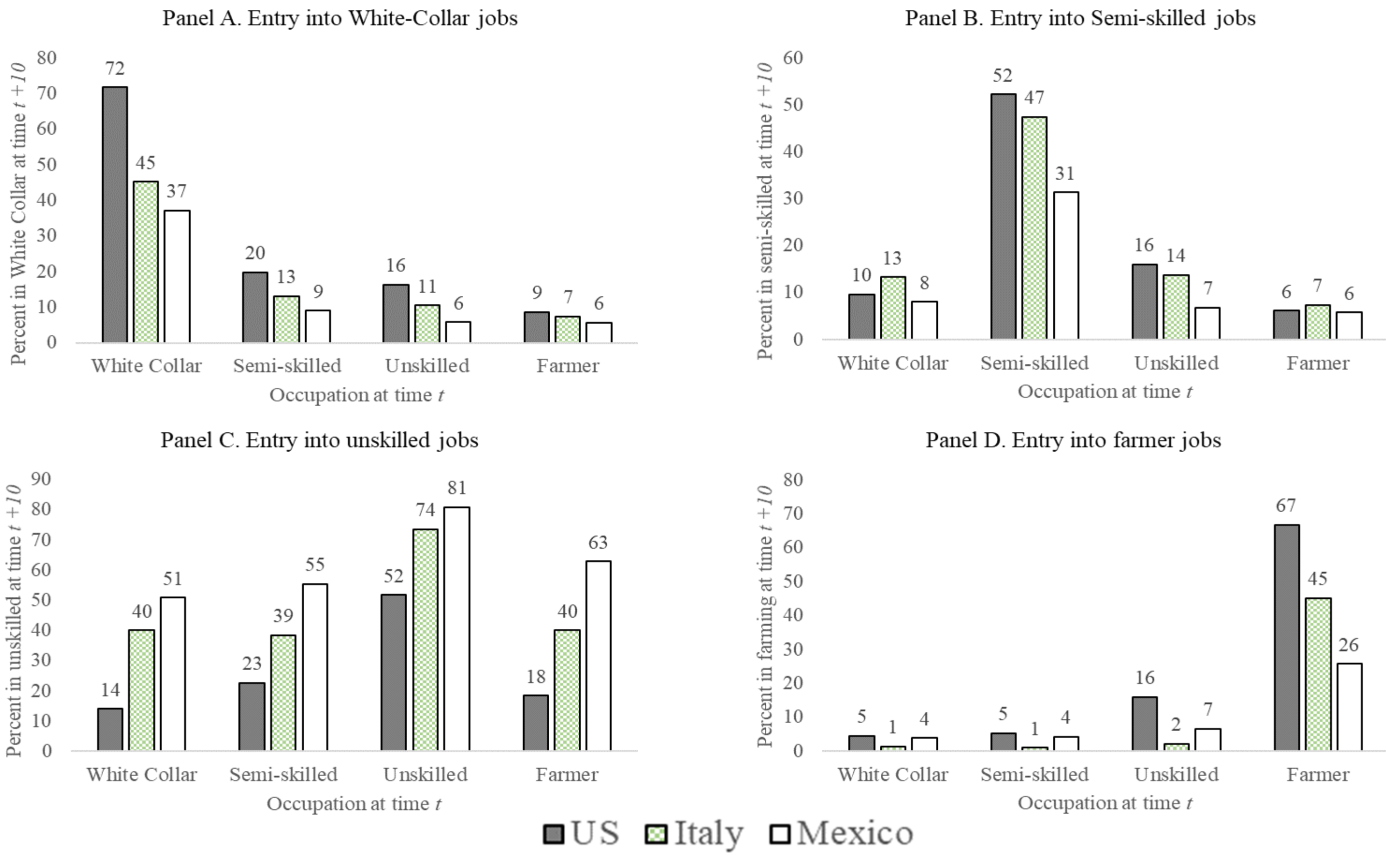


Panel B. Height distributions from arrival records



Notes: Data in Panel A are from the 1908-1930 *Annual Report of the Commissioner General of Immigration* and the 1910, 1920 and 1930 full-count census (Ruggles et al. 2020). Data in Panel B are from Escamilla-Guerrero (2020, Kosack and Ward (2014) and Spitzer and Zimran (2018). Heights are for those 22-65-year-old males who arrived in 1907, 1908 and 1920.

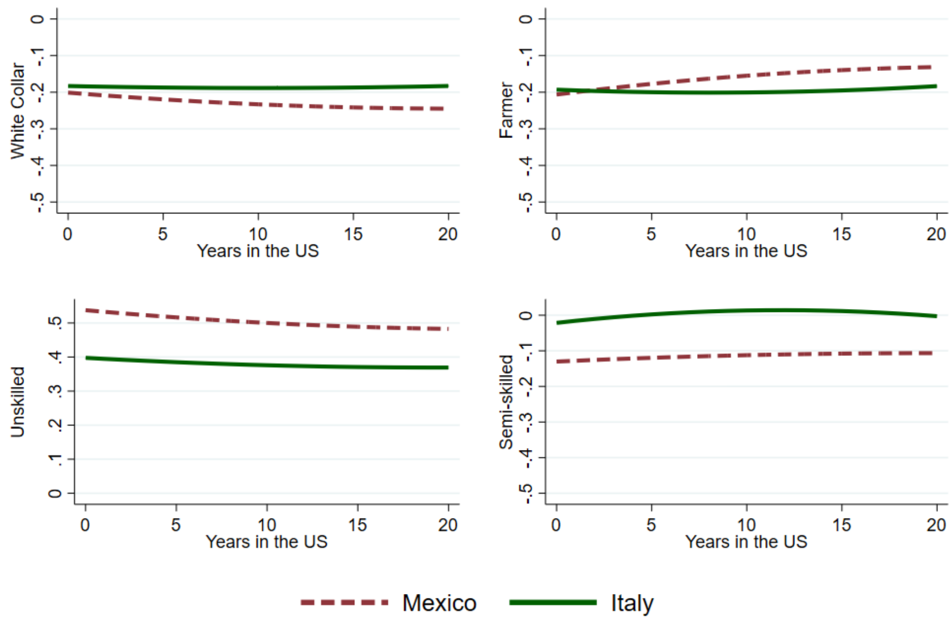
Figure 2. Conditional on first occupation, entry into occupational category, US white v. Mexico v. Italy



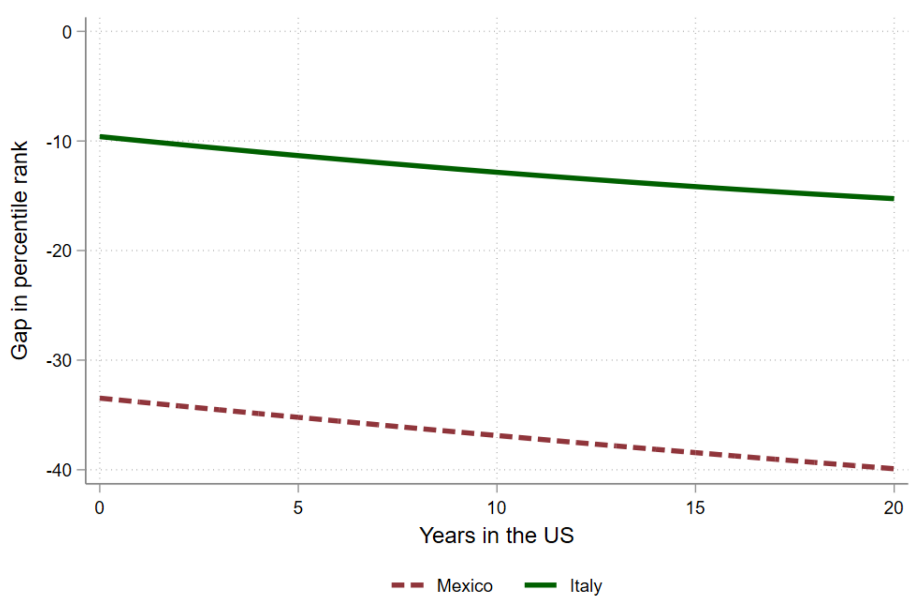
Notes: Data are from the 1910-1920, 1920-1930, 1930-1940 linked data. Each figure shows the average number of individuals in a job category at second observation, conditional on category at first observation. Sum a given occupational category and country of birth across panels to 100. For example, for Mexico-born white-collar workers at first observation, 37 percent ended in white-collar jobs (top left), 8 percent in semi-skilled jobs (top right), 51 percent in unskilled jobs (bottom left), and 4 percent in farmer jobs (bottom right

Figure 3. Mexican and Italian assimilation profiles for occupational categories and percentile ranks

Panel A. Occupational Categories

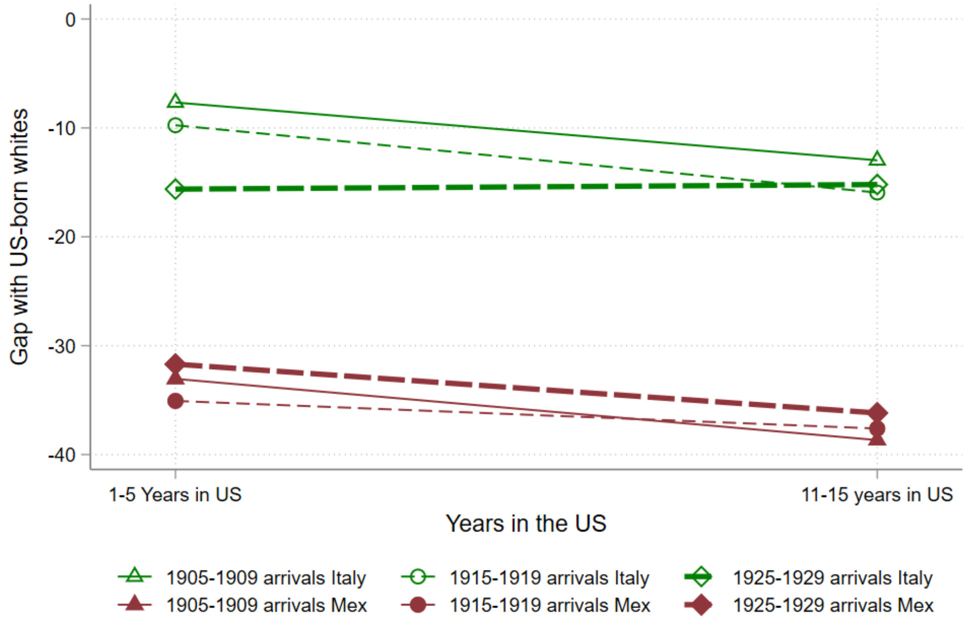


Panel B. Percentile ranks



Notes: Data are from 1910-1920, 1920-1930, and 1930-1940 linked samples. Panel A plots the predicted convergence in occupational categories for the Mexico-born and Italy-born 1910-1919 arrival cohort. Panel B show the assimilation profile after assigning earnings score to each individual based on their occupation, country of birth and census division, and then percentile ranking within the sample by year and birth cohort.

