



UNIVERSITY OF OXFORD

**Oxford Economic and Social  
History Working Papers**

**Number 183, September 2020**

---

**Life after Crossing the Border:  
Assimilation during the First Mexican Mass Migration**

**DAVID ESCAMILLA-GUERRERO, EDWARD KOSACK,  
AND ZACHARY WARD**

---

# Life after Crossing the Border: Assimilation during the First Mexican Mass Migration

September 2020\*

David Escamilla-Guerrero  
University of St Andrews  
University of Oxford

Edward Kosack  
Xavier University

Zachary Ward  
Baylor University

**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. We show that despite these large shifts, the rate of economic assimilation was remarkably similar across arrival cohorts. We find that the average Mexican immigrant held a lower-paying job than US-born whites near arrival and further *fell* behind in the following decade. However, 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 earning score, Mexicans had a slower growth rate than both US-born whites and Italians. We argue that Mexican-specific structural barriers help to explain why Mexican progress was slower than other groups and why different Mexican arrival cohorts had limited variation in outcomes despite the large shocks to migration.

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

**Keywords:** assimilation, immigration, Mexico, mobility, mob violence

\* david.escamilla-guerrero@st-andrews.ac.uk, School of Economics and Finance, University of St Andrews, Castlecliffe, The Scores, St Andrews, KY16 9AZ, Scotland. Oxford Centre for Economic and Social History, University of Oxford. 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.

“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)<sup>1</sup>

Mexican immigration to the United States underwent a significant transformation during the early twentieth century. What had been a small flow turned into a mass migration as the American Southwest boomed, deadly conflict in Mexico raged, and migration restrictions on Europe made Mexico one of the most important source countries. The rapid increase in migration was met with an equally rapid decrease in the 1930s as the Great Depression and a nativist backlash led to the deportation and repatriation of hundreds of thousands of Mexicans and Mexican Americans. In this paper, we analyze the economic outcomes of Mexican immigrants during this tumultuous period and document how they fared both across arrival cohorts and relative to other groups in the United States.

To understand whether the early twentieth century Mexican experience was unique, we contrast it with another important group: Italians. Historical Italians are often compared to modern-day Mexicans because both groups arrived relatively unskilled (Abramitzky 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.<sup>2</sup> This comparison allows us to uncover how Mexican immigrants fared relative to a group

---

<sup>1</sup> Adapted from song of the Cotton Harvest (Dickey 2006).

<sup>2</sup> 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.

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 the Mexico-US-born gap in economic outcomes *widened* with more years of stay. That is, Mexican immigrants experienced negative (or reverse) assimilation. After imputing earnings based on occupation, country of birth, and region, our baseline estimate is that Mexican immigrants fell behind US-born whites by 12 log points after one decade. This estimate comes from new linked data for 1900-1929 arrivals, which allows us to accurately estimate migrant assimilation without bias from selective return migration (Borjas 1985, Abramitzky et al. 2014). While the average Mexican fell behind US-born whites, we also find that the average Italian did as well. In fact, both groups fell behind US-born whites by the same amount. This result suggests that Mexicans and Italians did indeed economically assimilate at the same pace in the past. A negative assimilation rate is surprising because most modern-day research finds that immigrants catch up, in part due to acquiring United States-specific human capital (Borjas 2015, Chiswick 1978, Lubotsky 2007).

Despite the large shocks to Mexican migration during this period (e.g., the Panic of 1907, the Mexican Revolution, the Quota Acts, the Cristero War, and the Great Depression), we find that the assimilation rate was strikingly consistent across Mexican arrival cohorts. When we split the 1900-1929 arrivals into 5- or 10- year arrival cohorts, we find that all cohorts fell behind US-born whites by a similar amount. Not only were the growth rates alike across cohorts, but also the arrival gaps with US-born whites were similar. This pattern provides 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.

Observable differences across Mexicans and US-born whites 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. While Mexicans fell behind US-born whites by 12 log points on average, they fell behind by 32 log points when compared with those starting in the same location, with the same human capital, and the same initial earnings score. Further, conditioning on observables (such as urban status, residential segregation and networks) causes a gap in assimilation rates to appear between Mexicans and Italians, with Mexicans falling behind by 18 log points. One reason why the difference may occur is because of differential selection from source countries. However, selection does not appear to be the reason for the slower Mexican growth rate since Mexicans were positively selected from the source whereas Italians were negatively selected (Escamilla-Guerrero 2020, Kosack and Ward 2014, Spitzer and Zimran 2018).

We argue that part of the reason why Mexicans fell behind is because of Mexican-specific barriers to progress. Mexican immigrants experienced numerous barriers in the form of labor market discrimination, expropriation of land and resources, exclusion from public goods or assistance, ethnic animus, and mob violence (Gratton and Merchant 2015). For example, Mexicans were deported *en masse* during the Great Depression, while Italians were not. To provide direct evidence for the existence of barriers, we turn to Mexican lynching data from Carrigan and Webb (2013) and estimate how lynching was associated with economic mobility. To the extent that lynching can be documented, the Mexican lynching rate was about three-fourths of the Black lynching rate (Carrigan and Webb 2003, Table 1). We find that Mexicans who lived in a county with a documented Mexican lynching further fell behind by an additional 16 log points. Lynching is an extreme form of anti-Mexican animus, so we interpret the negative association as reflective

of a broader environment that limited Mexican progress. This finding of a negative association between lynching and economic mobility contributes to a lynching literature that predominantly focuses on Black victims (Cook 2013, Tolnay and Beck 1995, Williams 2017).

This paper complements the *intergenerational* analysis of Mexican outcomes in Kosack and Ward (2020). The primary difference is that we focus on *intragenerational* assimilation, like in Hatton (1997), Minns (2003), Abramitzky et al. (2014), and Collins and Zimran (2019). This paper also covers more key periods of the Mexican mass migration since the data observes 1900-1929 cohorts, while Kosack and Ward (2020) only observe pre-Revolutionary immigrants. When combining our results, it suggests that Mexican progress was slow: gaps *widened* within the first generation, and *zero* convergence occurred across generations. Only after World War II did gaps across Mexican Americans and US-born whites start to converge (Gratton and Merchant 2015).

Our main contribution is to document Mexican assimilation patterns during a key period of mass migration, but we also contribute to the 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). We find that Mexican assimilation was constant throughout the period and so had little association with sudden shifts in the migrant composition (whether due to violent conflict in Mexico or changes induced by US migration policy). 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.

## I. A brief overview of the rise of Mexican mass migration in the early 20<sup>th</sup> century

Before the Mexican Revolution and the sharp rise in immigration, a mix of push and pull factors shaped immigration during the 1900s. Employment opportunities in agriculture, mining,

and railroads increased with the economic boom of the American Southwest (Cardoso 1980, Clark 1908, Gratton and Merchant 2015, Kuntz 1995, Taylor 1929). These conditions attracted mainly unskilled laborers from the north and central plateau of Mexico (Henderson 2011, López-Alonso 2007, Oñate 1991, Rosenzweig 1965). As migration networks developed, potential migrants gained access to information and assistance which further reduced migration costs (Escamilla-Guerrero 2019, Morales 2016).<sup>3</sup> A system of labor recruiting, the *enganche*, also led to mass migration (Escamilla-Guerrero 2020). American employers “hooked” laborers in Mexico by offering wages in advance and covering transportation costs (Brass 1990, Durand 2016). However, since much of the immigration was thought to be temporary, US officials did not pay much attention to Mexican arrivals, despite the rapid growth in population.<sup>4</sup>

Mexican immigration further increased 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).<sup>5</sup> 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). Many refugees were high-skilled (Kosack and Ward 2014, Hall 1982, p. 27); however, it is unclear how many of these high-skilled refugees stayed long after the war ended. For instance, Alanis (2003, p. 411) documents that in 1917 the Mexican government eliminated the customs duty, which created incentives for refugees to return.

---

<sup>3</sup> Three rail lines connected central Mexico to the border by 1900, which reduced significantly traveling times towards the United States (Woodruff and Zenteno 2007). Although it is not clear if train travel was affordable for the impoverished working class (Coatsworth 1979), Mexican immigration increased sharply and expanded its range of settlement (see Figure 1).

<sup>4</sup> For example, the Dillingham Commission 1907-1910 has relatively scarce information on the Mexican immigrant (Benton-Cohen 2011).

<sup>5</sup> Gamio (1930) and Cardoso (1980) place the emigration estimate higher at about 700,000 and 2 million, respectively.

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). As part of the program, this exemption lasted until 1921, when the open-door policy for Mexican immigration effectively ended. Soon after, Mexican immigrants were required to pay a ten-dollar visa fee, raising the costs of crossing the border through official entrance ports and creating, for the first time in history, the incentives for undocumented immigration (Cardoso 1980, Henderson 2011, French et al. 1930).<sup>6</sup>

While admission costs rose, the Immigration Acts of the 1920s did not impose quotas on Mexican natives; subsequently, Mexican immigration boomed as unskilled labor from Europe was halted. Thousands moved northward to cities like Chicago, replacing the excluded labor from Europe (Abramitzky et al. 2019, Innis-Jimenez 2013). Soon came a nativist backlash against the increasing flows (Hoffman 1976). Unauthorized entry was criminalized, and the Labor Appropriation Act of 1924 established the Border Patrol to control areas between entry stations.<sup>7</sup> Yet, the Cristero War (1926-29) led to another exodus of refugees from Mexico.<sup>8</sup> Although most religious exiles—priests and nuns—filled skilled occupations in the United States, 42% of the flow consisted of wives or minors without occupations, leaving unclear whether the skill composition of arrival cohorts changed during the conflict (Alanis 2017, p. 14).

---

<sup>6</sup> Yet the 1918 Immigration Commissioner Report noted that undocumented entries likely increased following the 1917 law (1918 Report, p. 319).

<sup>7</sup> The Border Patrol also worked in areas up to a hundred miles away from the border, sometimes arresting immigrants without a warrant (Ngai 2003).

<sup>8</sup> The conflict induced new waves of immigrants mostly from the central plateau of Mexico, intensifying the preexisting regional composition of the flow (Young 2012).



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 first and second-plus generation Mexican Americans were repatriated or deported (Gratton and Merchant 2013, Verduzco 1995).<sup>9</sup> 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.

Throughout this entire period between 1900 and 1940, Mexicans faced systematic labor market discrimination and mob violence. Mexican immigrants were paid lower wages than US-born white workers of the same class (Clark 1908, Gamio 1930, Guerin-Gonzales 1994). Further, based on college students' surveys, Mexicans were also among the least favored ethnic group (Kosack and Ward 2020). Carrigan and Webb (2013) document that at least 296 Mexicans were lynched from 1880 to 1928 in the American Southwest, most of the time for unconfirmed crimes. Hence, barriers to progress from discrimination could have limited the economic mobility of Mexican immigrants.

## II. Estimating economic assimilation

To measure the economic assimilation of Mexican immigrants, we compare the age- and year-adjusted outcomes of Mexicans to that of US-born whites. To do so, we first estimate the US-born white lifecycle profile:

$$\ln(y_{it}) = f(\text{Age}_{it}) + \theta_t + \varepsilon_{ict} \quad (1)$$

---

<sup>9</sup> Balderrama and Rodríguez (2006) estimate the return flow at more than one million.

where  $y_{ict}$  is the log earnings score of individual  $i$  in year  $t$  and  $f(Age_{it})$  is a quadratic in age.<sup>10</sup> We also include a control for census year. We then predict the residuals for Mexicans (and Italians) to measure how far behind immigrants are relative to US-born whites for a given age and year.<sup>11</sup>

Our interest is how the gap between Mexicans, Italians, and US-born whites changes across arrival cohorts and within arrival cohorts. We group cohorts into five-year intervals (1900-1904, 1905-1909, ... , 1925-1929) and estimate the gaps when they are observed between the 1910 and 1940 censuses. If large differences in the relative gap across cohorts emerge, then this would capture differences in cohort “quality” (Borjas 1985). For instance, if the Mexico/US-born white gap is smaller for the 1910-1914 cohort relative to other cohorts, then this could reflect that refugees had better outcomes because they were more positively selected from Mexico.

If within-cohort growth is greater for immigrants than for natives such that they close the gap, then migrants economically “assimilated.” However, if growth is smaller for migrants, then they experienced negative or reverse assimilation. The expectation is that immigrants started below the US-born and had faster growth throughout the lifecycle (Chiswick 1978). This pattern is often attributed to greater human capital investment by immigrants, who need to acquire English fluency or US-specific qualifications. However, the Mexican profile may deviate from this expectation. For instance, early 20<sup>th</sup> century Europeans did not gain on US-born whites (Abramitzky et al. 2014). Moreover, when controlling for the initial occupational income, Europeans fell behind (Collins and Zimran 2019). One reason why immigrants may have fallen behind is that they were

---

<sup>10</sup> While the modern-day literature often calculates gaps with all US-born individuals (including non-white), we follow the historical literature and calculate it with respect to the white US-born population. As opposed to others (e.g., Abramitzky et al. 2014, Collins and Zimran 2019), we include southern states because many Mexicans resided in Texas. We will also show assimilation estimates when controlling for location.

<sup>11</sup> We will later estimate a standard assimilation specification that includes a quadratic in years in the United States and arrival cohort fixed effects, where the US-born whites are the excluded group. However, this specification allows us to more flexibly measure differences in within-cohort growth across arrival cohorts.

negatively selected on unobserved ability (Abramitzky et al. 2012, Borjas 1987).<sup>12</sup> Another reason is that structural barriers halted migrant progress in the labor market relative to those starting in the same occupations.

Accurately estimating economic assimilation requires panel data. Otherwise, repeated cross-sections conflate within-cohort growth with selective return migration (Abramitzky et al. 2014, Borjas 1985). Mexican return migration rates were high, around 44 percent between 1920-1930 (Kosack and Ward 2014). To avoid the problem of selective return migration, we build a new longitudinal data of Mexican immigrants, which we discuss in the next section.

### III. Data

#### *Linking Mexican immigrants.*

To estimate Mexican immigrant's rate of economic assimilation, we build a linked dataset from the 1910, 1920, 1930 and 1940 complete-count United States Censuses, accessed at the National Bureau of Economic Research (Ruggles et al. 2020).<sup>13</sup> We follow a machine-learning approach (Feigenbaum 2016). First, we hand linked a random sample of 2,000 Mexican immigrants across the 1910-1920 census. 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 then 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

---

<sup>12</sup> Historical studies suggest that migrants were mostly negatively or neutrally selected relative to the source population (Abramitzky and Boustan 2017, Hatton and Ward 2018), but it is unclear how this selection compares to the US-born population.

<sup>13</sup> 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 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.

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.<sup>14</sup> 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).<sup>15</sup> We describe the full linking method in detail in Appendix B. Note that the assimilation rates are robust to the automated linking methods from 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.<sup>16</sup>

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 within the last ten years at first observation and who are between 16 and 40 years old.<sup>17</sup> 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. This leaves us with a sample of 4,628 Mexicans linked between 1910 and 1920, 13,269 between 1920 and 1930, and 10,163 between 1930 and 1940. The total number is 28,060.

We compare the occupational upgrading of Mexicans to that of Italians and US-born whites. The Italian dataset comes from Ward (2020) and the US-born white datasets are from Ward

---

<sup>14</sup> We do not use the year of arrival for predicting the 1930-1940 link.

<sup>15</sup> Pérez (2017) does a similar progress for linking immigrants in Argentina.

<sup>16</sup> 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.

<sup>17</sup> We do not link those who arrived in the same year as the census since we do not observe the full arrival cohort.

(2019, forthcoming).<sup>18</sup> All datasets were linked using the same machine-learning approach and, importantly, have the same precision rate such that the links are expected to be of equal quality.<sup>19</sup> (That is, the expected false positive rate is 10 percent.) Therefore, differences in outcomes across groups should not be due to differences in false positives (Bailey et al. forthcoming). The Mexican immigrant linking rate (9-11 percent) is lower than that of the main comparison group of US-born whites (32-36 percent) but is higher than for Italians (4-6 percent). The comparison group of US-born whites includes 13,516,131 observations, and the Italian group includes 60,043.

A major concern with the linked samples is that they are unrepresentative of the permanent migrant population. To account for unrepresentativeness, we weight our linked sample using inverse probability weights based on age fixed effects; 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). 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 cohort in case the weights vary across arrival 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.

---

<sup>18</sup> 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.

<sup>19</sup> 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.

*Earnings score proxy.*

Due to data limitations in censuses before 1940, we estimate earnings score convergence, rather than wage assimilation. Closely following the process in Collins and Wanamaker (2017), our earnings proxy primarily relies on the 1940 Census information on wage income for wage workers. However, self-employed income is not included in the 1940 Census. Therefore, we impute self-employed income in the 1940 Census based on the 1960 Census, where we assume that the ratio of total income to wage income by occupation is the same in 1960 as it was in 1940. We multiply this ratio by the average wage income by occupation in the 1940 Census for self-employed workers. This gives us an estimate of the total income for self-employed workers in 1940. We also add perquisites for farmer and farm laborer income using information from the United States Department of Agriculture (USDA). We then regress this adjusted earnings measure (that includes estimated self-employment income and perquisites) on occupation (3-digit code), region of residence, a quadratic in age, country of birth, country of birth interacted with region, occupation (1-digit code) interacted with country of birth, and occupation (1-digit code) interacted with region.<sup>20</sup> The mean predicted income by 3-digit occupation, region and country of birth is used as our earnings proxy. To provide a clearer understanding of these scores, Online Appendix Table A1 shows the average earnings in our dataset for 1-digit occupational codes. The estimates suggest that Mexicans earned less than US-born whites and Italians within broad occupational group; for example, average professional earnings are estimated to be \$2,054 for US-born whites, \$1,346 for Italian immigrants and \$1,247 for Mexican immigrants.

---

<sup>20</sup> This regression is similar in spirit to Abramitzky et al. (2019) and Saavedra and Twinam (2019).

Note that earnings score assimilation 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). Another issue is that we apply 1940 information to earlier censuses between 1910 and 1930. To address this issue we estimate results for percentile ranks, which may be more accurate if an occupation, country of birth, and region's position in the distribution is similar between 1910 and 1940. Finally, we will also estimate assimilation for occupational categories, which do not rely on earning score assumptions.

#### IV. The assimilation of Mexican immigrants

The descriptive statistics in Table 1 show that Mexican immigrants started with lower earnings scores than US-born whites and Italians. The gaps were large such that at first observation Mexicans had earnings scores that were 54 log points (42 percent) lower than US-born whites and 51 log points lower than Italians. Alternatively, Mexicans started off 28 percentiles lower than US-born whites and 19 percentiles lower than Italians. This pattern is partially explained by lower earnings for Mexicans within occupation, but Mexicans were more heavily concentrated in lower-skilled occupations. For instance, 83 percent of Mexican immigrants were in an unskilled occupation at first observation, in contrast to 69 percent of Italians and 42 percent of US-born whites.

In addition to the size of the initial gap, another striking pattern is that Mexicans had slower earning score growth from the first to the second observation. While US-born whites gained 13 log points, Mexico-born immigrants gained *zero* points. A zero gain is surprising and suggests that opportunities for upward mobility were nonexistent. Due to this zero gain, Mexicans fell behind US-born whites in relative terms. Table 1 also shows that Italian immigrants only gained 3 log points from the first to the second observation, which suggests that Italians also fell behind US-

born whites. Therefore, Mexicans and Italians appear to have had similar growth rates from census to census; yet, they had large differences in their initial earning score.

Further illustrating the slow growth rate for Mexicans, Figure 2 shows how the distribution of Mexican percentile ranks changed between the first and second observation. Mexicans who had stayed 11-20 years were lower in the distribution than those staying 1-10 years (percentile ranks are calculated within birth cohort). Note that Mexicans rarely made it above the 60<sup>th</sup> percentile and had a bimodal distribution around the 10<sup>th</sup> percentile (mostly farm laborers) and 30<sup>th</sup> percentile (general laborers and operatives).

The gaps in Table 1 potentially mask important differences across arrival cohorts. In Table 2, we present across- and within-cohort outcomes by five-year arrival cohorts. This table shows that the Mexico/US-born white gap *widened* over time for *all* cohorts of arrival. Therefore, it was not a few key cohorts or years that led to Mexicans falling behind. Rather, all cohorts experienced negative assimilation. Moreover, the magnitude of the gap near arrival, which indirectly measures “cohort quality,” does not vary much across arrival cohorts. Those who had stayed less than 5 years had a gap between negative 47 and 53 log points, whereas those who remained 6-10 years had a gap of negative 57 to 61 log points. Similar to this lack of variation in the initial gap, there is also little variation in the within-cohort growth rate. For cohorts that had already been in the United States for 6-10 years (1900-1904; 1910-1914; 1920-1924), the rate of within cohort is at negative 11 to 15 log points. For cohorts who had been in the United States 1-5 years (1905-1909; 1915-1919; 1925-1929), the within-cohort growth rate is between negative 17 and 20 log points. The fact that more recent arrivals fell behind at a greater rate suggests that Mexicans did not improve their outcomes rapidly after arrival.



Falling behind contrasts with results from Feliciano (2001, Table 1) that 1910 Mexicans *completely* closed the gap after five years of stay. There are numerous differences between our analyses, but a major one is the data. Feliciano uses average earnings by industry-ethnic group rather than census microdata, where her averages are taken from the 1911 Dillingham Commission. These data are advantageous in that one does not need to resort to the 1940 Census to estimate earnings. However, the Commission only covered a specific set of industries and a specific set of states. For instance, the Commission data did not have earnings estimates for agricultural workers and only surveyed states that covered 36 percent of the 1910 Mexican population. Therefore, we view our data as more complete and thus more accurate for Mexican assimilation.

Overall, the results show a remarkable consistency of negative assimilation for Mexican immigrants in the early 20<sup>th</sup> century. This finding suggests that the Revolutionary-era cohort (1910-1919) was not unique, as it had similar within-cohort growth rates to other Mexican cohorts. This result is notable because arrival records from the 1910s show that the flow became more skilled and literate, especially from 1914 to 1915 (Kosack and Ward 2014, Hall 1982). One explanation for the similar assimilation rate of the Revolutionary cohort is that most (positively-selected) refugees had returned home by the 1920 census, when we are first able to observe the cohort. If so, then our results may capture the assimilation of economic migrants rather than refugees, and the lack of a difference in assimilation rates may not be that surprising.

To test for selection into return migration, we follow the method of Abramitzky et al. (2014) and compare within-cohort growth rates from panel data to that from repeated cross-

sectional data.<sup>21</sup> The within-cohort growth rate is slightly *faster* in the repeated cross-sections, which suggests *negative* selection into return migration (see Figure A1). In other words, part of the reason why there is a faster growth rate in the cross-sections is that low-skilled Mexicans returned home. Back of the envelope calculations suggest that return migrants had 3.4 to 18.4 percent lower earnings scores than permanent migrants at first observation (see Appendix Table A2).<sup>22</sup> Negative selection appears to have been stronger during the Mexican Revolution, which contrasts with the story that high-skilled refugees returned home. Once again, however, it may be that positively selected refugees had returned before 1920, and we are measuring negative selection of return migrants who had already stayed until 1920. Future research could determine whether high-skilled refugees returned home by using border crossing records between 1910 and 1920 and linking to the Mexican and United States Censuses.

Panel B of Table 2 shows that Mexicans were not unique in their negative assimilation, but that Italian immigrants also fell behind US-born whites. The magnitude of within-cohort growth was similar to Mexicans between 1900 and 1919, where Italians fell behind by 15 to 17 log points (in contrast with 14 to 20 log points for Mexicans). After the 1920s, assimilation patterns diverged where the Italian assimilation rate increased in relative terms. At the same time, the initial gap dropped for 1920s Italians, which is surprising because the quotas increased the skill composition of arrivals (Massey 2018). These conflicting patterns may be because low-skilled individuals who entered under the quota system were more likely to remain until 1930 rather than returning home;

---

<sup>21</sup> The 1940 census does not include the year of immigration, so we cannot calculate within-cohort growth rates between 1930 and 1940.