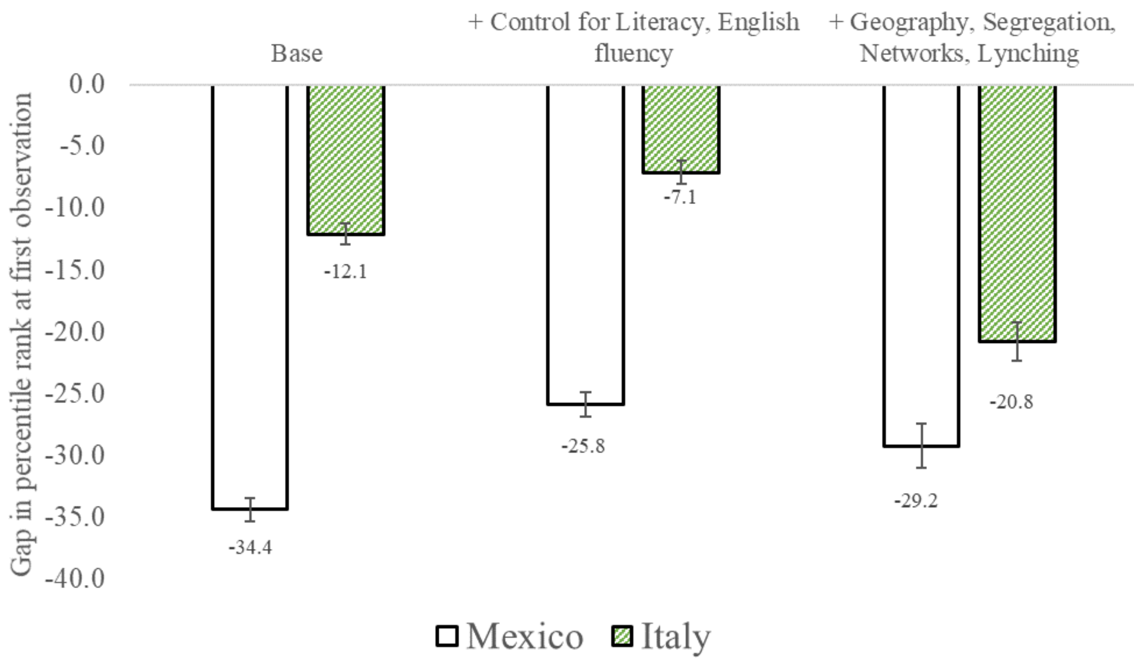
Figure 4. Raw gaps in percentile ranks with US-born for Mexican and Italian immigrants



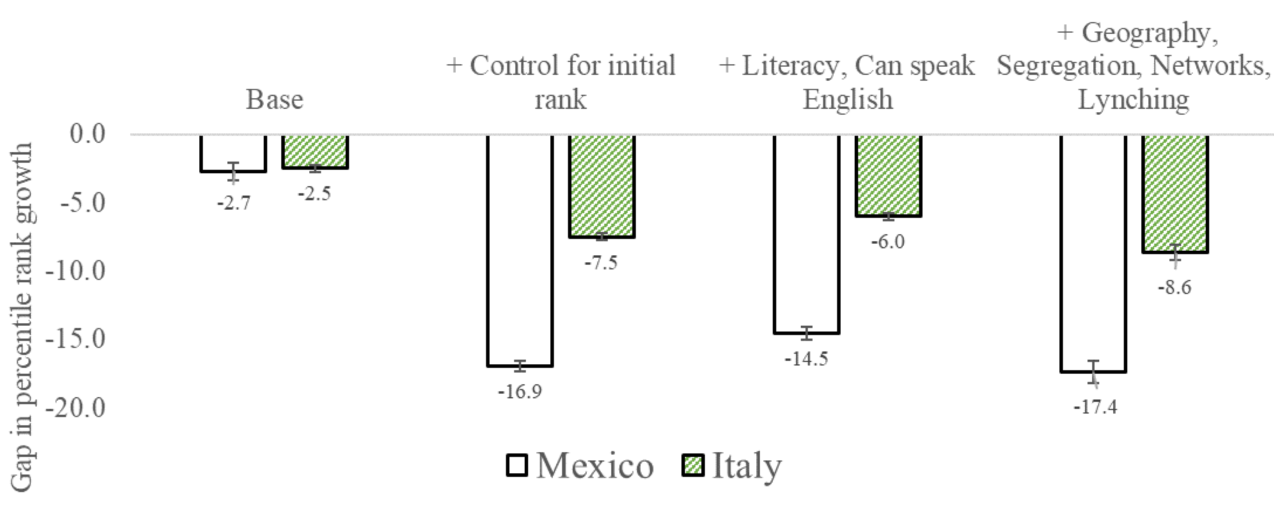
Notes: Data are from the 1910-1920, 1920-1930, and 1930-1940 linked samples. The figure plots the raw gap in percentile rank with US-born whites. The figure shows that Mexicans had wider gaps with the US-born than Italians, and the gaps did not converge after arrival, and there was not much variation in arrival gaps or growth rates after ten more years of stay.

Figure 5. Gaps in initial position and in decadal growth for Mexican and Italian immigrants

Panel A. Gaps in initial percentile rank



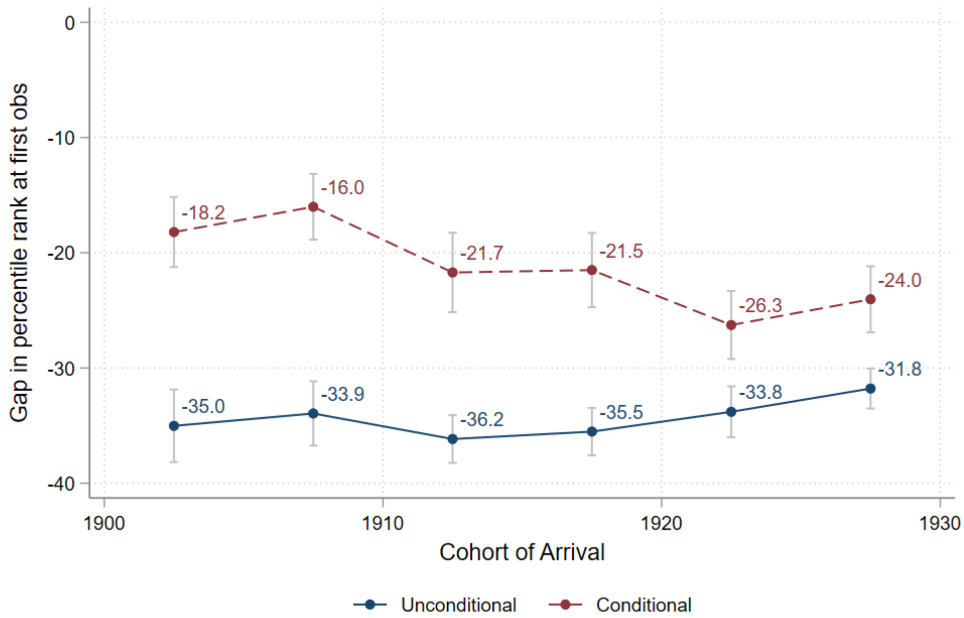
Panel B. Gaps in percentile rank growth from census to census



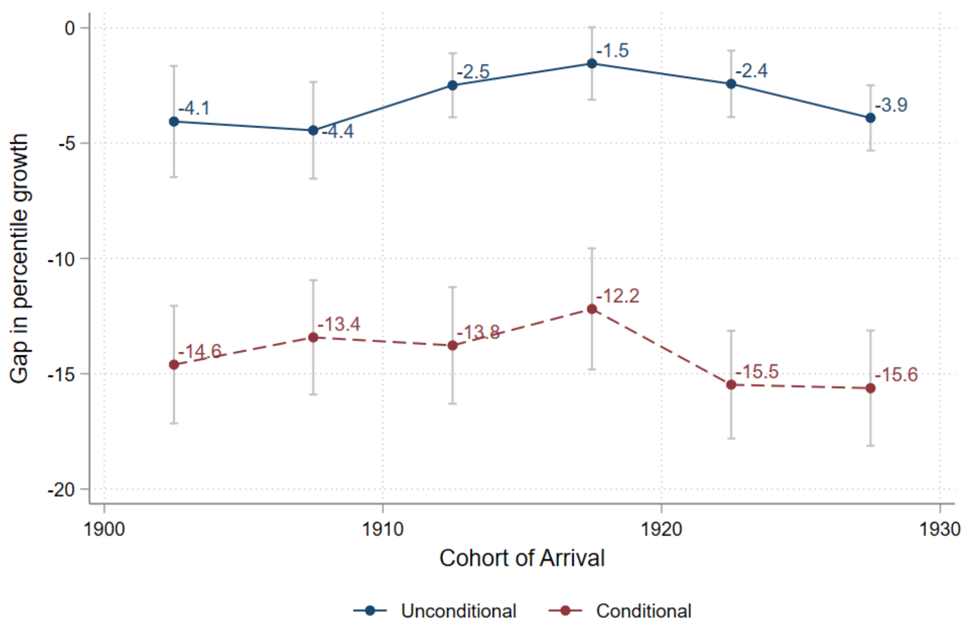
Notes: Data are from the 1910-1920, 1920-1930, and 1930-1940 linked samples. Panel A plots the difference in gaps at first observation with US-born whites. Panel B plots the gap in decadal change in percentile rank. Human capital controls are for literacy and ability to speak English at first observation. Geography, segregation and networks control for initial county fixed effects, segregation level from the US-born at the county level, urban, log county population, log number of immigrants in county from the country of birth, and whether a lynching occurred in the county.

Figure 6. Differences in outcomes across cohort of arrival for Mexican immigrants

Panel A. Gaps at arrival



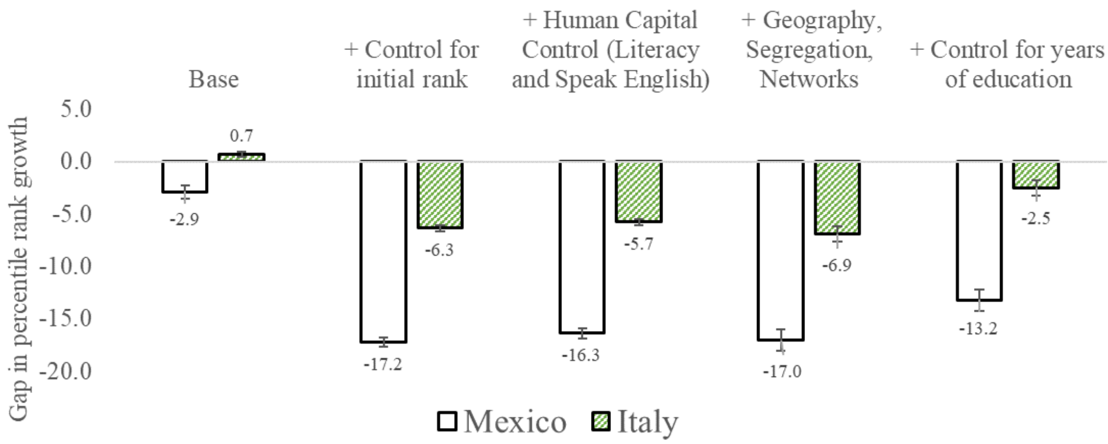
Panel B. Gaps in decadal growth



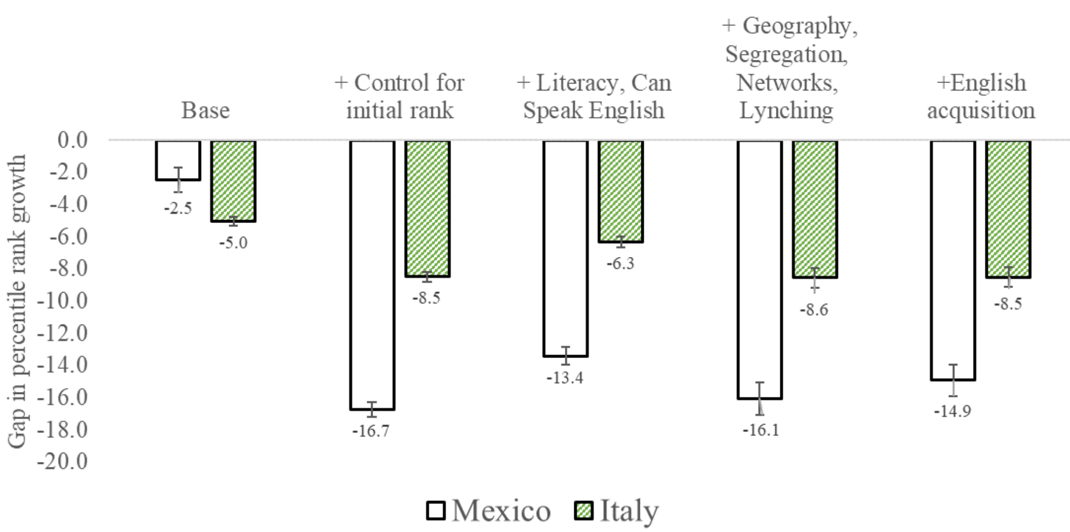
Notes: Data are from the 1910-1920, 1920-1930, and 1930-1940 linked samples. Panel A plots the difference in gaps at first observation by Mexican arrival cohort. Panel B plots the gap in decadal change in percentile rank. Unconditional differences plot the raw difference in percentile rank level or growth, while conditional differences control for the human capital and geography controls.