<sup>22</sup> Abramitzky et al. (2014, Figure 5) note that the RCS gap is a weighted average of the permanent migrant gap and temporary migrant gap. Therefore, the difference in earnings scores between permanent and temporary migrants can be calculated after dividing the difference between the growth rate in the panel relative to the growth in repeated cross-section with the return rate. For example, with the 1910-1914 cohort the selection of return migrants can be calculated as  $(0.151 - (-0.102)) / 0.44 = -0.111$ .

moreover, those lucky enough to enter had faster growth rates (Ward 2017). However, it is far from clear that the quotas are the sole reason for the increase in assimilation rates for the 1920s Italian cohort. For example, the Great Depression may have been more harmful to Mexicans than Italians.

The negative assimilation rate for Italians contrasts with results from Abramitzky et al. (2014) that Italians stagnated in relative terms between 1900 and 1920. The difference in results is not due to the linking method. When we use the publicly available links from the Census Linking Project and keep conservative NYSIIS links (Abramitzky et al. 2020), then we also find a negative assimilation rate for Italian immigrants.<sup>23</sup> Specifically, while our method finds that Italians fell behind by 10 log points, the conservative ABE method finds a fall of 7 log points. The assimilation rates for Mexicans are also similar across linking methods, where we find a 13-log point drop and the ABE method finds a 12-log point drop. We prefer our method of linking, in part because we locate 180 percent more Mexican immigrants (28,060 v. 9,970). One reason for a difference in assimilation rates may be because assimilation rates changed over time. We estimate growth rates for post-1900 arrivals, in contrast to the 1880-1900 arrivals in Abramitzky et al. (2014); it may be that post-1900 arrivals had slower relative growth rates because structural transformation steepened the lifecycle profile for natives (Collins and Zimran 2019).

Panel A of Figure 3 shows the predicted Mexican and Italian assimilation profiles based on the number of years in the United States.<sup>24</sup> From arrival to 20 years later, Mexicans are predicted to fall behind US-born whites by 29 log points. In percentage terms, the Mexico-US born gap

---

<sup>23</sup> To match our sample, we only allow the year of arrival to differ by at most 7 years for the ABE method.

<sup>24</sup> That is, we estimate  $\ln(y_{ict}) = g(\text{YearsInUS}_{ict}) + f(\text{Age}_{ict}) + \gamma_c + \theta_t + \varepsilon_{ict}$ . In this regression, the log earnings score 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.

increased from 38 percent upon arrival to 53 percent after 20 years of stay. This result reflects that the Mexican lifecycle profile was essentially flat, while the lifecycle profile for US-born whites increased. In contrast, Italian immigrants fell behind US-born whites by 24 log points (23%). If one uses percentile ranks instead of log earnings score, which may suffer less from issues when extrapolating 1940 earnings backward to early decades, Mexicans fell behind by 9.5 percentiles and Italians fell behind by 11 percentiles (see Figure A2).

It may be that the assimilation rate of Mexicans varied across destination choices. Immigrants who stayed closer to the border in the border states may have benefited from stronger immigrant networks and lower migration costs than those who moved further north.<sup>25</sup> As strong networks and low migration costs predict a negative selection of immigrants, we would expect a slower rate of assimilation in this area (McKenzie and Rapoport 2010). Further, the historical association with living in an enclave and occupational upgrading is negative, though modern-day evidence suggests a positive effect (Abramitzky et al. 2020, Edin et al. 2003, Eriksson 2020). When splitting the sample into border states and non-border states, we find that both groups experienced negative assimilation (see Panel B of Figure 3). However, negative assimilation was stronger in the border states: immigrants fell behind by 35 log points, while non-border state migrants fell behind by 23 log points.<sup>26</sup> The initial gap was also smaller in non-border states such that after 20 years of stay, Mexicans had a 36 percent earnings gap with US-born whites in non-border states, but a 55 percent earnings gap in border states. However, only 10 percent of Mexican

---

<sup>25</sup> Gratton and Merchant (2015) show that most persons of Mexican origin were in border states.

<sup>26</sup> We show the estimated assimilation rates when controlling for state or county of residence in Online Appendix Figure A3. When one splits the sample into border counties (within 100 kilometers of the border) and non-border counties, then there is little difference in assimilation rates across groups.

immigrants were in non-border states, which makes it unclear if these patterns across border and non-border states are due to different levels of human capital.

The similarity of assimilation rates for Italians and Mexicans may be surprising given the difference in the acquisition of basic English fluency. In panel A of Figure 4, we plot the gain in English fluency across Italian and Mexican immigrants in the early 20<sup>th</sup> century. Mexican immigrants arrived with only 16 percent able to speak English near arrival (according to the judgment of the enumerator) and after twenty years of stay the number increased to 56 percent, indicating a rapid increase in English fluency. Italian immigrants both started out with greater English ability and had a larger increase in English fluency, where after twenty years the level increased from 27 to 92 percent. A possible explanation for the similar assimilation rates across groups, despite different English acquisition rates, is that English fluency had little impact on earnings score (Ward 2020).<sup>27</sup>

The slow acquisition of English fluency may suggest that Mexican immigrants were more residentially segregated from the US-born than Italians, but this was not the case. Panel B of Figure 4 shows the residential segregation levels of Mexicans and Italians at the county level, based on the neighbor segregation measure from Eriksson and Ward (2019). The segregation index measures the difference between the expected number of US-born households that are next-door neighbors relative to the actual number, where a higher number indicates greater segregation (Eriksson and Ward 2019).<sup>28</sup> Mexican immigrants started with a segregation level of 0.44 near

---

<sup>27</sup> Controlling for geography (i.e., county of residence or urban status) or literacy does not narrow the English acquisition gap across groups (see Online Appendix Table A3).

<sup>28</sup> The measure is  $\eta_c = 1 - \frac{USborn_c}{E(USborn_c)}$  where  $USborn_c$  is the actual number of US-born next-door households in the county, and  $E(USborn_c)$  is the expected number under random assignment (Eriksson and Ward 2019). For a few examples, an index of one indicates complete segregation, zero indicates random assignment within the county, and a value of 0.75 indicates that the actual number of US-born neighbors is 25 percent of the expected number under random assignment.

arrival and ended with a level of 0.32 after twenty years. This decrease shows that spatial assimilation occurred (as opposed to the lack of earnings score growth). On the other hand, Italian immigrants started at a higher level of 0.52, which indicates that newly arriving Italian immigrants were more segregated than Mexican immigrants. Like the rate of English acquisition, the rate of spatial assimilation was faster for Italians as both groups ended with similar segregation levels after 20 years of stay.<sup>29</sup> In the next section, we will account for differences in location, English fluency and segregation to help understand the differences in assimilation patterns across Mexicans and Italians.

## V. Explaining the differences in levels and the differences in growth rates

### *Differences in earnings score levels at first observation*

In this section, we turn to the linked data to uncover why Mexicans started and fell behind US-born whites. First, we ask whether the gap at first observation between Mexicans, Italians, and US-born whites, as well as any differences across cohorts of Mexican arrivals, can be explained by observable characteristics. Then we will address why growth throughout the lifecycle was slower for Mexican immigrants than for other groups, and whether observable characteristics contribute to differences in growth across arrival cohorts.

To explain the economic gaps at first observation, we estimate the following specification:

$$\ln(y_i) = \gamma_0 + \gamma_1 \text{Mexico}_i + \gamma_2 \text{Italy}_i + \Pi'X_i + \varepsilon_i \quad (2)$$

Where we regress the first census's log earnings score on a dummy variable for Mexico and Italy. We add control variables for human capital (literacy, ability to speak English), location (urban, log

---

<sup>29</sup> Similarly, controlling for geography or human capital does not narrow the difference in the rate of spatial assimilation (see Online Appendix Table A4).

county population, border state/state of residence), networks (log number of Mexican immigrants in the county for Mexicans, and Italian immigrants for Italians), residential segregation (segregation from US-born) and census year.

A unique addition to Equation (2) is that we include a variable for whether the county had a documented lynching of someone of Mexican descent in the decade of arrival.<sup>30</sup> 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. Nevertheless, our view is that this variable proxies for severe ethnic animus against Mexicans and other discriminatory factors.

In column 1 of Table 3, we show the unconditional gaps across groups: Mexicans started 57 log points behind, while Italians had similar earnings score as US-born whites. Human capital differences do explain some of these gaps. Controlling for literacy and the ability to speak English, the US-Mexico gap fell from 57 to 33 log points, a drop of 42 percent. However, Mexican immigrants were still about 28 percent behind US-born whites. Interestingly, controlling for human capital predicts that Italian immigrants out-earned US-born whites by 13 log points. However, the relative increase of Mexican and Italian gaps leaves the Mexico-Italian gap at the same place; that is, the Mexican-Italian gap is not explained by observable human capital differences across groups.

While human capital does not help to explain why Mexicans and Italians had different outcomes near arrival, location does. In column 3 of Table 3, we show that after controlling for

---

<sup>30</sup> The lynching variable is zero for Italians and US-born whites.

various factors like being in a border state, residential segregation from the US-born, and the population of the county, the Mexico-Italy gap was only 13 log points.<sup>31</sup> This 13-point gap contrasts with a 55-point gap without the controls. Two of the most important controls are urban status and county population. The rural-urban divide is key for understanding the gaps across Italians and Mexicans, since Mexico was one of the most rural sources while Italy was one of the most urban (Eriksson and Ward forthcoming). If one accounts for the general urban premium in earning scores, then the groups' outcomes converged. Not all location variables mattered though. In Table 3, we also show that being in a border state and having a higher number of immigrants in the county are not correlated with initial outcomes. This result suggests that the border/non-border gap in Mexican outcomes is not necessarily due to network effects but could be due to a rural/urban divide. Mexican lynchings are also not correlated with the initial earnings score, but we will later show that this result does not hold for earnings score growth. Based on the full specification that includes county fixed effects in Column IV, the Mexico-Italy gap was 11 log points and the US-Mexico gap was 43 log points.

While we estimated before that 1900-1929 Mexican arrival cohorts had similar gaps near arrival, these estimates were unconditional on observable characteristics. Table 4 shows that the conditional gaps did vary.<sup>32</sup> For conciseness, we show that arrival gaps in Table 4 after splitting cohorts into the decade of arrival (see Table A5 for 5-year arrival cohorts). After controlling for observable characteristics, the Mexico-US gap increased in magnitude across arrival cohorts, from negative 37 log points for the 1900-1909 cohort, to negative 48 log points for the 1910-1919 cohort, to 54 log points for the 1920-1929 cohort. It is unclear why the conditional gap grew across

---

<sup>31</sup> That is, the difference between the Mexico and Italian dummy variables.

<sup>32</sup> We omit Italians from the regression.



arrival cohorts. It may be that as opportunities increased in the early 20<sup>th</sup> 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.

### *Differences in earnings score growth*

What happened in the following decades after these initial gaps? Now we turn to analyze the conditional differences in growth across US-born whites, Mexicans, and Italians:

$$(\ln(y_{i,t}) - \ln(y_{i,t-1})) = \gamma_0 + \gamma_1 \text{Mexico}_i + \gamma_2 \text{Italy}_i + \Pi' X_{i,t-1} + \varepsilon_i \quad (3)$$

We control for the same human capital, geography, and network variables as in Equation (2), but include a key variable: the initial earnings score ( $\ln(y_{i,t-1})$ ). Collins and Zimran (2019) show that differences in where immigrants and US-born whites start off have a large impact on their growth rate in the following years. Another difference is that the lynching variable is now for a lynching that occurred in between observations, rather than in the decade of arrival.

In the first column of Table 5, we show that Mexicans fell behind US-born whites by 12 log points. Italian immigrants fell behind by 11 log points, which is not statistically different from the Mexican growth rate. However, the gap between Mexicans and other groups diverges after adding the control for the initial log earning score. Conditional on the starting point, Mexican immigrants fell behind US-born whites by 37 log points (31 percent). Recall that Mexicans were concentrated at the bottom of the distribution. This result implies that decadal growth was high for the US-born who started low, but not high for Mexicans who started in a similarly low place. In other words, upward mobility existed for low-skilled, US-born whites, but not for low-skilled

Mexicans. Controlling for the initial earnings score also creates a wedge between Mexicans and Italians. Surprisingly, the Italian variable remains constant at negative 11 log points, indicating that their growth rate relative to US-born whites did not depend on where they started. In Figure 5, we show these gaps from the first to the second census when using percentile ranks, which confirms that there are intercept differences between Mexicans and the other groups. These differences suggest that there were barriers to Mexican progress that did not allow for them to escape low earnings score occupations.

Other observables, such as human capital and geography, do not fully explain the difference in growth rates across groups. When adding controls for the initial literacy level and ability to speak English, Mexican immigrants still fell behind US-born whites by 31 log points. When further adding controls for living in an urban area, living in a border state or the degree of residential segregation, Column IV shows that Mexican immigrants fell behind by 25 log points. Interestingly, the border state control variable does not explain the growth rate, which implies that the reason why we found slower growth in border states is because of differences in the initial earning score, human capital, or urban/rural status. A novel result in Table 5 is that segregation is negatively associated with earnings score growth, where going from random allocation of immigrants and natives in a county to full segregation is associated with a fall of 19 log points (Column IV). Yet a predicted gap remains across US-born and Mexicans for Mexicans who were in a completely integrated county. Note that this specification controls for the Mexican population in the county, which is also negatively associated with earning score growth.

The residual gap in Column IV suggests that unobservable variables that are correlated with Mexicans explain why they fell behind US-born whites and Italians. One reason for this residual is that structural barriers limited the economic progress of Mexicans. Structural barriers

would also account for the minimal difference across cohorts in the assimilation rate despite the shocks to selection in the short run from economic downturns or violence in Mexico. In the last row of Table 5, we attempt to capture the influence of structural barriers in the form of Mexican lynching. After additionally including county fixed effects, Mexicans in counties with a recorded lynching experienced a 16 log-point slower growth rate. This is a large effect. Further, the negative association between lynching and earning score *growth* contrasts with the result in Table 3 that lynching was *uncorrelated* with the initial earnings score *level*. Our interpretation is that violent acts halted the economic progress of Mexican immigrants, which would help to explain why Mexican immigrants who are observably similar to other groups still fell behind.

Finally, in Table 6, we estimate how earnings score growth varied across Mexican arrival cohorts. The growth results are similar to the level results: while the unconditional differences in growth rates are small across arrival cohorts, the conditional differences are larger (Columns II through V). In particular, the 1920-1929 cohort, which arrived just before the Great Depression, is estimated to have the slowest growth rate relative to US-born whites. These results are suggestive that the Great Depression had a large and negative impact on Mexican immigrants relative to US-born whites, given county fixed effects, and having the same observable human capital. A stronger impact of a negative shock on migrants is consistent with evidence from Canada that immigrants were hit harder by the Great Depression than natives (Inwood et al. 2016). However, this interpretation is only speculative. An alternative explanation, for instance, is that the selection of Mexican immigrants was less positive in the 1920s.

In Online Appendix Table A6, we further narrow our focus on specific years of arrival to test whether cohorts most likely to reflect refugees had different assimilation rates. For example, fighting during the Mexican Revolution was most intense between 1914 and 1915, which are the

exact years when the composition of arrivals shifted toward professional jobs, going from about 10 percent of arrivals in 1912 to almost 40 percent in 1915 (Kosack and Ward 2014). However, this group did not have a different assimilation rate or different earnings score level at first arrival. The cohort which arrived during intense periods of the Cristero War (1927-1928) also did not have a difference in assimilation rates. Besides these refugee cohorts, we do not find a difference in assimilation rates for those who arrived during the Panic of 1907 or the year afterward (1907-1908), which has been shown to influence selection in the short run (Escamilla-Guerrero and López-Alonso 2020). In our view, these results provide additional evidence that shifts in the composition of arrivals did not necessarily translate into different assimilation outcomes, perhaps because Mexicans were not able to escape a concentrated set of occupations.

An important caveat for our entire set of results is that earnings score estimates may not accurately capture earnings, especially when moving further back in time. In Figure 6, we show estimates of mobility based on broad occupational groups (white collar, semi-skilled, unskilled and farmer). For instance, Panel A shows that of the white-collar workers at first observation, 70 percent of US-born whites, 43 percent of Italians, and 36 percent of Mexicans ended up in a white-collar job ten years later. These gaps across groups are consistent with the earnings score results that Mexicans fell behind conditional on the first occupation. Further, the results suggest that most Mexicans fell off the occupational ladder: 64 percent of Mexicans who were in a white collar job at first observation ended up in a lower-ranked occupation 10 years later. In panels B, C and D, we show where the white-collar Mexican workers ended up: 8 percent in semi-skilled jobs, 52 percent in unskilled jobs and 4 percent in farming. The key result here for Mexican migrants can be seen in panel C where we show that no matter in which job category a Mexican started, the majority ended up in an unskilled job ten years later.

## VI. Conclusions

In this paper, we show that for the first wave of Mexican mass migration, immigrant outcomes were remarkably consistent across cohorts. Mexican arrivals between 1900 and 1929 all held lower-paid jobs than both US-born whites and Italian immigrants at arrival. After arrival, both Mexican and Italian immigrants fell behind US-born whites by about 12 log points. If one controls for characteristics at first observation, such as initial earnings score, literacy and county of residence, then Mexican immigrants fell behind both US-born whites and Italian immigrants. The results that assimilation rates were similar across cohorts and that Mexicans fell behind observably similar Italians and US-born whites are consistent with Mexican-specific structural barriers limiting the opportunities for upward mobility. While evidence for structural barriers is difficult to pinpoint statistically, we do show that Mexicans who lived in counties where a documented Mexican lynching occurred – the most severe form of repression – fell behind US-born whites by an additional 16 log points.

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 validity to this claim. Both groups had similar assimilation profiles, though differences do emerge when accounting for observable differences in geography, human capital, or starting jobs. However, our view is that simple comparisons between immigrant groups could mask far more complex realities, including specific economic contexts and barriers to progress that limit upward mobility. Therefore, one should not be too quick to assume that holding the economy constant would lead to similar assimilation rates across groups with distinct experiences in the United States.

## Bibliography

- Abramitzky, Ran, Philipp Ager, Leah Boustan, Elicor 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.
- 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, Katherine Eriksson, James J. Feigenbaum, and Santiago Pérez. Automated linking of historical data. *The Journal of Economic Literature*, forthcoming
- 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.
- Alanís, Fernando Saúl. "No cuenten conmigo: La política de repatriación del gobierno mexicano y sus nacionales en Estados Unidos, 1910—1928." *Mexican Studies* 19, no. 2 (2003): 401-431.
- 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.
- 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.
- 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, 2019.
- 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.
- Cook, Lisa D. "Violence and economic activity: evidence from African American patents, 1870–1940." *Journal of Economic Growth* 19, no. 2 (2014): 221-257.
- 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.
- 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.
- 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.
- 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, and Zachary Ward. "Immigrants and Cities during the Age of Mass Migration" *Regional Science and Urban Economics*, forthcoming.
- 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.
- 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.
- 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.
- 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.

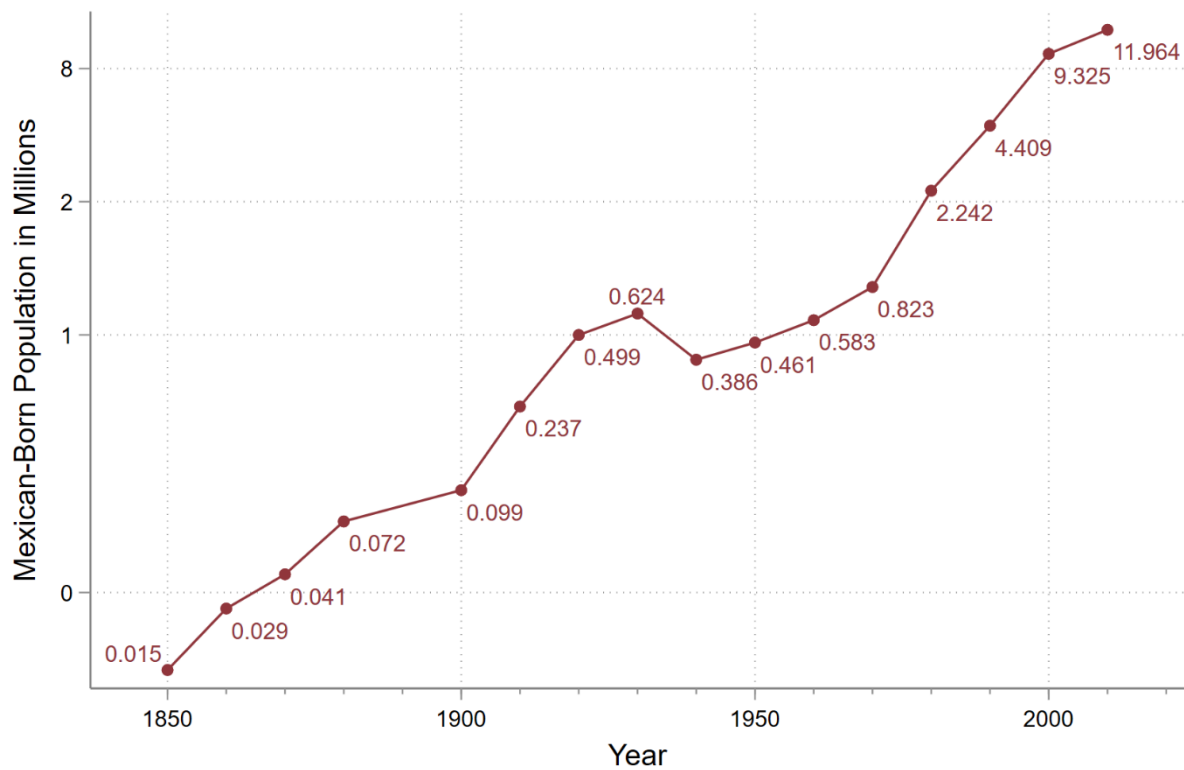


- Hacker, J. David. "New estimates of census coverage in the United States, 1850–1930." *Social Science History* 37.1 (2013): 71-101.
- Hall, Linda B. "El Refugio: Migración Mexicana a los Estados Unidos: 1910-1920." *Boletín del Instituto de Investigaciones Históricas* 8 (1982): p. 23-44.
- Hatton, 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, Hupert Michael (eds) *Handbook of Cliometrics*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-40458-0\\_41-1](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. "Reverse assimilation? Immigrants in the Canadian labour market during the Great Depression." *European Review of Economic History* 20, no. 3 (2016): 299-321.
- 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.
- 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.
- 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.
- 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.

- 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.
- Lee, Jongkwan, Giovanni Peri, and Vasil Yassenov. *The Employment Effects of Mexican Repatriations: Evidence from the 1930's*. No. w23885. National Bureau of Economic Research, 2017.
- 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.
- 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.
- 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* (2019): 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.
- 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.