Figure 7. Gaps in initial position and in decadal growth for Mexican and Italian immigrants including additional control variables for human capital

Panel A. Control for years of education with the 1930-1940 panel



Panel B. Control for English acquisition with the 1910-1920 and 1920-1930 panels



Notes: For Panel A: Data are from the 1930-1940 linked sample. Education can only be observed in the 1940 census. The first four columns are the same estimates as in the main sample. The last column includes a control for dummy variables for years of education, to flexibly control for the return to education. For Panel B: Data are from the 1910-1920 and 1920-1930 linked samples. The change in English Acquistion can only be controlled for in the 1910-1920 and 1920-1930 panels. The first four columns are the same estimates as in the main sample. In the last column, we additionally control for the change in English fluency.

Table 1. Descriptive statistics for US-born whites, Mexico-born and Italy-born immigrants, 1910-1940

	US-born white			Mexico-born			Italy-born		
	First Census	Second Census	Diff over 10 years	First Census	Second Census	Diff over 10 years	First Census	Second Census	Diff over 10 years
Percentile rank	51.29	50.47	-0.82	17.20	13.53	-3.67	39.26	36.00	-3.26
	(28.97)	(28.94)	(0.01)	(15.09)	(14.88)	(0.15)	(14.99)	(15.82)	(0.11)
Log earnings score	6.97	7.06	0.09	6.43	6.42	-0.01	6.86	6.89	0.03
	(0.57)	(0.48)	(0.00)	(0.38)	(0.40)	(0.00)	(0.20)	(0.21)	(0.00)
Literate	0.98	0.98	-0.00	0.62	0.65	0.03	0.74	0.74	-0.00
	(0.15)	(0.15)	(0.00)	(0.49)	(0.48)	(0.01)	(0.44)	(0.44)	(0.00)
Can speak English	0.98	0.99	0.00	0.28	0.49	0.21	0.67	0.89	0.22
	(0.12)	(0.11)	(0.00)	(0.45)	(0.50)	(0.01)	(0.47)	(0.32)	(0.00)
Age	30.71	40.71	10.01	30.16	40.29	10.14	30.84	40.96	10.12
	(5.40)	(5.42)	(0.00)	(5.31)	(5.39)	(0.05)	(5.24)	(5.26)	(0.04)
Age at arrival	0.00	0.00	0.00	24.61	24.75	0.14	23.31	23.42	0.12
	(0.00)	(0.00)	(0.00)	(5.79)	(5.86)	(0.06)	(5.32)	(5.33)	(0.04)
Border state	0.09	0.10	0.01	0.78	0.82	0.04	0.07	0.07	0.00
	(0.28)	(0.30)	(0.00)	(0.42)	(0.38)	(0.00)	(0.26)	(0.26)	(0.00)
White Collar	0.27	0.30	0.03	0.06	0.08	0.02	0.11	0.15	0.04
	(0.45)	(0.46)	(0.00)	(0.24)	(0.27)	(0.00)	(0.31)	(0.35)	(0.00)
Farmer	0.22	0.22	0.00	0.04	0.07	0.03	0.02	0.03	0.01
	(0.41)	(0.41)	(0.00)	(0.20)	(0.26)	(0.00)	(0.13)	(0.16)	(0.00)
Skilled	0.18	0.19	0.00	0.07	0.09	0.01	0.21	0.21	-0.01
	(0.39)	(0.39)	(0.00)	(0.26)	(0.28)	(0.00)	(0.41)	(0.41)	(0.00)
Unskilled	0.33	0.29	-0.04	0.83	0.76	-0.06	0.66	0.62	-0.04
	(0.47)	(0.45)	(0.00)	(0.38)	(0.42)	(0.00)	(0.47)	(0.49)	(0.00)
Lives in Urban area	0.50	0.51	0.01	0.52	0.57	0.05	0.84	0.85	0.01
	(0.50)	(0.50)	(0.00)	(0.50)	(0.50)	(0.01)	(0.37)	(0.36)	(0.00)
Observations	11,013,388	11,013,388		23,411	23,411		51,947	51,947	

Notes: Data are from the 1910-1920, 1920-1930, 1930-1940 linked samples. Age and age at arrival may mismatch across ten years because linking algorithm allows for a 2-year difference in age and 7-year difference in years in the United States. Literacy is the ability to read and write in any language, which is not included in the 1930-1940 linked sample since the 1940 Census does not include it. Speak English is also not in the 1930-1940 sample for the same reason. Standard deviations are reported for the means, while standard errors are reported for the differences over 10 years.

Table 2. Main occupations for US-born whites, Mexico-born and Italy-born immigrants, 1910-1940

Mexico			United States			Italy		
Occupation	Percent	Percentile Rank	Occupation	Percent	Percentile Rank	Occupation	Percent	Percentile Rank
Laborers	39.1	13.5	Farmer, owner	12.5	36.7	Laborers	28.6	27.0
Farm laborers	23.9	6.1	Farmer, tenant	9.3	20.4	Operative workers	10.8	40.9
Miners	5.0	21.4	Managers	8.7	91.8	Managers	7.1	60.9
Farmer, tenant	4.9	6.6	Laborers	7.4	18.6	Mine operatives	5.9	34.5
Operative workers	3.4	31.8	Farm laborers	5.9	8.3	Barbers	3.4	40.3
Managers	2.6	42.0	Salesmen	5.6	73.3	Tailors and tailoresses	3.1	41.2
Salesmen	1.4	29.3	Operative workers	5.1	49.5	Brick masons	2.9	38.8
Carpenters	1.3	11.2	Clerical workers	3.0	70.4	Farm laborers	2.4	22.1
Truck and tractor drivers	1.0	17.9	Carpenters	2.6	41.4	Carpenters	2.2	32.5
Farmer, owner	0.7	20.5	Machinists	2.3	65.2	Shoemakers	2.1	36.2
Service workers	0.7	9.9	Mine operatives	1.9	35.7	Salesmen	1.8	54.2
Barbers	0.6	18.8	Foremen	1.8	79.6	Machinists	1.3	59.5
Cooks	0.6	16.4	Truck drivers	1.7	41.0	Farmer, owner	1.3	61.5
Tailors and tailoresses	0.6	18.5	Mechanics	1.3	44.7	Longshoremen	1.2	45.9
Painters	0.6	10.3	Painters	1.1	35.8	Cooks	1.2	36.6
Bakers	0.5	30.0	Electricians	1.0	70.1	Painters	1.2	25.0
Machinists	0.5	44.5	Bookkeepers	0.9	72.3	Truck and tractor drivers	1.1	35.9
Gardeners	0.5	23.3	Stationary engineers	0.9	76.4	Waiters and waitresses	1.0	30.7
Clerical workers	0.5	48.4	Plumbers	0.9	58.9	Bakers	1.0	52.8
Mechanics	0.5	18.5	Teachers	0.8	77.2	Foremen	0.9	75.8

Notes: Data are from the 1910-1920, 1920-1930, 1930-1940 linked samples. The above are the most common occupations by country of birth in our sample. Occupations are determined by the *occ1950* code from IPUMS. Farmers are separated into owners and tenants based on the home ownership variable. Percentile ranks are based on imputed earnings form one's occupation, census division and country of birth. Percentile ranks are calculated within the sample b y year and birth cohort.

Table 3. Across and within-cohort percentile rank differences for Mexican and Italian immigrants

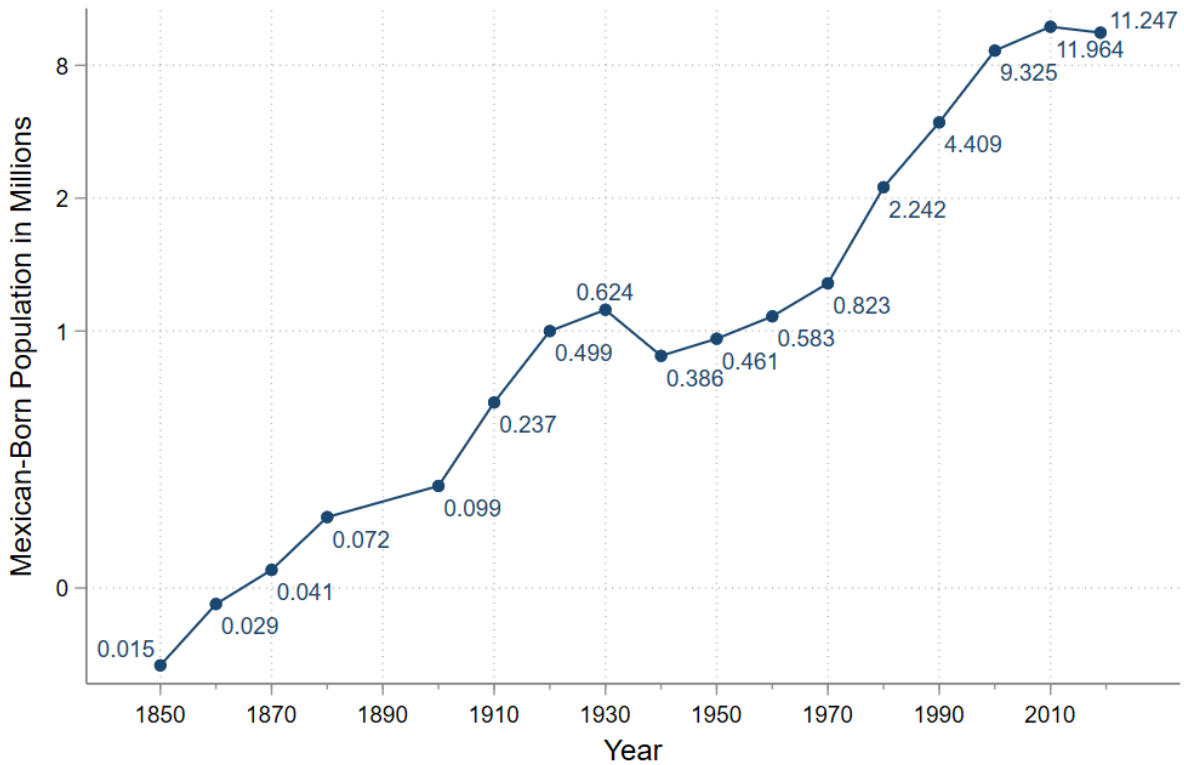
	Census Year				
	1910	1920	1930	1940	
Panel A. Mexico-born					
1900-1904	-34.057	-39.170			-5.113
	(0.500)	(0.490)			(0.700)
1905-1909	-33.108	-38.622			-5.514
	(0.316)	(0.288)			(0.427)
1910-1914		-35.540	-39.168		-3.628
		(0.224)	(0.193)		(0.295)
1915-1919		-34.859	-37.651		-2.792
		(0.180)	(0.172)		(0.248)
1920-1924			-33.488	-36.830	-3.342
			(0.221)	(0.228)	(0.317)
1925-1929			-31.437	-36.277	-4.840
			(0.284)	(0.293)	(0.408)
Panel B. Italy-Born					
1900-1904	-5.956	-11.700			-5.744
	(0.177)	(0.178)			(0.250)
1905-1909	-7.770	-12.938			-5.168
	(0.168)	(0.174)			(0.241)
1910-1914		-10.223	-16.652		-6.429
		(0.106)	(0.109)		(0.152)
1915-1919		-9.578	-16.037		-6.459
		(0.358)	(0.386)		(0.527)
1920-1924			-15.593	-15.960	-0.367
			(0.142)	(0.164)	(0.216)
1925-1929			-15.409	-15.041	0.368
			(0.501)	(0.561)	(0.752)

Notes: Data are from linked censuses (1910-1920; 1920-1930; 1930-1940). The tables report the gap in percentile gaps with US-born whites for each year and arrival cohort. The last column calculates the ten-year within-cohort growth rate in percentile rank. Percentile ranks are based on earnings scores and are calculated within the sample by year and birth cohort.

Online Appendix

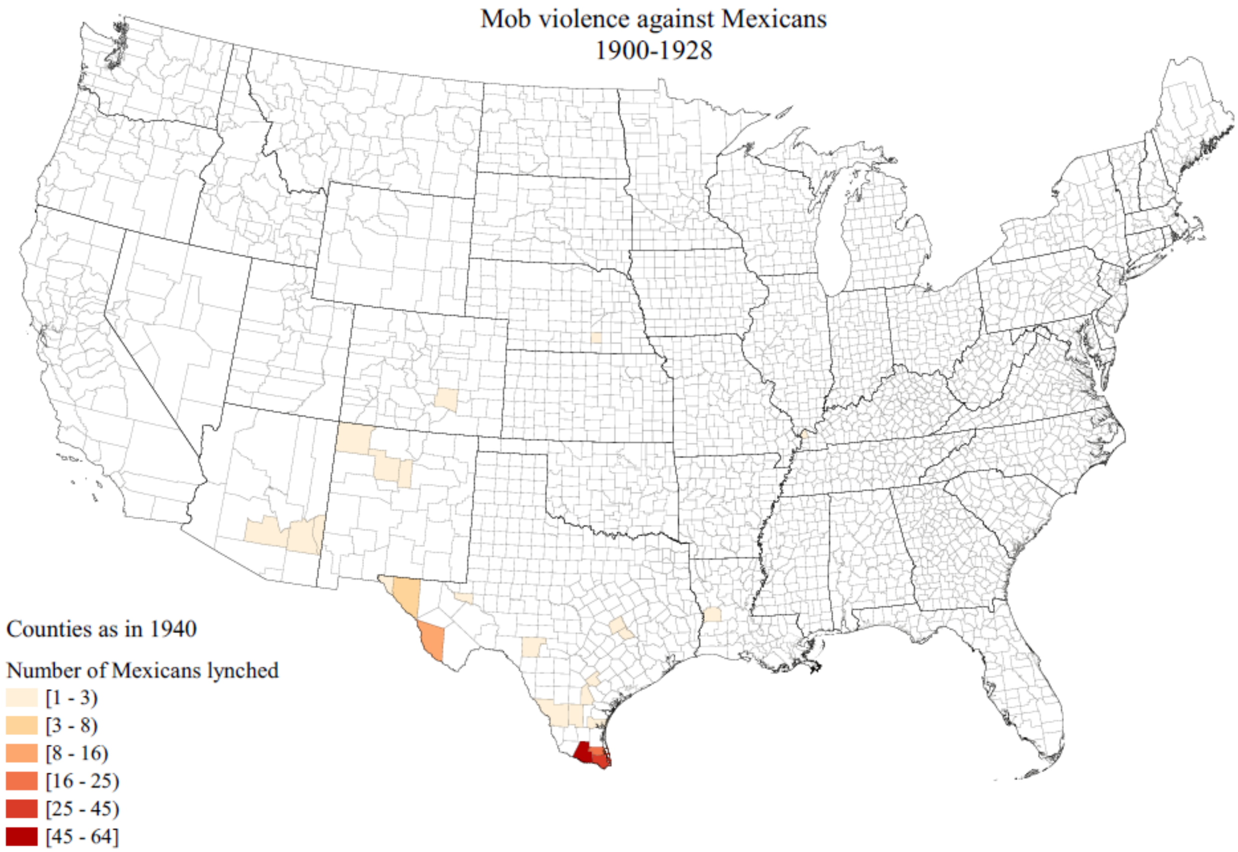
Life after Crossing the Border:
Assimilation during the first Mexican Mass Migration

Figure A1. Stock of Mexican-born population in United States, 1850 to 2019



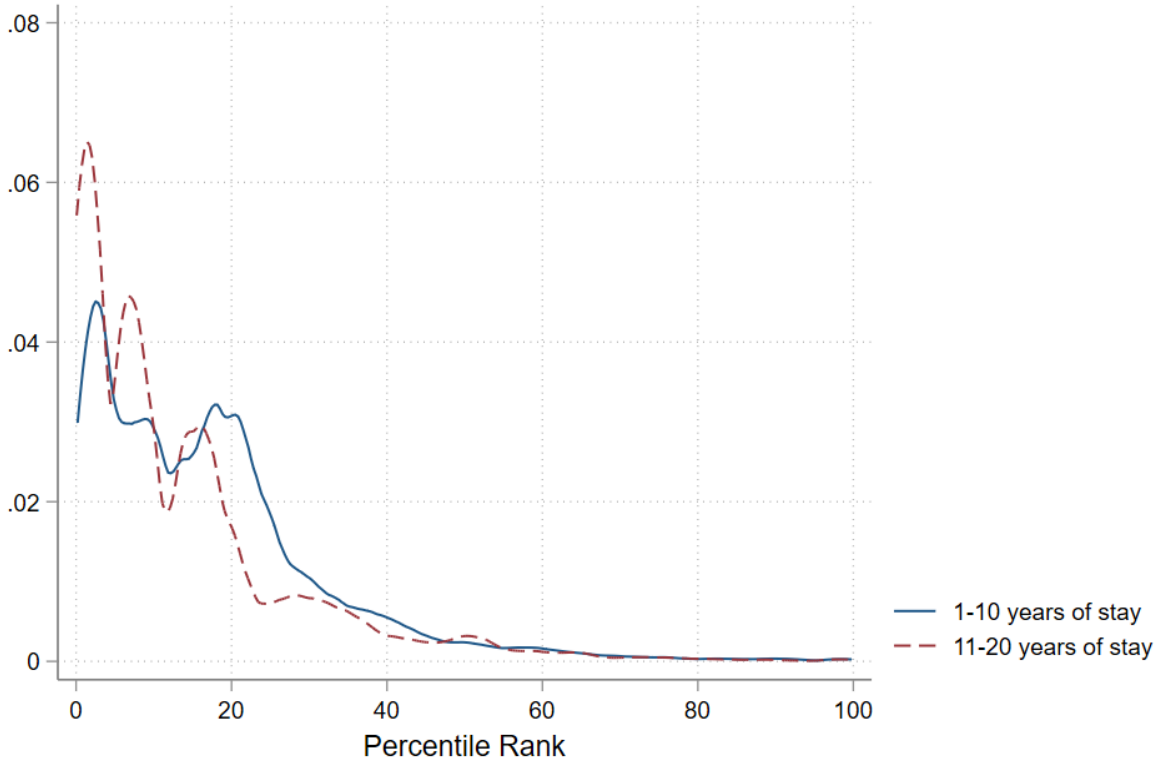
Notes: Data is from the 1850 to 2010 US Census samples and the 2019 ACS from IPUMS (Ruggles et al. 2020). In this study we examine cohorts who arrived between 1900-1929 and are observed in the 1910 to 1940 Censuses.

Figure A2. Map of Mexican lynchings between 1900 and 1928.



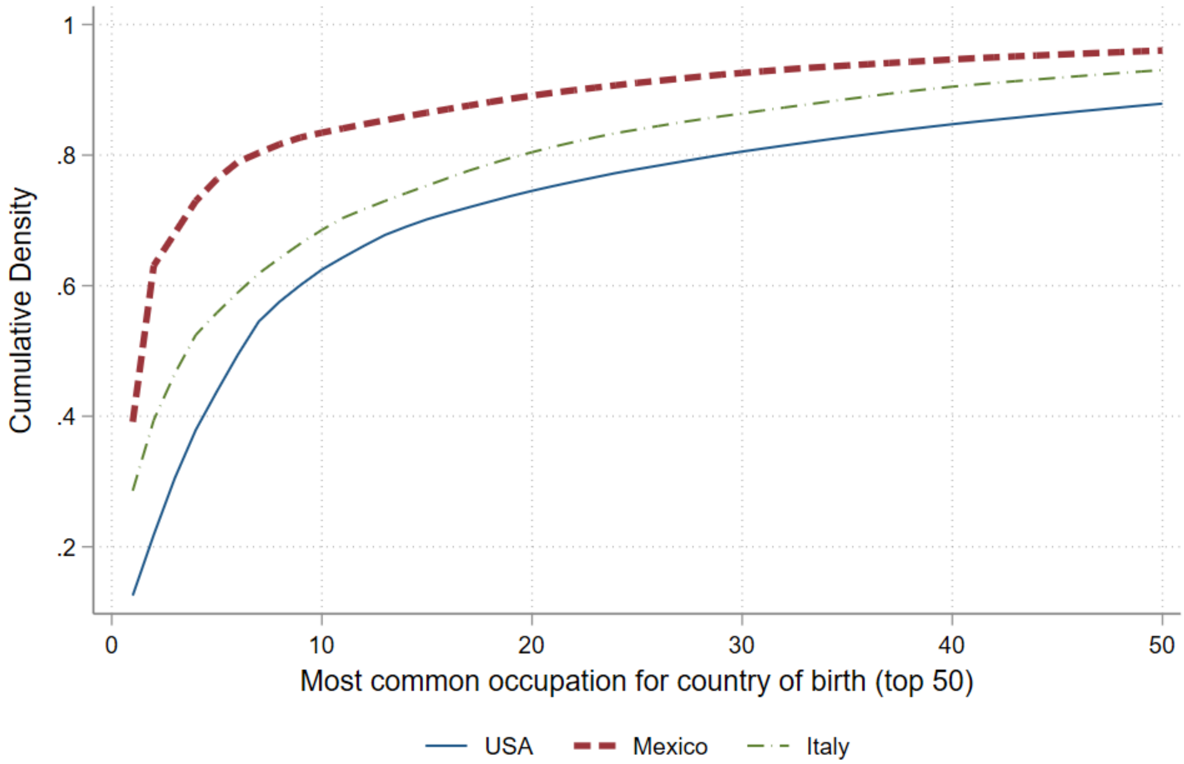
Notes: Data are from Carrigan and Webb (2013).

Figure A3. Mexican-born location in the earnings score distribution in the early 20th century



Notes: Data are from the linked censuses between 1910-1920, 1920-1930 and 1930-1940. Percentile ranks are calculated within the data by year and birth cohort. Ties are given equal percentile rank.

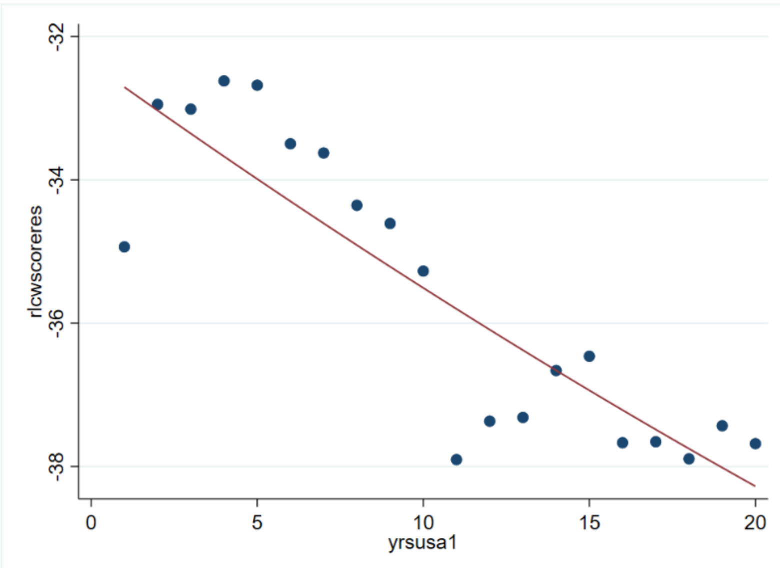
Figure A4. Cumulative density of the top 50 occupations for US-born whites, Mexico-born and Italy-born immigrants, 1910-1940



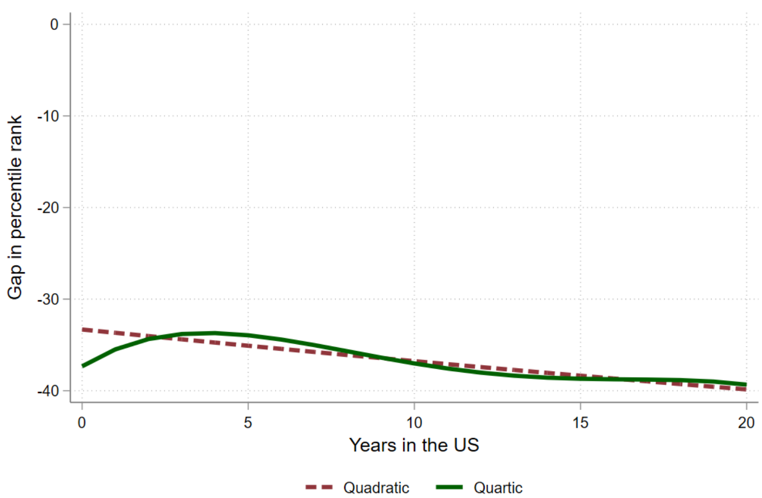
Notes: Data are from the linked censuses between 1910-1920, 1920-1930 and 1930-1940. The figure shows the cumulative fraction of the sample in the top 50 occupations (x-axis).