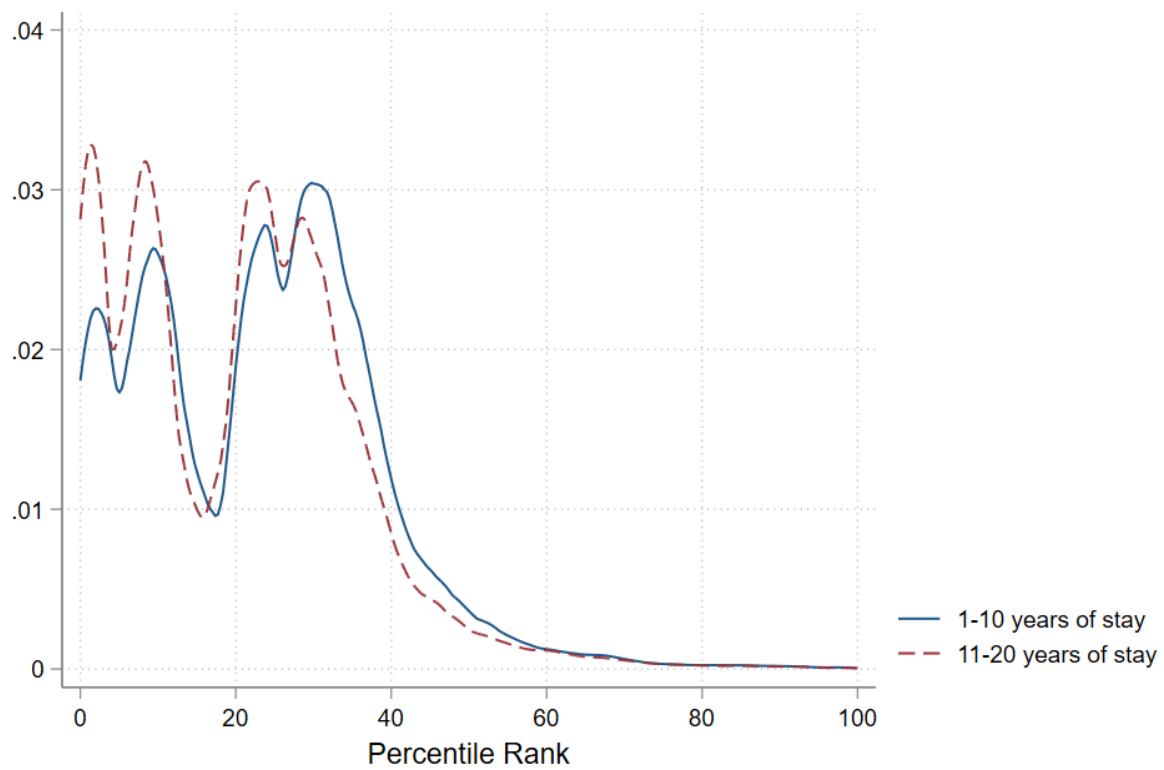
- Verduzco, Gustavo. "La migración mexicana a Estados Unidos: recuento de un proceso histórico." *Estudios Sociológicos* 13, no. 39 (1995): 573-594.
- Ward, Zachary. "The low return to English fluency during the Age of Mass Migration." *European Review of Economic History* 24.2 (2020): 219-242.
- Ward, 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.
- Williams, Jhacova. *Historical lynchings and contemporary voting behavior of blacks*. Technical report, Louisiana State University, 2017.
- 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.** Stock of Mexican-Born Population in United States, 1850 to 2010



Notes: Data is from the 1850 to 2010 US Census samples and the 2010 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.

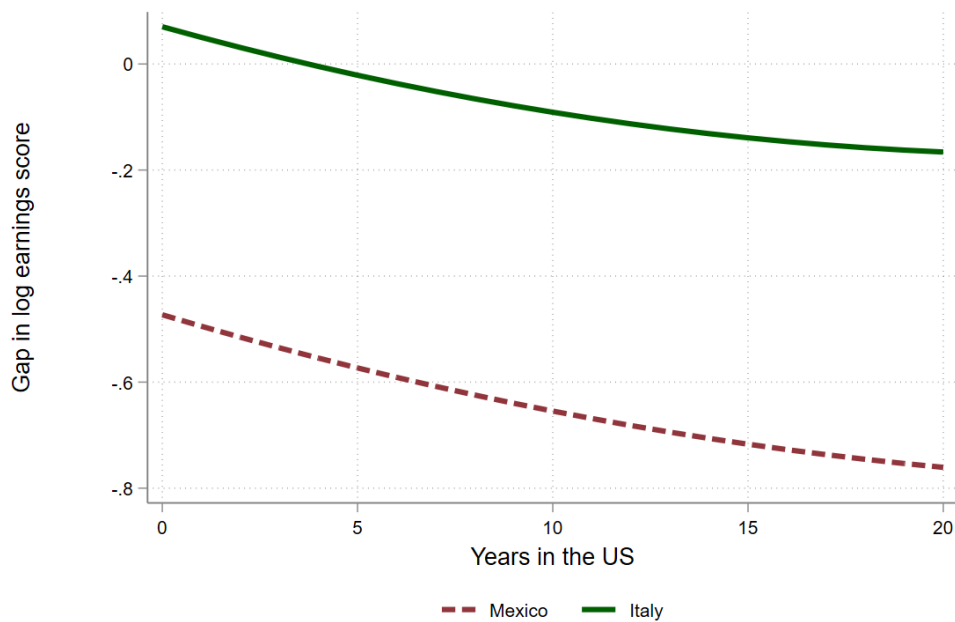
**Figure 2.** Mexican-born location in the earnings score distribution in the early 20<sup>th</sup> century



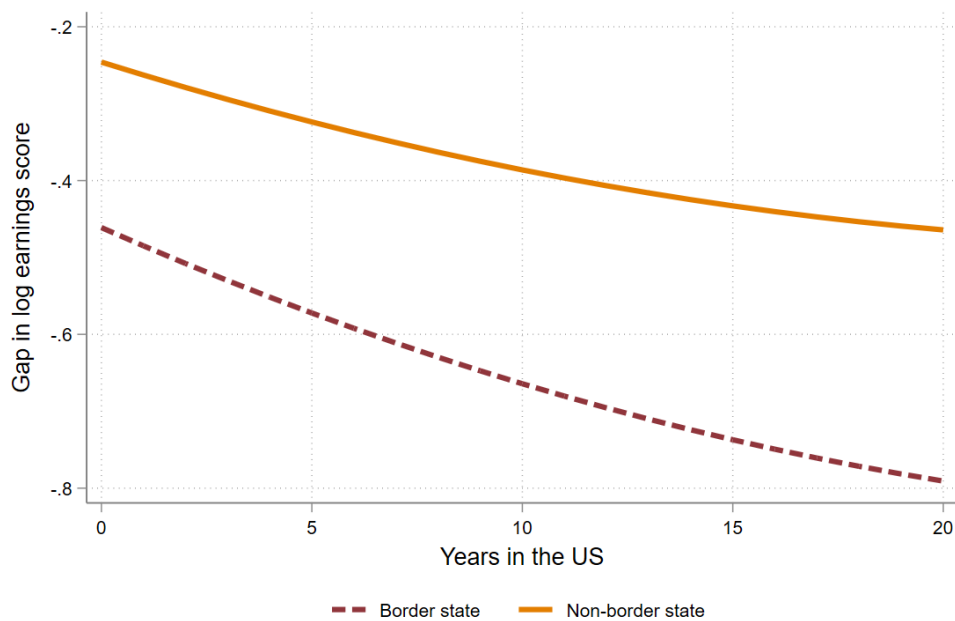
Notes: Data are from the linked censuses between 1910-1920, 1920-1930 and 1930-1940. Percentile ranks are calculated within the data by year and birth cohort. Ties are given equal percentile rank.

**Figure 3.** Mexican immigrants diverged from US-born whites after arrival but so did Italians

Panel A. Log earnings score



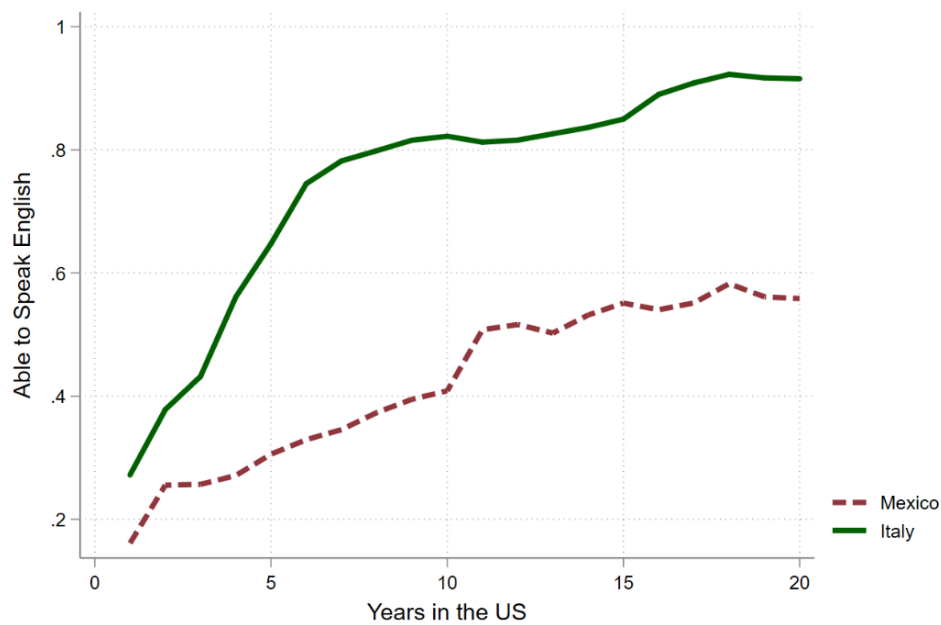
Panel B. Mexican assimilation in border versus non-border states



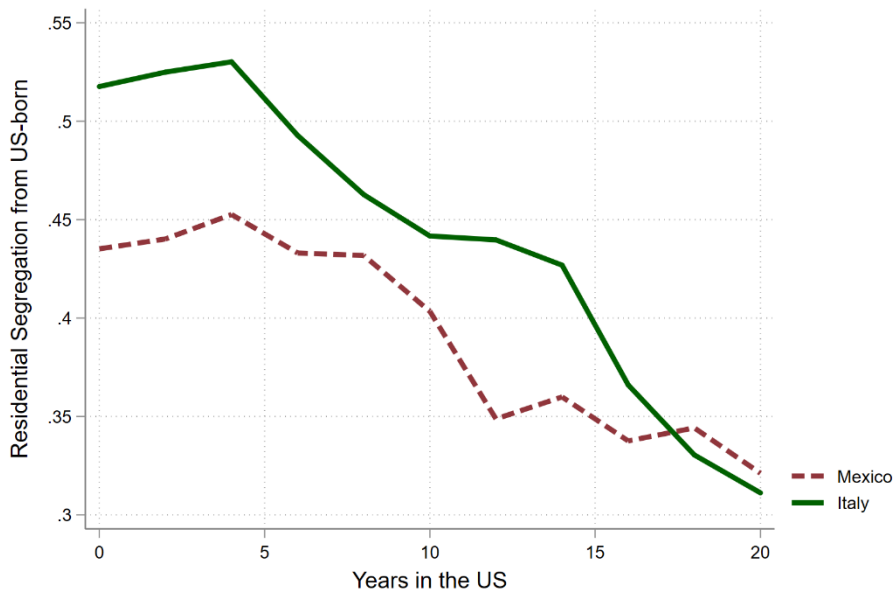
Notes: Data are from 1910-1920, 1920-1930, and 1930-1940 linked samples. Panel A plots the predicted assimilation profile for the Mexico-born and Italy-born 1910-1919 arrival cohort. Panel B splits the Mexican sample into those living in a border state or non-border state.

**Figure 4.** English acquisition and spatial assimilation were slower for Mexican immigrants

*Panel A. Ability to speak English*

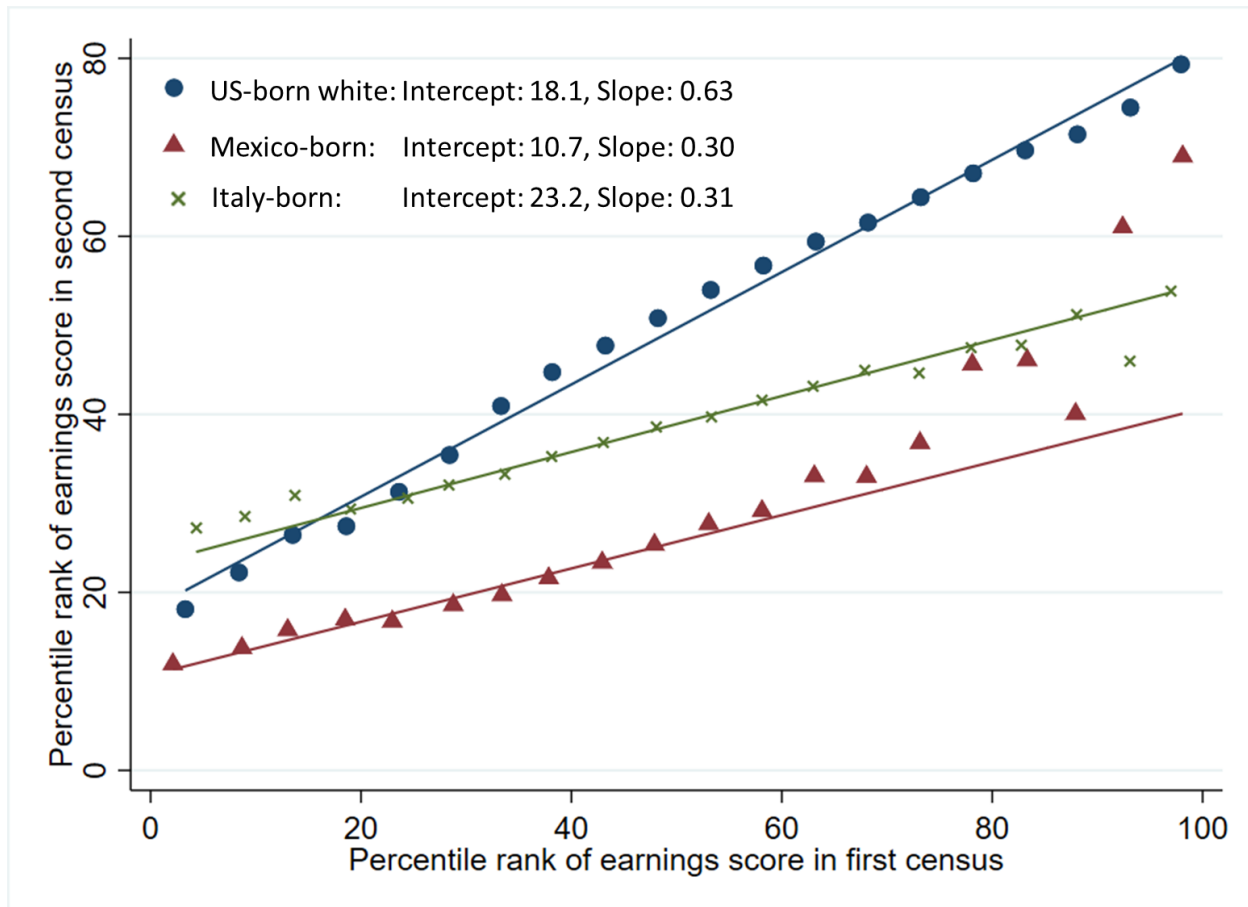


*Panel B. Spatial Assimilation*



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.

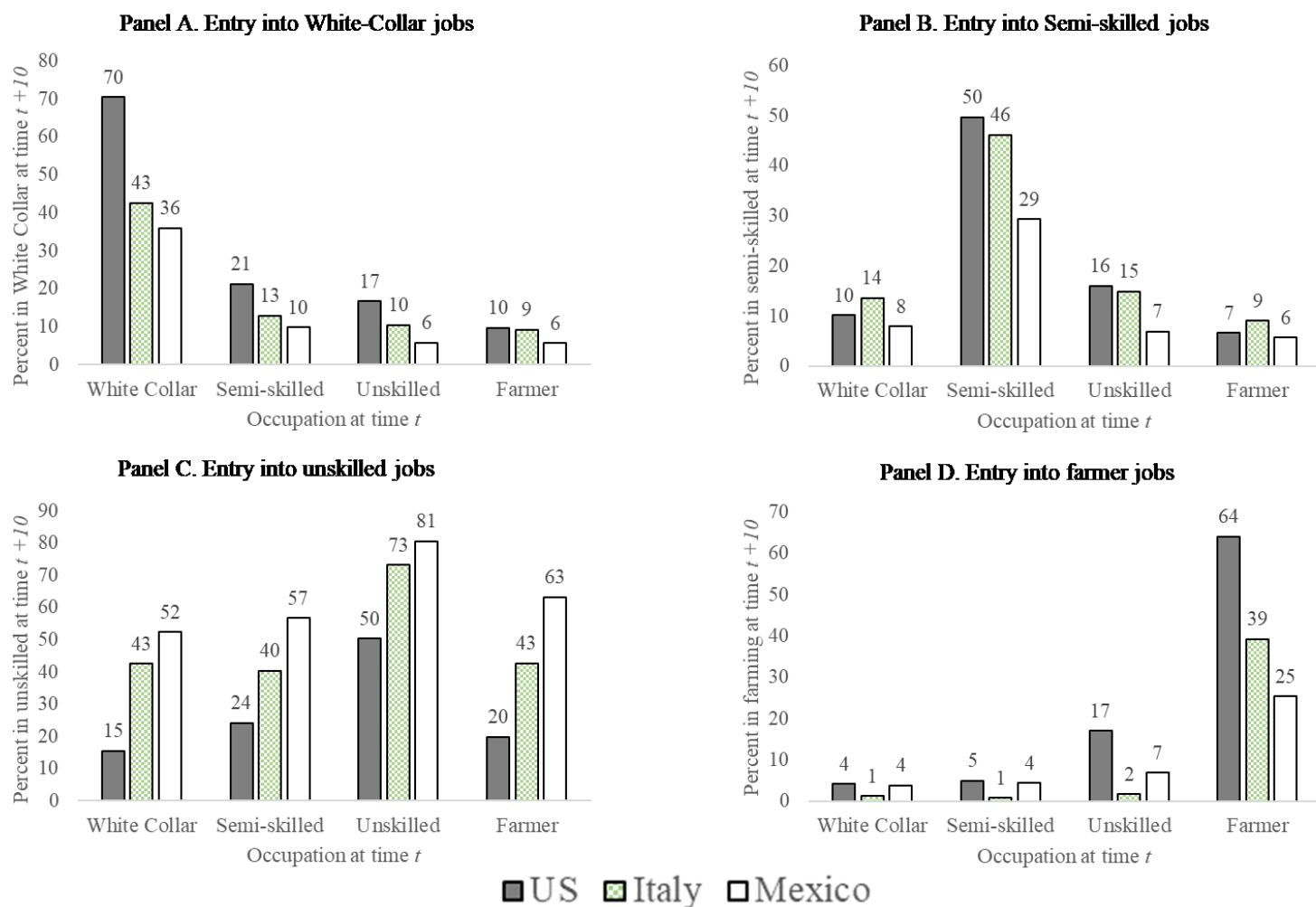
**Figure 5.** Rank-rank associations across groups



Note: Data are from the 1910-1920, 1920-1930, 1930-1940 linked censuses. The figure plots the rank-rank association in percentile rank of earning score by Mexican and US-born white. Note that there are few Mexico-born above the 60<sup>th</sup> percentile (see Figure 2).



**Figure 6.** 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, 36 percent ended in white-collar jobs, 8 percent in semi-skilled jobs, 52 percent in unskilled jobs and 4 percent in farmer jobs.

**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
Log earning score	6.87 (0.66)	7.01 (0.58)	0.13 (0.00)	6.33 (0.48)	6.33 (0.48)	-0.00 (0.00)	6.84 (0.20)	6.88 (0.21)	0.03 (0.00)
Percentile rank	50.90 (28.76)	50.24 (28.83)	-0.66 (0.01)	23.02 (15.03)	17.60 (14.13)	-5.42 (0.13)	41.92 (11.14)	36.37 (10.76)	-5.55 (0.07)
Literate	0.98 (0.14)	0.98 (0.13)	0.00 (0.00)	0.63 (0.48)	0.69 (0.46)	0.05 (0.01)	0.76 (0.43)	0.76 (0.42)	0.00 (0.00)
Can speak English	0.98 (0.12)	0.99 (0.11)	0.00 (0.00)	0.30 (0.46)	0.54 (0.50)	0.24 (0.01)	0.68 (0.47)	0.88 (0.32)	0.20 (0.00)
Age	27.80 (6.85)	37.77 (6.90)	9.97 (0.00)	27.19 (6.61)	37.18 (6.83)	9.99 (0.06)	28.05 (6.43)	38.15 (6.56)	10.10 (0.05)
Age at arrival	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	21.61 (6.93)	21.60 (7.12)	-0.01 (0.07)	21.28 (6.34)	21.38 (6.45)	0.10 (0.05)
Border state	0.09 (0.28)	0.10 (0.30)	0.01 (0.00)	0.80 (0.40)	0.83 (0.38)	0.03 (0.00)	0.07 (0.25)	0.07 (0.26)	0.00 (0.00)
White Collar	0.25 (0.43)	0.30 (0.46)	0.05 (0.00)	0.06 (0.24)	0.08 (0.27)	0.02 (0.00)	0.10 (0.30)	0.14 (0.35)	0.04 (0.00)
Farmer	0.16 (0.36)	0.19 (0.39)	0.03 (0.00)	0.04 (0.19)	0.07 (0.26)	0.03 (0.00)	0.01 (0.11)	0.02 (0.14)	0.01 (0.00)
Skilled	0.17 (0.38)	0.19 (0.39)	0.02 (0.00)	0.07 (0.25)	0.08 (0.28)	0.01 (0.00)	0.20 (0.40)	0.21 (0.41)	0.01 (0.00)
Unskilled	0.42 (0.49)	0.32 (0.47)	-0.10 (0.00)	0.83 (0.38)	0.76 (0.42)	-0.06 (0.00)	0.69 (0.46)	0.63 (0.48)	-0.06 (0.00)
Lives in Urban area	0.50 (0.50)	0.53 (0.50)	0.03 (0.00)	0.51 (0.50)	0.57 (0.50)	0.05 (0.00)	0.83 (0.37)	0.86 (0.35)	0.03 (0.00)
Observations	13,516,131	13,516,131		28,060	28,060		60,043	60,043	

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.

**Table 2.** Across and within-cohort earnings score differences for Mexican immigrants

	Census Year				Within- Cohort Growth
	1910	1920	1930	1940	
<i>Panel A. Mexico-born</i>					
1900-1904	-0.589 (0.014)	-0.743 (0.013)			-0.154 (0.020)
1905-1909	-0.506 (0.009)	-0.702 (0.009)			-0.196 (0.013)
1910-1914		-0.607 (0.007)	-0.746 (0.007)		-0.140 (0.010)
1915-1919		-0.534 (0.005)	-0.699 (0.005)		-0.165 (0.007)
1920-1924			-0.570 (0.006)	-0.683 (0.006)	-0.113 (0.009)
1925-1929			-0.471 (0.007)	-0.666 (0.008)	-0.195 (0.011)
<i>Panel B. Italy-Born</i>					
1900-1904	0.045 (0.002)	-0.101 (0.002)			-0.146 (0.004)
1905-1909	0.053 (0.002)	-0.120 (0.002)			-0.172 (0.003)
1910-1914		-0.038 (0.002)	-0.179 (0.001)		-0.141 (0.002)
1915-1919		0.019 (0.005)	-0.155 (0.004)		-0.174 (0.008)
1920-1924			-0.097 (0.002)	-0.130 (0.002)	-0.034 (0.003)
1925-1929			-0.036 (0.005)	-0.102 (0.005)	-0.065 (0.007)

Notes: Data are from linked censuses (1910-1920; 1920-1930; 1930-1940). The tables report the age- and year-adjusted log earnings score difference with US-born whites. The last column calculates the ten-year within-cohort growth rate.