Figure A5. Functional form for the assimilation profile of Mexican immigrants

Panel A. Binscatter relationship between the gap with natives and years in the United States

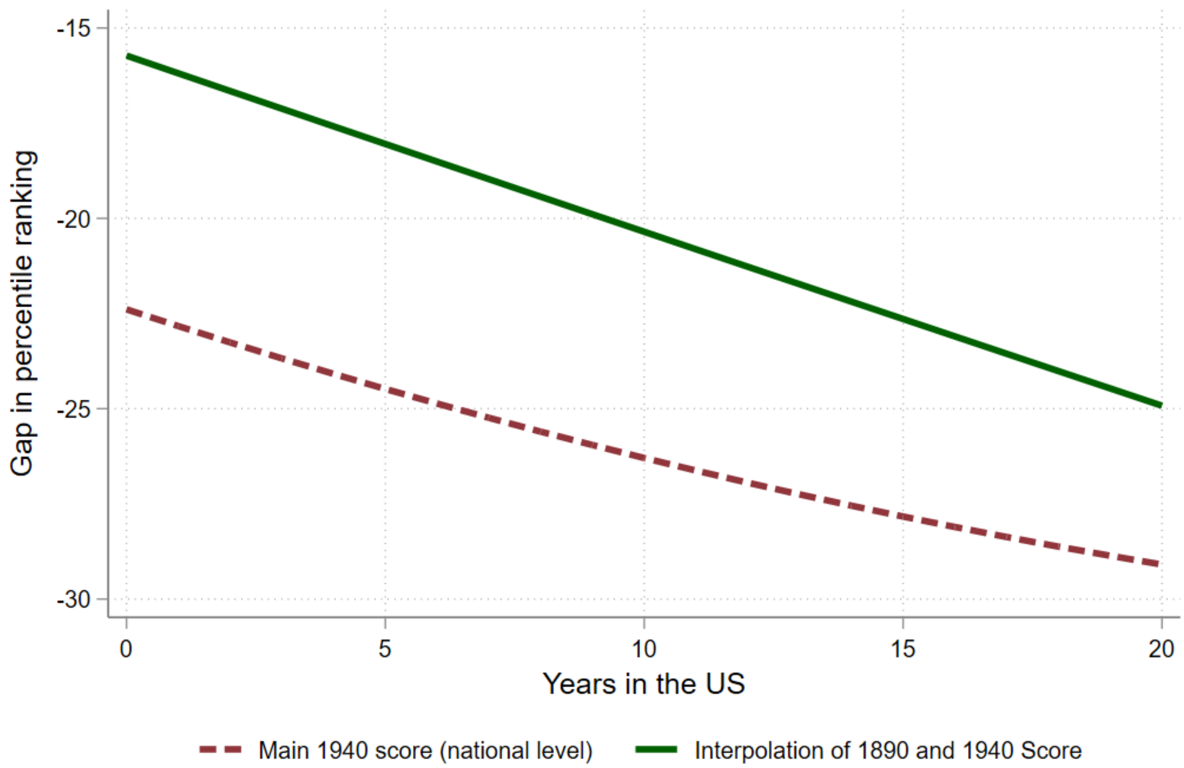


Panel B. Quadratic versus quartic assimilation profile.



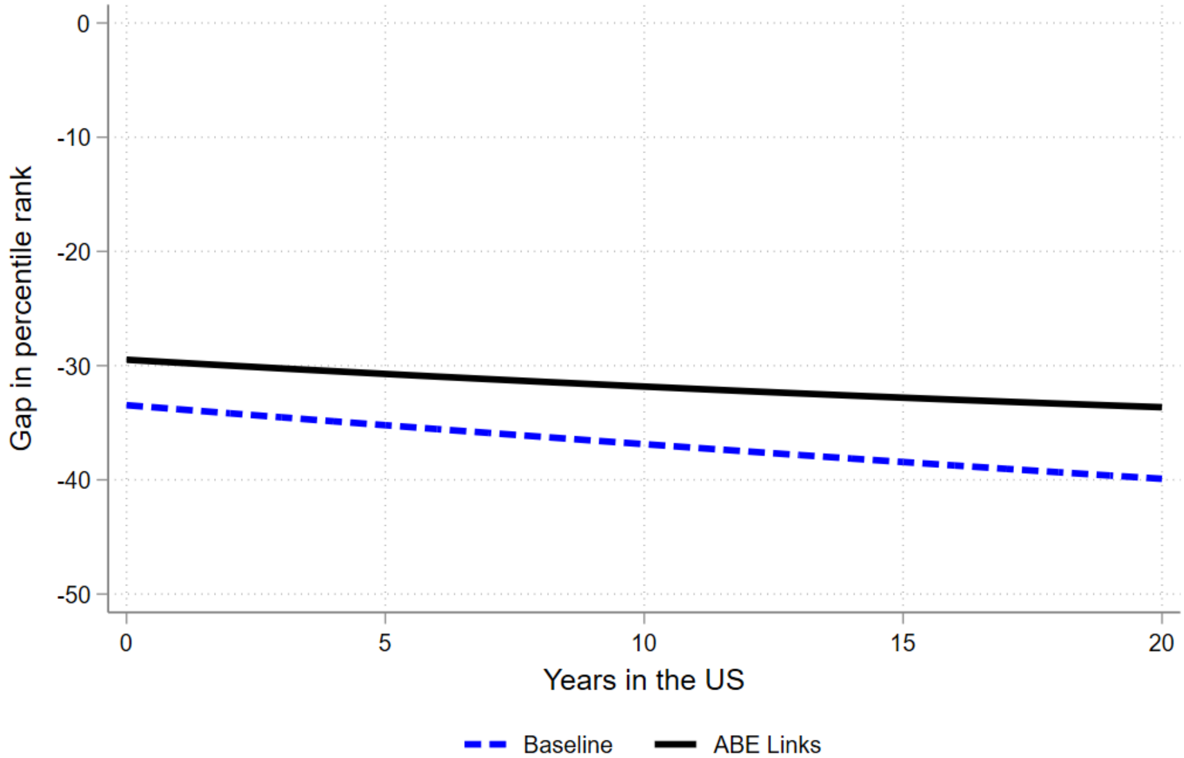
Notes: Data are from the 1910-1920, 1920-1930, and 1930-1940 linked samples. Panel A plots the binscatter relationship between the residual gap with US-born whites and years in the United States. Panel B estimates the assimilation profile where a quartic is used for years in the United States, rather than a quadratic. Both profiles suggest that there was some catch up between year 1 and year 2 of stay before falling behind.

Figure A6. Mexican assimilation profile using linear interpolation of 1890 and 1940 scores



Notes: Data are from the 1910-1920, 1920-1930, and 1930-1940 linked samples. The interpolation of the 1890 and 1940 Score is based on Sobek (1996) for 1890 and our score for 1940 (without accounting for country of birth or region differences).

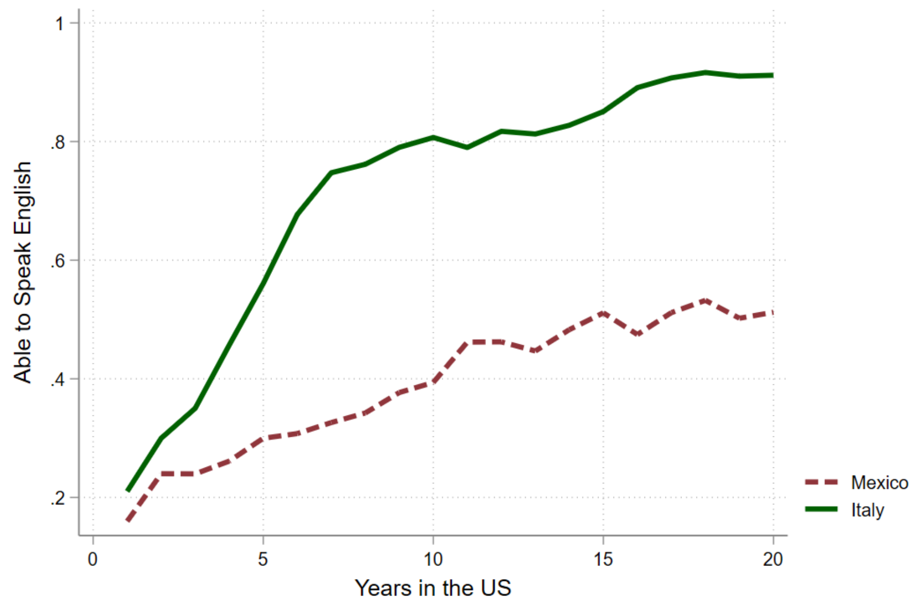
Figure A7. Mexican assimilation profile using data from the Census Linking Project.



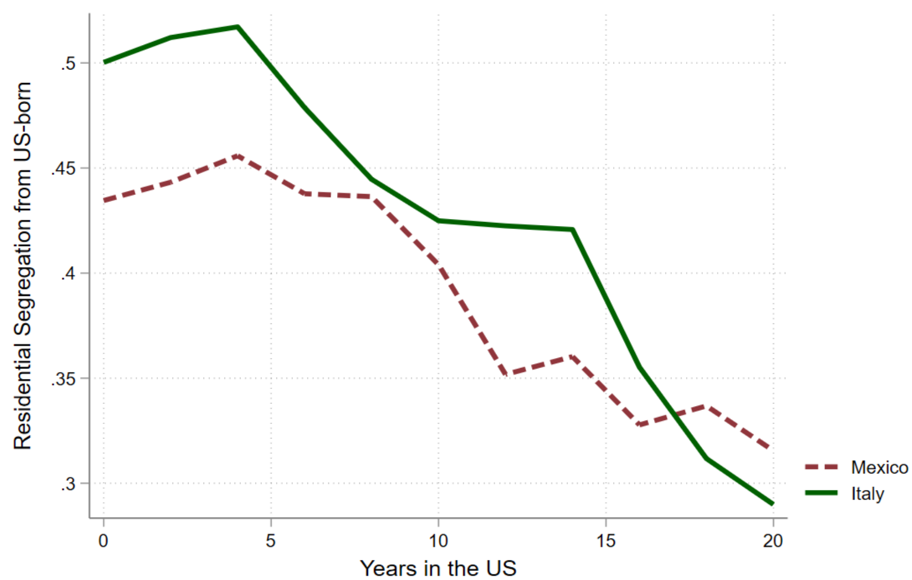
Notes: The baseline data is the linked data we create in the paper for 1910-1920, 1920-1930, and 1930-1940. The ABE links are from the Census Linking Project (Abramitzky et al. 2020)

Figure A8. English acquisition and spatial assimilation for Mexican and Italian immigrants

Panel A. Able to speak English



Panel B. Spatial Assimilation



Notes: Panel A's data are from linked samples between 1910-1920 and 1920-1930. It plots the raw means in the ability to speak English for the first twenty years after arrival. Panel B's data are from linked samples between 1910-1920, 1920-1930 and 1930-1940. It plots the estimate from a regression of the foreignness index of the second-generation child's name on mother's years in the United States. Ethnicities are based on mother tongue, where Mexican is Mexico-born individuals whose mother tongue is Spanish.