**Table 3.** Differences in initial gaps between Mexicans, Italians and US-born whites

	I	II	III	IV
Mexico	-0.572 (0.038)	-0.329 (0.032)	-0.367 (0.050)	-0.432 (0.031)
Italy	-0.020 (0.021)	0.127 (0.020)	-0.242 (0.053)	-0.322 (0.033)
Literate		0.659 (0.015)	0.388 (0.007)	0.179 (0.003)
Can speak English		0.114 (0.012)	0.027 (0.008)	0.026 (0.005)
Urban area			0.545 (0.014)	0.481 (0.011)
Border State			-0.016 (0.030)	
Segregation level			-0.091 (0.111)	-0.175 (0.065)
Log source country population			-0.003 (0.005)	-0.006 (0.002)
Log county population			0.103 (0.007)	0.033 (0.009)
Mexican lynching in county			0.014 (0.021)	0.009 (0.018)
Year fixed effects	Y	Y	Y	Y
County fixed effects	N	N	N	Y
F-test p-value (Mexico = Italy)	0	0	0.009	0
Observations	13,603,890	13,603,890	13,603,890	13,603,890
R-squared	0.015	0.033	0.385	0.516

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 4.** Differences in arrival outcomes across Mexican cohorts

	I	II	III	IV
1900-1909 arrivals	-0.550 (0.055)	-0.157 (0.057)	-0.245 (0.073)	-0.365 (0.050)
1910-1919 arrivals	-0.585 (0.037)	-0.282 (0.031)	-0.420 (0.074)	-0.484 (0.051)
1920-1929 arrivals	-0.539 (0.038)	-0.361 (0.033)	-0.537 (0.065)	-0.538 (0.043)
Literate		0.681 (0.017)	0.406 (0.007)	0.197 (0.003)
Can speak English		0.125 (0.013)	0.032 (0.008)	0.030 (0.005)
Urban area			0.531 (0.014)	0.471 (0.010)
Border State			-0.017 (0.030)	
Segregation level			0.176 (0.171)	0.029 (0.110)
Log source country population			-0.005 (0.005)	-0.007 (0.002)
Log county population			0.107 (0.007)	0.030 (0.008)
Mexican Lynching in county			0.004 (0.020)	-0.008 (0.015)
Year fixed effects	Y	Y	Y	Y
County fixed effects	N	N	N	Y
Observations	13,543,974	13,543,974	13,543,974	13,543,974
R-squared	0.064	0.083	0.436	0.565

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). Standard errors are clustered at the county level.

**Table 5.** Differences in growth rates between Mexicans, Italians and US-born whites

	I	II	III	IV	V
Mexico	-0.123 (0.023)	-0.372 (0.010)	-0.305 (0.011)	-0.253 (0.016)	-0.318 (0.015)
Italy	-0.106 (0.006)	-0.114 (0.005)	-0.072 (0.005)	-0.085 (0.021)	-0.136 (0.017)
<i>Observables from first observation</i>					
Log earnings score		-0.437 (0.005)	-0.442 (0.005)	-0.540 (0.003)	-0.613 (0.002)
Literate			0.192 (0.003)	0.187 (0.003)	0.144 (0.002)
Can speak English			0.032 (0.004)	0.016 (0.004)	0.013 (0.003)
Urban area				0.116 (0.004)	0.133 (0.003)
Border State				0.012 (0.010)	
Segregation level				-0.192 (0.033)	-0.243 (0.027)
Log source country population				-0.007 (0.002)	-0.006 (0.001)
Log county population				0.045 (0.002)	-0.003 (0.005)
Lynching in county b/w time $t-1$ and $t$				-0.127 (0.017)	-0.152 (0.014)
Year fixed effects	Y	Y	Y	Y	Y
County fixed effects	N	N	N	N	Y
F-test p-value (Mexico = Italy)	0.471	0	0	0	0
Observations	13,603,890	13,603,890	13,603,890	13,603,890	13,603,890
R-squared	0.004	0.299	0.301	0.328	0.351

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 6.** Differences in growth rates across Mexican cohorts

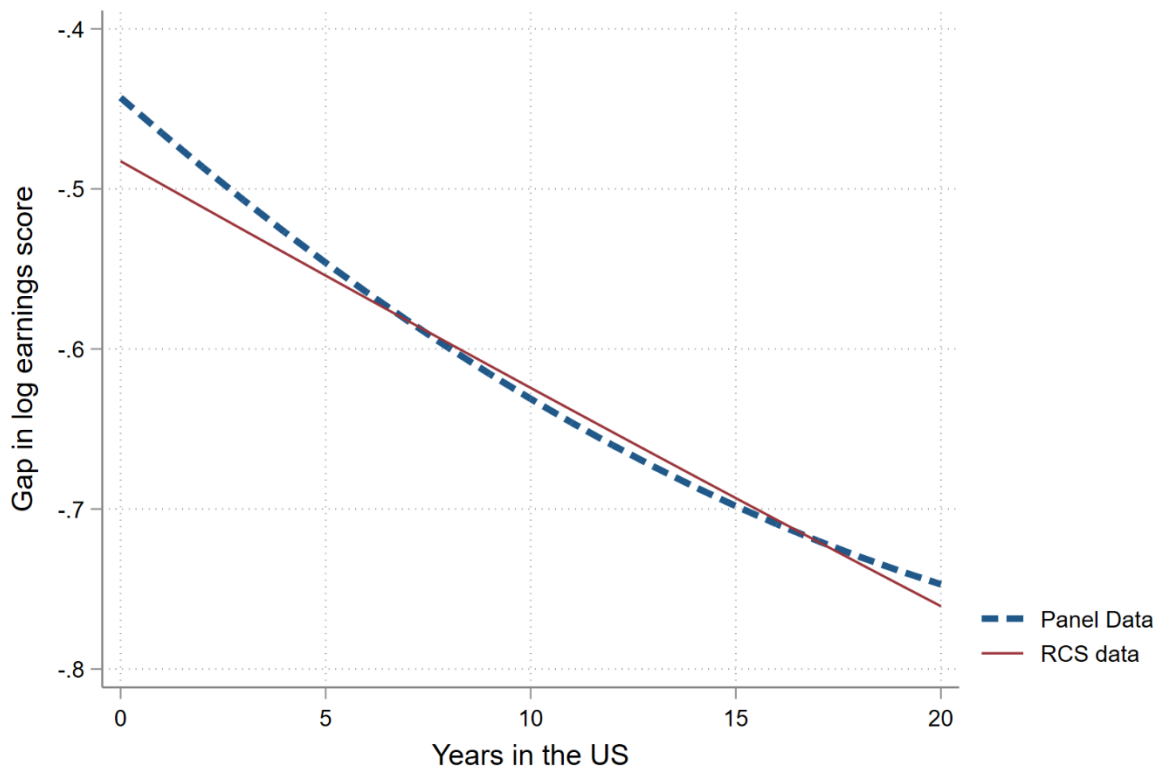
	I	II	III	IV	V
1900-1909 arrivals	-0.146 (0.036)	-0.388 (0.016)	-0.276 (0.015)	-0.199 (0.026)	-0.283 (0.024)
1910-1919 arrivals	-0.099 (0.024)	-0.369 (0.012)	-0.284 (0.014)	-0.230 (0.028)	-0.304 (0.026)
1920-1929 arrivals	-0.136 (0.023)	-0.371 (0.011)	-0.322 (0.012)	-0.296 (0.025)	-0.345 (0.025)
<i>Observables from first observation</i>					
Log earnings score		-0.437 (0.005)	-0.442 (0.005)	-0.540 (0.003)	-0.613 (0.002)
Literate			0.198 (0.003)	0.192 (0.003)	0.148 (0.002)
Can speak English			0.037 (0.005)	0.021 (0.004)	0.017 (0.003)
Urban area				0.116 (0.004)	0.133 (0.003)
Border State				0.012 (0.010)	
Segregation level				-0.172 (0.058)	-0.228 (0.051)
Log Mexican population				-0.007 (0.002)	-0.006 (0.001)
Log county population				0.045 (0.002)	-0.003 (0.005)
Lynching in county b/w time $t-1$ and $t$				-0.132 (0.017)	-0.155 (0.014)
Year fixed effects	Y	Y	Y	Y	Y
County fixed effects	N	N	N	N	Y
Observations	13,543,974	13,543,974	13,543,974	13,543,974	13,543,974
R-squared	0.002	0.306	0.309	0.336	0.359

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.

# Online Appendix

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

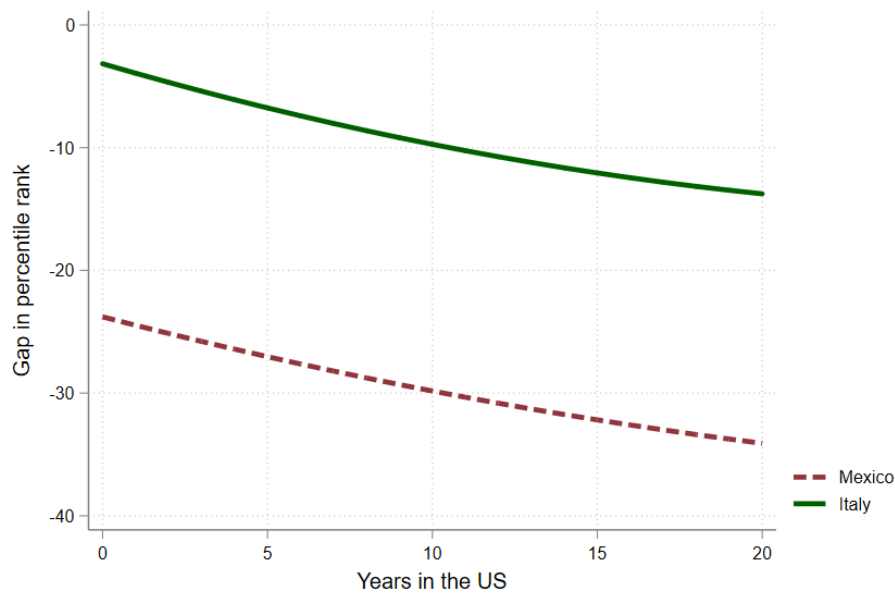
**Figure A1.** The assimilation profile for panel and repeated cross-sectional data



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 figure shows the predicted assimilation rate for the 1910-1919 arrival cohort. The figure suggests negative selection into return migration because the assimilation rate is faster (less negative) for the RCS data than panel data.