Table A1. Share of Mexican and Italian immigrants by occupation

	Mexico	Italy

	Year	Arrival Records	Census	Upgrade/ Downgrade	Arrival Records	Census	Upgrade/ Downgrade
White-Collar	1910	0.050	0.015	-0.035	0.013	0.018	0.005
	1920	0.080	0.026	-0.054	0.046	0.068	0.022
	1930	0.256	0.062	-0.195	0.081	0.083	0.002
Unskilled	1910	0.864	0.949	0.084	0.894	0.923	0.029
	1920	0.814	0.917	0.103	0.761	0.781	0.019
	1930	0.570	0.886	0.316	0.692	0.748	0.057
Skilled	1910	0.077	0.018	-0.059	0.084	0.056	-0.028
	1920	0.082	0.033	-0.049	0.180	0.140	-0.039
	1930	0.145	0.038	-0.107	0.183	0.164	-0.019
Farmer	1910	0.009	0.018	0.009	0.008	0.002	-0.006
	1920	0.024	0.024	-0.001	0.013	0.011	-0.002
	1930	0.028	0.014	-0.014	0.044	0.004	-0.039

Notes: Data are from the 1910, 1920, and 1930 Annual Report of the Commissioner General of Immigration and the 1910, 1920 and 1930 full-count censuses (Ruggles et al., 2020). The Annual Reports list the total number of immigrants by occupation and race/ethnicity. The Annual Report covers fiscal years (e.g., July 1909-June 1910 arrivals in the 1910 Annual Report). The census includes arrivals with less than or equal to one year of stay (e.g., 1909-1910 in the 1910 Census). Therefore, there is slippage in the comparison. Mexican is included for "Mexico" and "South Italian" and "North Italian" are combined for Italy.

Table A2. Inferred selection into return migration for Mexican immigrants.

Cohort of Arrival	Within-cohort growth rate		Difference	Implied selection of return migrants
	Panel	RCS		
1900-1904	-5.11	-3.18	-1.93	-4.39
1905-1909	-5.51	-3.87	-1.64	-3.73
1910-1914	-3.63	-2.86	-0.77	-1.75
1915-1919	-2.79	-1.74	-1.05	-2.39
1920-1924	-3.34			
1925-1929	-4.84			

Notes: Panel data are from linked samples for 1910-1920 and 1920-1930. Repeated cross-sectional data are from the 1910-1930 full-count censuses (Ruggles et al. 2020). The 1940 census cannot be included in the RCS data because it does not include year of arrival. The implied selection of return migrants is calculated as the difference in within-cohort growth rate across the panel and RCS divided by the return rate (see Abramitzky et al. 2014). The decadal return rate is assumed to be 44 percent, which is based on the return rate when linking arrival records to the US and Mexican censuses (Kosack and Ward 2014).

Table A3. Differences in initial gaps between Mexicans, Italians and US-born whites

	I	II	III
Mexico	-34.351	-25.845	-29.214
	(0.931)	(0.983)	(1.823)
Italy	-12.079	-7.095	-20.757
	(0.870)	(0.918)	(1.555)
Literate		24.764	11.261
		(0.482)	(0.269)
Can speak English		2.894	0.670
		(0.525)	(0.181)
Urban			18.807
			(0.369)
Segregation level			-8.250
			(3.065)
Log source country population			-0.265
			(0.065)
Log county population			2.614
			(0.429)
Lynching in prior 10 years			-0.751
			(0.806)
Lynching in prior 10 years x Mexico			0.301
			(2.458)
Year Fixed Effects	Y	Y	Y
County fixed effects	N	N	Y
F-test p-value (Mex=Ita)	0	0	0.005
Observations	11,088,451	11,088,451	11,088,451

Notes: Data are from linked censuses (1910-1920; 1920-1930; 1930-1940). The dependent variable is the log earnings score at first observation (1910, 1920 or 1930). Urban is defined as living in a city or incorporated area of more than 2,500 people. Border state indicates living in California, Arizona, New Mexico or Texas. The segregation levels are taken from Eriksson and Ward (2019); US-born whites are assigned a level of zero. Log source country population is the log number of Mexican immigrants for Mexicans and Italian immigrants for Italians; US-born whites are assigned zero. Mexican lynchings are taken from Carrigan and Webb (2003). Standard errors are clustered at the county level.

Table A4. Differences in growth rates between Mexicans, Italians and US-born whites

	I	II	III	IV
Mexico	-2.701	-16.903	-14.522	-17.374
	(0.629)	(0.384)	(0.487)	(0.843)
Italy	-2.486	-7.480	-5.989	-8.643
	(0.224)	(0.239)	(0.271)	(0.552)
Observables from first observation:				
Percentile Rank		-0.413	-0.419	-0.551
		(0.004)	(0.004)	(0.001)
Literate			8.135	5.485
			(0.218)	(0.168)
Can speak English			0.579	-0.158
			(0.173)	(0.108)
Urban area				5.687
				(0.131)
Segregation level				-9.774
				(1.136)
Log source country pop.				-0.308
				(0.030)
Log county pop.				0.115
				(0.166)
Lynching in county in next 10 years				-2.317
				(0.429)
Lynching in county in next 10 years x Mexico				3.775
				(1.221)
Year Fixed Effects	Y	Y	Y	Y
County fixed effects	N	N	N	Y
F-test p-value (Mex=Ita)	0.747	0	0	0
Observations	11,088,451	11,088,451	11,088,451	11,088,451

Notes: Data are from linked censuses (1910-1920; 1920-1930; 1930-1940). The dependent variable is the log earnings growth between observations. Standard errors are clustered at the county level.

Table A5. Assimilation profile of Mexican immigrants using the full-count 1910 Census

	I	II	III	IV	V	VI
	Average percentile Rank			Percent unskilled		
Less than 5 years	-10.840	-13.812	-13.780	-0.183	-0.080	-0.065
	(1.609)	(1.977)	(1.961)	(0.037)	(0.036)	(0.037)
More than 5 years	-14.737	-15.986	-15.972	-0.130	-0.042	-0.033
	(1.483)	(1.792)	(1.778)	(0.033)	(0.036)	(0.037)
Literate	16.822	10.975	11.039	-0.496	-0.304	-0.288
	(1.586)	(1.719)	(1.724)	(0.034)	(0.034)	(0.034)
Can Speak English	20.857	17.592	17.629	-0.400	-0.276	-0.261
	(1.126)	(1.243)	(1.249)	(0.028)	(0.028)	(0.028)
Age	-1.319	-1.604	-1.579	-0.016	-0.027	-0.032
	(0.221)	(0.291)	(0.289)	(0.007)	(0.009)	(0.009)
Age squared	0.020	0.023	0.022	0.000	0.000	0.000
	(0.003)	(0.004)	(0.004)	(0.000)	(0.000)	(0.000)
Broad industry dummies	No	No	Yes	No	No	Yes
Industries	All	Dillingham	Dillingham	All	Dillingham	Dillingham
States	Dillingham	Dillingham	Dillingham	Dillingham	Dillingham	Dillingham
Observations	5,253	2,473	2,473	5,254	2,473	2,473
R-squared	0.280	0.314	0.317	0.142	0.113	0.182

Notes: Data are from the 1910 full-count census (Ruggles et al. 2020). Our aim is to mimic the specification of Feliciano (2001, Table 1), but show that gaps do not converge after five years of stay in 1910. First, we take the 1910 Census, assign earnings scores and percentile ranks, and then collapse the data to calculate the mean percentile rank and percent unskilled by industry (based on the *ind1950* variable) and ethnicity (based on the *bpl* variable). Feliciano (2001) instead uses mean wages as the dependent variable, as calculated from the 1911 Dillingham Commission. The regression is the mean percentile rank/percent unskilled on years in the United States (grouped into less than five years and more than 6 years), fraction literate, fraction fluency in English, age and age squared. We also control for years in the United States for Europeans, which is unreported. We limit to sample to states in the Dillingham Commission (Arizona, California, Colorado, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming). Dillingham industries include mining (ind1950: 206-239), Manufacturing (306-499), and Transportation and Utilities (506-568). The broad industry dummies are at the 1-digit level. The main point of the table is that the gap between Mexicans and US-born whites does not close after five years of stay.

Table A6. Differences in initial gaps across Mexican cohorts

	I	II	III
1900-1904 arrivals	-34.270	-17.244	-18.201
	(1.655)	(1.745)	(1.552)
1905-1909 arrivals	-32.805	-13.821	-15.727
	(1.508)	(1.750)	(1.436)
1910-1914 arrivals	-36.091	-23.159	-21.676
	(1.136)	(0.908)	(1.755)
1915-1919 arrivals	-35.427	-20.523	-21.477
	(1.114)	(0.936)	(1.637)
1920-1924 arrivals	-34.554	-26.432	-26.266
	(1.045)	(0.900)	(1.505)
1925-1929 arrival	-32.209	-23.241	-24.206
	(0.985)	(1.089)	(1.435)
Cristero (1927-1928)	-0.749	-0.872	0.393
	(0.525)	(0.738)	(0.474)
Revolutionary (1914-1915)	-1.060	-1.830	-0.757
	(0.669)	(0.752)	(0.662)
Panic of 1907 (1907-1908)	0.530	-0.215	-0.124
	(0.456)	(0.747)	(0.369)
Literate		29.491	13.001
		(0.648)	(0.165)
Can speak English		6.093	2.016
		(0.483)	(0.201)
Urban			19.056
			(0.364)
Segregation level			-19.511
			(3.355)
Log source country population			-0.083
			(0.057)
Log county population			2.474
			(0.439)
Lynching in prior 10 years			-0.668
			(0.598)
Lynching in prior 10 years x Mexico			-0.466
			(2.197)
Year Fixed Effects	Y	Y	Y
County fixed effects	N	N	Y
Observations	11,036,607	11,036,607	11,036,607

Notes: Data are from linked censuses (1910-1920; 1920-1930; 1930-1940). The dependent variable is the log earnings score at first observation (1910, 1920 or 1930). Urban is defined as living in a city or incorporated area of more than 2,500 people. Border state indicates living in California, Arizona, New Mexico or Texas. The segregation levels are taken from Eriksson and Ward (2019); US-born whites are assigned a level of zero. Log source country population is the log number of Mexican immigrants for Mexicans; US-born whites are assigned zero. Mexican lynchings are taken from Carrigan and Webb (2003). A quartic for years in the United States is included. Standard errors are clustered at the county level.

Table A7. Differences in growth rates across Mexican cohorts

	I	II	III	IV
1900-1904 arrivals	-4.640	-18.739	-13.764	-16.291
	(1.232)	(0.764)	(0.746)	(1.282)
1905-1909 arrivals	-5.071	-18.567	-12.979	-15.459
	(1.126)	(0.669)	(0.660)	(1.348)
1910-1914 arrivals	-2.076	-16.924	-13.200	-14.903
	(0.774)	(0.488)	(0.615)	(1.366)
1915-1919 arrivals	-1.145	-15.720	-11.405	-13.519
	(0.854)	(0.542)	(0.720)	(1.431)
1920-1924 arrivals	-2.294	-16.510	-14.287	-16.572
	(0.713)	(0.425)	(0.499)	(1.272)
1925-1929 arrivals	-3.587	-16.838	-14.352	-16.471
	(0.870)	(0.611)	(0.588)	(1.525)
Cristero arrivals (1927-1928)	-0.421	-0.730	-0.815	-0.528
	(0.649)	(0.547)	(0.544)	(0.582)
Revolution arrivals (1914-1915)	-0.352	-0.134	-0.369	-0.425
	(0.578)	(0.430)	(0.511)	(0.520)
Panic of 1907 arrivals (1907-1908)	0.129	-0.307	-0.563	-0.454
	(0.687)	(0.557)	(0.646)	(0.704)
Log earnings score		-0.411	-0.418	-0.551
		(0.004)	(0.004)	(0.001)
Literate			9.658	6.430
			(0.142)	(0.101)
Can speak English			1.527	0.628
			(0.189)	(0.125)
Urban area				5.771
				(0.130)
Segregation level				-12.097
				(2.830)
Log source country population				-0.287
				(0.034)
Log county population				0.091
				(0.180)
Lynching in county in next 10 years				-2.307
				(0.389)
Lynching in county in next 10 years x Mexico				3.675
				(1.136)
Year effects	Y	Y	Y	Y
County fixed effects	N	N	N	Y
Observations	11,036,607	11,036,607	11,036,607	11,036,607
R-squared	0.000	0.206	0.209	0.253

Notes: Data are from linked censuses (1910-1920; 1920-1930; 1930-1940). The dependent variable is the log earnings growth between observations. Standard errors are clustered at the county level. Italy is dropped from the regression.

Table A8. Determinants of census to census change in English fluency for Mexican immigrants

	I	II	III	IV	V
Mexico	-0.059	-0.375	-0.370	-0.344	-0.317
	(0.007)	(0.005)	(0.005)	(0.008)	(0.010)
Observables from t-1:					
Can speak English		-0.909	-0.923	-0.927	-0.929
		(0.004)	(0.004)	(0.004)	(0.004)
Literate			0.055	0.054	0.055
			(0.004)	(0.004)	(0.004)
Urban			0.009	0.017	0.019
			(0.004)	(0.004)	(0.005)
Segregation level				-0.038	-0.044
				(0.012)	(0.025)
Years in US FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
State FE	N	N	N	Y	Y
County FE	N	N	N	N	Y
Observations	51,315	51,315	51,315	51,315	51,315
R-squared	0.057	0.589	0.590	0.592	0.602

Notes: Data are from linked samples for 1910-1920 and 1920-1930. The 1940 Census does not include the English fluency variable. The dependent variable is the change in English fluency from *t-1* to *t*. The negative coefficient on "Can speak English" is because those who can speak English at first observation cannot acquire English fluency by the second observation.

Table A9. Determinants of census to census change in segregation level for Mexican immigrants

	I	II	III	IV	V
Mexico	0.040	0.038	0.041	0.066	0.072
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Observable from *t-1:*					
Literate			-0.003	-0.003	-0.002
			(0.001)	(0.001)	(0.001)
Segregation level		-0.674	-0.681	-0.720	-0.773
		(0.003)	(0.003)	(0.003)	(0.005)
Urban			0.009	0.006	0.003
			(0.001)	(0.001)	(0.001)
Years in US FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
State FE	N	N	N	Y	Y
County FE	N	N	N	N	Y
Observations	74,706	74,706	74,706	74,706	74,706
R-squared	0.082	0.520	0.520	0.528	0.560

Notes: Data are from linked samples for 1910-1920, 1920-1930 and 1930-1940. The dependent variable is the change in segregation from the US-born from *t-1* to *t.* The segregation measure is taken from Eriksson and Ward (2019).

Table A10. Association between Mexican lynching and percentile rank growth between *t-1 and t*

	I	II	III

Lynching in next 10 years *(between t-1 and t)*	-2.317	-1.606	
	(0.429)	(0.642)	
Lynching in next 10 years x Mexico	3.775	3.050	
	(1.221)	(1.031)	
By timing of lynching (prior or future lynching):			
Lynching in prior 10 years *(between t-2 and t-1)*		0.487	
		(0.835)	
Lynching in prior 10 years x Mexico		1.217	
		(1.157)	
Future lynching in county *(between t and t+1)*		-8.657	
		(0.787)	
Future lynching in county x Mexico		4.037	
		(1.290)	
By migration status:			
Moved away from a lynching county			-1.625
			(0.892)
Moved away from a lynching county x Mexico			4.294
			(1.532)
Stayed in a lynching county			-2.950
			(0.983)
Stayed in a lynching county x Mexico			1.648
			(1.612)
Moved into a lynching county			-1.867
			(0.560)
Moved into a lynching county x Mexico			-3.238
			(0.852)
Baseline controls from main regression	Y	Y	Y
County Fixed Effects	Y	Y	Y
Year Fixed Effects	Y	Y	Y
Observations	11,088,451	11,088,451	11,088,451
R-squared	0.257	0.257	0.257

Notes: Data are from the 1910-1920, 1920-1930 and 1930-1940 linked samples. The dependent variable is the change in percentile rank from *t-1* to *t*. The first column shows the main lynching coefficients in Table A5, which measure the association for lynching on incumbent percentile rank growth. The second column tests leads and lags: lynching in the prior ten years occurs, for example, between 1900 and 1909 for the 1910-1920 panel. Future lynching is if a lynching occurred between 1920-1929 for the 1910-1920 panel. The third column splits the results by those who moved away from a lynching county (the starting county had a lynching but the ending county did not) and those who stayed in a lynching county (both the starting and ending counties had a lynching). 68 percent of Mexicans in our sample moved out of a lynching county and 32 percent stayed. Results are also estimated for in-migrants: moved into a lynching county coefficients are for those who started in a county that would have a lynching but did not end up in one. Of those ending up in a lynching county, 62 percent were in-migrants and 38 percent were stayers. All columns include baselines controls for initial percentile rank, human capital, and other county-level controls.

Appendix B. Linking Mexican immigrants

Building the set of potential matches.

To study the assimilation of Mexican immigrants, we create a new sample that tracks Mexican immigrants across censuses. In this section, we provide further detail on how the sample was created.

First, we extract the population of 16- to 40-year-old Mexican-born males from the 1910, 1920 and 1930 censuses, who also had been in the United States between 1 and 10 years. We then search for all possible matches in the census ten years later based on first name, last name, country of birth, year of birth and year of arrival. Before searching, we create a new variable in both censuses that reflect the Anglicized version of each first name (e.g. Jose to Joe, Jorge to George) in case immigrants took on an American name after arrival (Biavaschi et al. 2017). For those that do not have an obvious Anglicization, we use their original first name. We will match based on Anglicized names, which means we favor a George to George match equally with a Jorge (anglicized to George) to George match and a Jorge (anglicized to George) to Jorge (anglicized to George) match.

After anglicizing names, we extract a sample of recently arrived males and search for all potential matches ten years later that meet the following criteria:

1) The first letter of anglicized name matches OR the first letter of last name matches
2) Jaro-Winkler distance of the first name is less than 0.20 (0 indicates perfect match)
3) Jaro-Winkler distance of the last name is less than 0.25
4) The year of birth is less than or equal to a three-year difference
5) The year of arrival is less than or equal to a seven-year difference has been in the United States at least 7 years and at most 23 years

Note that when linking the 1930 and 1940 censuses, we cannot use the year of arrival condition since the 1940 Census does not report the year of arrival.

Choosing the best link.

Now we have several candidate links that meet the above criteria, with it being unclear which candidate is the best link.[35] One metaphor for this stage of the linking process is that there is a "police line up" of candidate links, and now we need to pick the best match (Bailey et al. forthcoming).

We follow the method of Feigenbaum (2016) to choose the best link, where we estimate a probit model from a hand-linked dataset that aims to mimic a human choosing the best match. Feigenbaum (2016) argues that you only need to hand-link from a random sample of 500 people before the probit model converges to its estimated coefficients; instead of 500, we take a more conservative approach and instead draw a sample of 2,000 Mexican immigrants from the 1910 census to hand-link to the 1920 census. From this set of 2,000, we only find 1,657 with potential matches in the 1920 census that meet the above criteria, suggesting that the best possible linking rate would be 82.9 percent. Of these 1,657 with at least one potential match, there are 22,149 candidate links, or an average 13.4 people to choose from in the police line up metaphor.

To build a dataset of true links, we create two sets of hand-linked data from two independent linkers, and then reconcile conflicting links after review. Ultimately, we link 644 of the original 2000 for a linking rate of 32.2 percent, or 38.8 percent of those that have potential matches. This linking rate is close to that of Italians, where Ward (2020) uses a similar method

[35] For those with more than 25 matches, we give a match score to each link based on Feigenbaum's (2016) probit model to keep the best 25 potential matches. We are unconcerned that this dropped many links since we often handpick matches that are highly rated by Feigenbaum's algorithm.

and hand links 31.0 percent in his training data. Failure to link could be due to a variety of reasons, including having a common name, anglicizing one's name in a non-obvious manner, return migration, death, errors by enumerators in writing names, failure to enumerate in the second census, and errors by data transcribers who digitized the name. Common names are especially problematic for Mexican immigrants, where a true link cannot be reliably determined; moreover, it is clear that enumerators entered in poor phonetic translations of Spanish names. Return and repeat migration were especially prevalent for Mexican immigrants, where the return rate for 1910 arrivals by 1930 was 44 percent and moving back and forth across the border was very common (Kosack and Ward 2014). Morever, Hacker (2013) notes that under enumeration in the early 20th century censuses were between 5 to 6 percent for native-born whites, which may have been much higher for Mexican-born immigrants.

Modeling the best link.

With our hand-linked dataset of 22,149 observations where 644 of the observations are coded as a link, we can estimate a probit model that determines the best link. We estimate a probit where we include variables that we determined were important given our experience linking individuals. These include variables in the Feigenbaum method such as Jaro-Winkler distance of first name and last name, the difference in year of birth, whether the link is an exact match on first name string, and last name string. We add additional variables such as the difference in year of arrival, the total number of candidate links that had exact matches for first and last name, and the total number of candidate links that had exact match on NYSIIS last name.[36] Table B1 shows the results from the probit. One interesting result is that linking immigrants relies more heavily on a

[36] NYSIIS stands for New York State Identification and Intelligence System. It is an algorithm that codes names based on their phonetics, such that a name like "John" would match with "Jon". This has been used extensively in the linking literature (e.g., Abramitzky et al. 2014).

close match of the last name than close match of the first name. Year of arrival also provides meaningful information about links. The second column shows probit estimate when not including the year of arrival, which is for the link between 1930 and 1940 censuses.

The probit provides a predicted probability for each candidate link in our hand linked data; importantly, we can also use the probit coefficients to score matches for the entire set of links from full-count census to full-count census. However, for the full to full-count data, it is still unclear which predicted links should be kept in the linked dataset. To do so, we must determine two critical values for one to be coded as a predicted link. First, a cut-off for the predicted probability where candidate links above the predicted probability are included in the dataset. However, if two hits that are above the predicted value then we could potentially code two or more links for a given individual. Therefore, we also determine a cut-off for the ratio between the first-best and second-best link.

Table B1. Probit model for linking Mexico-born individuals.

	Linked	Linked
Jaro-Winkler Distance, First name	-3.249***	-3.196***
	(0.454)	(0.442)
Jaro-Winkler Distance, Last name	-12.04***	-11.64***
	(0.884)	(0.858)
Year of Birth Difference=1	-0.275**	-0.260**
	(0.122)	(0.118)

Year of Birth Difference=2	-0.426***	-0.409***
	(0.131)	(0.127)
Year of Birth Difference=3	-0.549***	-0.542***
	(0.140)	(0.135)
Years in US Difference	-0.239***	
	(0.0335)	
Years in US Difference squared	0.0148***	
	(0.00338)	
Total hits	-0.0642***	-0.0640***
	(0.0234)	(0.0229)
Total hits squared	0.000509	0.000534
	(0.000766)	(0.000746)
Unique and Exact NYSIIS Match, first name	0.822***	0.716***
	(0.173)	(0.172)
Unique and Exact NYSIIS Match, last name	-0.0152	-0.0640
	(0.234)	(0.229)
Unique and Exact NYSIIS Match, first and last name	1.570***	1.562***
	(0.136)	(0.135)
Unique and Exact Match, last name	0.491***	0.457**
	(0.190)	(0.185)
Middle initial match	0.988*	0.998*
	(0.523)	(0.513)
First letter of last name match	0.00182	0.0647
	(0.137)	(0.136)
First letter of first name match	0.162	0.113
	(0.149)	(0.143)
1 hit with exact last name NYSIIS match	-0.146	-0.126
	(0.132)	(0.127)
1+ hit with exact last name NYSIIS match	-0.536***	-0.565***
	(0.172)	(0.167)
1 hit with exact last name match	-0.628***	-0.608***
	(0.149)	(0.144)
1+ hit with exact last name match	-1.355***	-1.272***
	(0.123)	(0.119)
One hit	0.242	0.274
	(0.190)	(0.185)
Absolute difference in length of last name	-0.299***	-0.287***
	(0.0482)	(0.0468)
Exact NYSIIS last name match and year of birth diff==0	0.876***	0.886***
	(0.191)	(0.186)
Exact NYSIIS last name match and year of birth diff==1	0.695***	0.676***
	(0.196)	(0.191)
Exact NYSIIS last name match and year of birth diff==2	0.462**	0.451**
	(0.203)	(0.197)
Constant	1.316***	0.720***
	(0.288)	(0.274)
Observations	22,149	22,149

Notes: Column I is used for 1910-1920, 1920-1930 match, Column II for 1930-1940 match.

To determine these cutoffs, we choose a positive predictive value of 0.90 or let the ratio of true links to total links be 0.90 in our training data; this to reduce the number of false positives in our linked data. A cost of keeping the positive predictive value at a high level is that it throws out a lot of potential links that do not meet the strict criteria, whether because there is a second match that is close in value or because the predicted probability is not high enough. After setting the PPV at 0.90, we do a grid search to determine the maximum true positive rate (TPR), which is 0.36. The TPR is the ratio of predicted links to true links in our training data, suggesting that our probit only captures 36 percent of links we find when hand linking. This is a low level of efficiency, reflecting difficulty matching immigrants; for example, in Feigenbaum's training data linking native-born Iowans to the 1940 Census, he has a TPR of 0.881 and a PPV of 0.854. Ultimately our cut-off points are that the predicted probability must be above 0.642 and the ratio of predicted probabilities of the first- and second-best link must be above 1.6.

Finally, we apply the probit and cutoff values to the full to full-count links, ending up with the linking rates shown in Online Appendix Table A1. The expectation is that since we hand linked 32 percent of the sample and the TPR is 0.36, then we should approximately link 11.5 percent of immigrants. We are slightly below this target, linking approximately 10 percent of immigrants. A low linking rate occurs because, for example between 1910 and 1920, person A in 1910 may be linked to person B in 1920, but person B in 1920 may also be linked to person C in 1910. We use the same ratio cut off rule, where we only keep either the A-B or C-B link if the ratio of predicted probabilities is more than 1.6 times the second-best match.

Table B2. Linking Rates for Mexican Immigrants for One-Generation Sample

I Start Year	II Set to link	III End Year	IV Linked Number	V Linking Rate
1910	49,574	1920	5,399	0.108
1920	149,103	1930	15,064	0.101
1930	123,684	1940	11,934	0.096

Notes: The sample are those linked from the census in Column I to the census in Column III. The linked number is not the same in the final sample because we keep only those who have observable occupations, English ability and literacy rates. This drops approximately 2 to 5 percent of the linked sample.

Weighting

To weight the data, we use the inverse probability method (Bailey et al. forthcoming). To implement this method, we take the linked dataset of the 1900-1929 cohorts, append the full-count census data of the same cohorts, and then estimate what observable characteristics predict a link. Importantly, we only use observables from the *second* observation of the linked data: that is, 1920 for the 1900-1909 cohort, 1930 for the 1910-1919 cohort and 1940 for the 1920-1929 cohort. We do not use the first observation/census because the first cross-section includes temporary migrants and we only want to weight to match the outcomes of permanent migrants. However, one issue with the 1940 census is that we cannot separate the 1920-1929 arrival cohort in the full-count census data, so we simply use all Mexican-born.

To create the weights for Mexicans and the US-born, we estimate a probit model (1=in linked data, 0=in cross-sectional data) with the following controls: (1) age fixed effects, grouped into 5-year bins, (2) first digit of 3-digit occ1950 code, (3) ability to read and write (or have more than 8 years of education in the 1940 census) (4) and region of residence. We interact the age controls with the country of birth and the region with the country of birth in case linking probability varies across these subgroups. We also interact the occupation control and literate control with 5-year arrival cohort (e.g., 1900-1904 and 1905-1909 for the 1920 census). We separately weight by 1900-1909, 1910-1919, and 1920-1929 cohorts in case of selection into the sample varying across time.

Appendix C. Imputed earnings score

We measure the gaps across Mexicans, US-born whites and Italians by imputing their earnings and then percentile ranking the earning score by census and age. In this appendix, we provide more detail on how we impute earnings.

Our approach to imputing earnings borrows heavily from Collins and Wanamaker (2020) and Abramiztky et al. (2021)/Saavedra and Twinam (2019). From Collins and Wanamaker (2020), we implement their method of imputing income for self-employed workers, and adding perquisites for farmers and farm laborers. From Abramitzky et al. (2021) and Saavedra and Twinam (2019), we use their method of relying on a regression to predict earnings; in our instance, we mainly use information on occupation, census division, and country of birth.

First, we take the complete-count 1940 census and keep males aged 25-55 with an occupational response. We then replace the top-coded income of $5,000 with 1.4 times the top code, which follows Goldin and Margo (1992). We will later use this income in a regression to predict the earnings by occupation, census division and country of birth. But a major issue is that we do not have income for self-employed workers, which we have to impute. Imputing self-employed income will be important for farmers. Before imputing their income, we separate the farmer occupational code (occ1950=100) into farmer owners and tenant farmers based whether they own their home.

To impute the income for self-employed workers, we turn to the 1960 Census's information on wage, business, and farm income. First, we calculate a total earnings variable by summing the wage, business and farm income. Then for each 1950 occupational code, we estimate the mean earnings for self-employed workers and for wage workers. The ratio of these earnings within occupation provides an estimate of the self-employed worker income relative to wage income by occupation. Going back to the 1940 Census, we then calculate the mean wage income by occupation, country of birth and region for wage workers, and then multiple this wage income by the 1960 ratio to predict self-employed income.

For agricultural workers, perquisites in the form of room and board are not included in the census earnings. For farm laborers, we boost 1940 wage income by 26 percent to reflect these perquisites. This number is taken from Collins and Wanamaker (2020), who use data from a 1957 USDA report *Major Statistical Series of the U.S. Department of Agriculture: Volume 3, Gross and Net Farmer Income*. For farmers, we do not have the base amount of income data to boost from since the 1940 census only has wage income. To calculate income for farmers, we assume that the ratio of farmer income (inclusive of perquisites) to farm laborer income (inclusive of perquisites) is the same in 1960 as it was in 1940.[37] We then calculate the mean farm laborer income at the country of birth and census division level, and then set farmer income to be equal to the product of the 1960 ratio of farmer/farm laborer income and 1940 farm laborer income.

At this point we have an "adjusted" income in the 1940 census, which is the actual income for wage workers, the imputed income for self-employed workers after adjusting for self-employed earnings, the imputed income for farm laborers after boosting wage income to reflect perquisites, and the imputed income for farmer owners and tenants based on the ratio of farmer income to farm laborer income. Following a similar approach to Abramitzky et al. (2021) and Saavedra and Twinam (2019), we regress the adjusted income on a quadratic in age, occupational fixed effects (at the 3-digit level), census division fixed effects, fixed effects for the interaction between country of birth and census division, fixed effects for the interaction between country of birth and 1-digit occupation, and the interaction between 1-digit occupation and division fixed effects:

[37] 1960 perquisites for farmers (35 percent) and farm laborers (19 percent) are also taken from Collins and Wanamaker (2020).

$$\text{AdjIncome} = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Age}^2 + \text{3digitOcc}_{\text{FE}} + \text{Division}_{\text{FE}} + \text{CountryBirth}_{\text{FE}} + (\text{CountryBirth} \times \text{Division})_{\text{FE}} + (\text{CountryBirth} \times \text{1digitOcc}_{\text{FE}}) + (\text{1digitOcc}_{\text{FE}} \times \text{Division})_{\text{FE}} \quad \text{(C1)}$$

Our score is then the average predicted score in each 3-digit occupation, country of birth and census division. Note that Black males are not included for the US-born calculation.

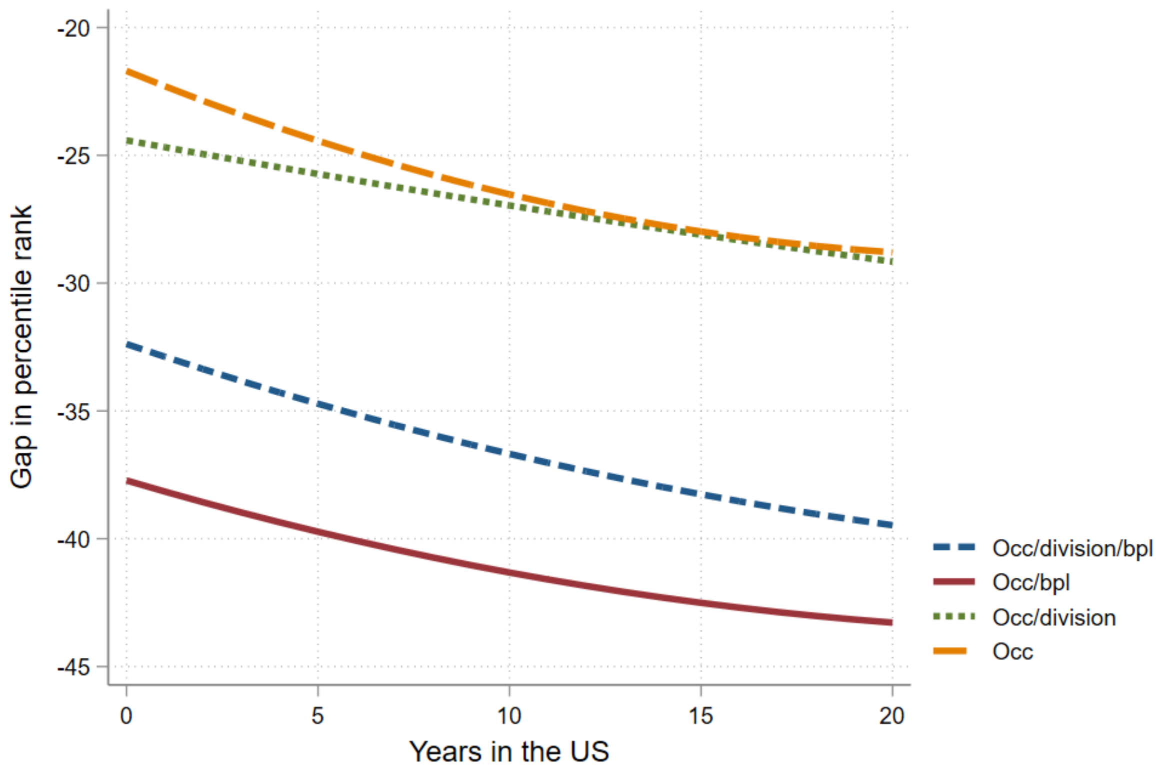
Incorporating more information besides occupation to infer economic status is relatively new in the literature, where most research has relied on occupational income scores. To test the importance of including country of birth and census division in the score, we create alternative scores that their drop fixed effects in Equation (C1). The results are shown in Figure C1.

Figure C1 shows that changing the earning score influences the gap in percentile rank with US-born whites, but does not have a strong influence on the slope. First, when including only the quadratic in age and the 3-digit occupational fixed effects, we still estimate that Mexican immigrants fell behind US-born whites. Note that this method is similar to using the 1950 occupational income score available in IPUMS (*occscore)*. When removing country of birth and census division, we estimate a smaller gap between Mexicans and US-born whites. Instead of starting with a 33-percentile gap, the occupational earnings score instead estimates about a 22-percentile gap. This result is consistent with Mexican immigrants earning less than US-born whites within occupation.

The key additions that lead to wider gaps are the country of birth fixed effects and their interaction with 1-digit occupational fixed effects. If one excludes division information, we actually find larger gaps with the US-born, at about 38 near arrival instead of 33. The combination of this information suggests that Mexicans are paid less than natives within occupation at the national level, but this gap is not as large when including census division.

Either way, a main point observed in Figure C1 is that negative assimilation for Mexicans holds no matter which earnings score one uses.

Figure C1. The gaps in percentile ranks differ when using various earnings scores, but negative assimilation holds



Notes: Data are from our linked sample between 1910-1920, 1920-1930 and 1930-1940. This figure estimates the main specification for assimilation in percentile ranks, but varies the earnings score imputation. Our main score uses information on occupation, census division and country of birth (Occ/division/bpl) to impute earnings. If one uses only occupation, the gaps narrow (Occ). If one uses occupation and census division (Occ/division), assimilation is similar to just using occupation. If one uses occupation and country of birth (Occ/bpl), then the gaps are the widest.