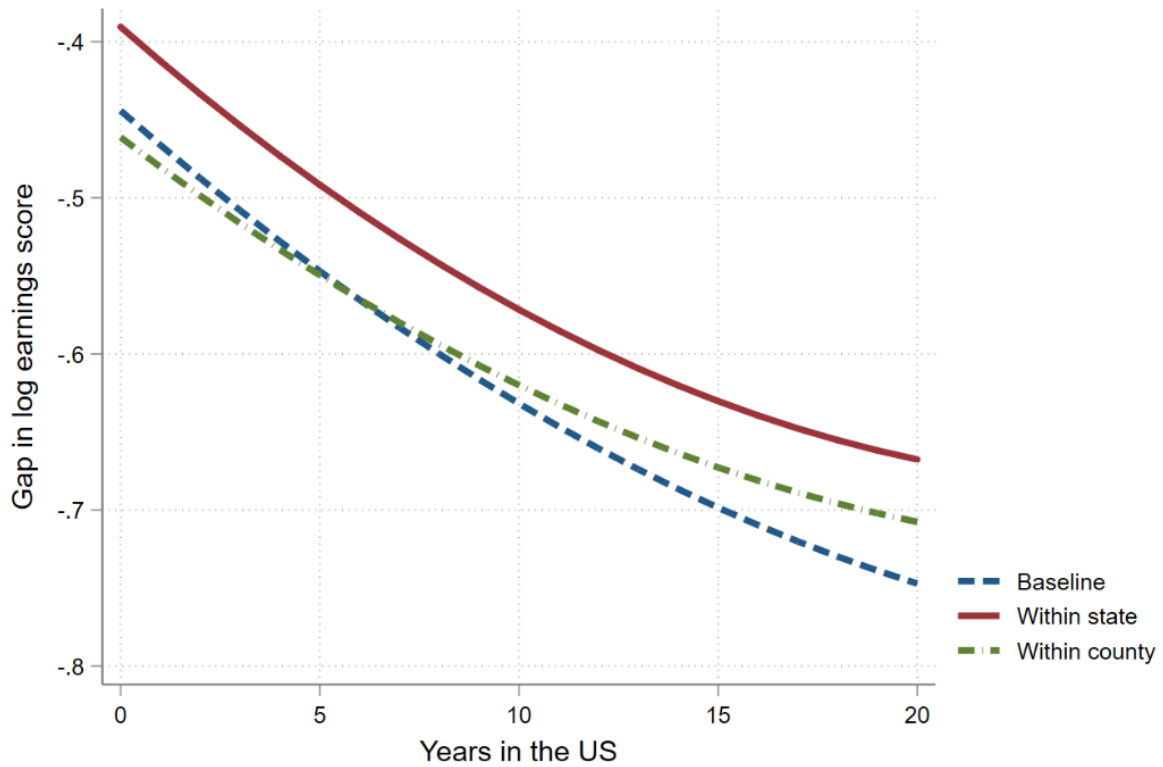


**Figure A2.** Negative assimilation holds when using percentile ranks.



Notes: Data are from 1910-1920, 1920-1930, and 1930-1940 linked samples. This figure uses the percentile rank as the dependent variable rather than log earnings score.

**Figure A3.** Mexican assimilation rates are negative when controlling for geography



Notes: Data are from linked samples for 1910-1920, 1920-1930, and 1930-1940. The figure shows the predicted assimilation rate for the 1910-1919 arrival cohort. The baseline regression only includes age- and year-effects. The within-state adds state fixed effects, and the within-county adds county fixed effects.

**Table A1.** Earnings scores for US-born white, Mexican and Italian immigrants

Occupational Group	US-born white	Mexico	Italy
<i>Panel A. Earnings Score</i>			
Professional, Technical	2,054	1,247	1,346
Farmers	537	329	658
Managers, Officials, and Proprietors	2,200	1,072	1,330
Clerical and Kindred	1,661	1,095	1,223
Sales Workers	1,730	852	1,133
Craftsmen	1,440	871	1,084
Operatives	1,187	835	973
Service (low-skilled)	1,283	644	958
Farm Laborers	406	357	591
Laborers	759	645	841
<i>Panel B. Percentile Rank</i>			
Professional, Technical	86	50	57
Farmers	20	3	27
Managers, Officials, and Proprietors	92	40	56
Clerical and Kindred	76	45	54
Sales Workers	78	34	47
Craftsmen	63	34	45
Operatives	51	32	41
Service (low-skilled)	54	22	39
Farm Laborers	16	7	27
Laborers	32	23	37
<i>Panel C. Percent of sample in occupation</i>			
Professional, Technical	5	1	2
Farmers	18	6	2
Managers, Officials, and Proprietors	8	2	6
Clerical and Kindred	7	1	1
Sales Workers	7	2	3
Craftsmen	17	8	20
Operatives	15	12	24
Service (low-skilled)	3	4	8
Farm Laborers	10	25	2
Laborers	10	39	32

Notes: Data are from the 1910-1920, 1920-1930, 1930-1940 linked samples. Earnings scores are based on information from occupation, country of birth and region. Scores are calculated at the 3-digit occ1950 level, but the table shows the averages within the data at the 1-digit level.

**Table A2.** Inferred selection into return migration.

Cohort of Arrival	Within-cohort growth rate			Implied selection of return migrants
	Panel	RCS	Difference	
1900-1904	-0.142	-0.106	-0.036	-0.082
1905-1909	-0.184	-0.169	-0.015	-0.034
1910-1914	-0.151	-0.102	-0.049	-0.111
1915-1919	-0.177	-0.096	-0.081	-0.184
1920-1924	-0.132			
1925-1929	-0.214			

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.** Determinants of the change in English fluency from census to census

	I	II	III	IV	V
Mexico	0.023 (0.005)	-0.313 (0.004)	-0.305 (0.004)	-0.286 (0.007)	-0.266 (0.009)
<i>Observables from t-1:</i>					
Can speak English		-0.893 (0.004)	-0.912 (0.004)	-0.918 (0.004)	-0.923 (0.004)
Literate			0.060 (0.004)	0.061 (0.004)	0.061 (0.004)
Urban			0.020 (0.004)	0.022 (0.004)	0.023 (0.005)
Segregation level				-0.018 (0.012)	-0.010 (0.021)
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	60,368	60,368	60,368	60,368	60,368
R-squared	0.038	0.498	0.500	0.501	0.515

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  $t$ .

**Table A4.** Determinants of the change in segregation level from census to census

	I	II	III	IV	V
Mexico	0.064 (0.001)	0.035 (0.001)	0.038 (0.001)	0.062 (0.002)	0.059 (0.002)
Observable from $t-1$ :					
Literate			-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Segregation level		-0.745 (0.003)	-0.749 (0.003)	-0.770 (0.003)	-0.790 (0.004)
Urban			0.009 (0.001)	0.006 (0.001)	0.003 (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	87,372	87,372	87,372	87,372	87,372
R-squared	0.111	0.555	0.555	0.561	0.585

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 A5.** Earnings score at first observation across 5-year cohorts and specific arrival cohorts

	I	II	III	IV
1900-1904 arrivals	-0.590 (0.058)	-0.137 (0.058)	-0.232 (0.104)	-0.336 (0.073)
1905-1909 arrivals	-0.530 (0.060)	-0.061 (0.061)	-0.148 (0.088)	-0.290 (0.061)
1910-1914 arrivals	-0.619 (0.044)	-0.258 (0.036)	-0.366 (0.107)	-0.424 (0.073)
1915-1919 arrivals	-0.585 (0.044)	-0.202 (0.033)	-0.358 (0.095)	-0.433 (0.067)
1920-1924 arrivals	-0.570 (0.050)	-0.314 (0.046)	-0.482 (0.095)	-0.480 (0.065)
1925-1929 arrivals	-0.519 (0.044)	-0.266 (0.038)	-0.463 (0.087)	-0.467 (0.060)
Cristero arrivals (1927-1928)	-0.003 (0.016)	-0.012 (0.018)	0.018 (0.017)	0.012 (0.013)
Revolution arrivals (1914-1915)	-0.007 (0.013)	-0.005 (0.015)	-0.006 (0.019)	0.000 (0.012)
Panic of 1907 arrivals (1907-1908)	-0.046 (0.022)	-0.059 (0.022)	-0.038 (0.022)	-0.015 (0.018)
Literate		0.682 (0.017)	0.406 (0.007)	0.197 (0.003)
Can speak English		0.127 (0.013)	0.033 (0.008)	0.031 (0.005)
Urban area			0.531 (0.014)	0.471 (0.010)
Border State			-0.017 (0.030)	
Segregation level			0.160 (0.166)	0.017 (0.106)
Log source country population			-0.005 (0.005)	-0.007 (0.002)
Log county population			0.107 (0.007)	0.030 (0.008)
Mexican lynching in county			0.003 (0.019)	-0.009 (0.014)
Year fixed effects	Y	Y	Y	Y
County fixed effects	N	N	N	Y
Observations	13,543,974	13,543,974	13,543,974	13,543,974
R-squared	0.059	0.077	0.423	0.548

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

**Table A6.** Earnings score growth, 5 year cohorts and specific arrival cohorts

	I	II	III	IV	V
1900-1904 arrivals	-0.130 (0.036)	-0.390 (0.016)	-0.287 (0.016)	-0.217 (0.028)	-0.302 (0.025)
1905-1909 arrivals	-0.163 (0.038)	-0.384 (0.019)	-0.264 (0.018)	-0.177 (0.027)	-0.262 (0.026)
1910-1914 arrivals	-0.089 (0.028)	-0.375 (0.013)	-0.297 (0.015)	-0.244 (0.029)	-0.317 (0.027)
1915-1919 arrivals	-0.099 (0.027)	-0.360 (0.014)	-0.267 (0.017)	-0.213 (0.030)	-0.286 (0.027)
1920-1924 arrivals	-0.110 (0.026)	-0.359 (0.011)	-0.312 (0.011)	-0.286 (0.024)	-0.338 (0.024)
1925-1929 arrivals	-0.167 (0.026)	-0.376 (0.016)	-0.324 (0.016)	-0.298 (0.030)	-0.345 (0.030)
Cristero arrivals (1927-1928)	-0.012 (0.016)	-0.025 (0.013)	-0.024 (0.013)	-0.016 (0.014)	-0.020 (0.014)
Revolution arrivals (1914-1915)	-0.021 (0.017)	-0.007 (0.012)	-0.012 (0.012)	-0.008 (0.013)	-0.009 (0.012)
Panic of 1907 arrivals (1907-1908)	0.012 (0.024)	-0.005 (0.020)	-0.010 (0.020)	-0.015 (0.020)	-0.012 (0.020)
Log earnings score		-0.436 (0.005)	-0.442 (0.005)	-0.540 (0.003)	-0.613 (0.002)
Literate			0.198 (0.003)	0.192 (0.003)	0.149 (0.002)
Can speak English			0.037 (0.005)	0.021 (0.004)	0.017 (0.003)
Urban area				0.116 (0.004)	0.133 (0.003)
Border State				0.012 (0.010)	
Segregation level				-0.175 (0.059)	-0.231 (0.052)
Log source country population				-0.007 (0.002)	-0.006 (0.001)
Log county population				0.045 (0.002)	-0.003 (0.005)
Mexican lynching in county b/w time $t-1$ and $t$				-0.132 (0.017)	-0.155 (0.014)
Year effects	Y	Y	Y	Y	Y
County fixed effects	N	N	N	N	Y
Observations	13,543,974	13,543,974	13,543,974	13,543,974	13,543,974
R-squared	0.004	0.299	0.301	0.328	0.351

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.



## **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 Untied 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.<sup>33</sup> 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

---

<sup>33</sup> 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.

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 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). Moreover, Hacker (2013) notes that under enumeration in the early 20<sup>th</sup> 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.<sup>34</sup> Table B1 shows the results from the probit. One interesting result is that linking immigrants relies more heavily on a 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.

---

<sup>34</sup> 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).

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

	Linked	Linked
Jaro-Winkler Distance, First name	-3.249*** (0.454)	-3.196*** (0.442)
Jaro-Winkler Distance, Last name	-12.04*** (0.884)	-11.64*** (0.858)
Year of Birth Difference=1	-0.275** (0.122)	-0.260** (0.118)
Year of Birth Difference=2	-0.426*** (0.131)	-0.409*** (0.127)
Year of Birth Difference=3	-0.549*** (0.140)	-0.542*** (0.135)
Years in US Difference	-0.239*** (0.0335)	
Years in US Difference squared	0.0148*** (0.00338)	
Total hits	-0.0642*** (0.0234)	-0.0640*** (0.0229)
Total hits squared	0.000509 (0.000766)	0.000534 (0.000746)
Unique and Exact NYSIIS Match, first name	0.822*** (0.173)	0.716*** (0.172)
Unique and Exact NYSIIS Match, last name	-0.0152 (0.234)	-0.0640 (0.229)
Unique and Exact NYSIIS Match, first and last name	1.570*** (0.136)	1.562*** (0.135)
Unique and Exact Match, last name	0.491*** (0.190)	0.457** (0.185)
Middle initial match	0.988* (0.523)	0.998* (0.513)
First letter of last name match	0.00182 (0.137)	0.0647 (0.136)
First letter of first name match	0.162 (0.149)	0.113 (0.143)
1 hit with exact last name NYSIIS match	-0.146 (0.132)	-0.126 (0.127)
1+ hit with exact last name NYSIIS match	-0.536*** (0.172)	-0.565*** (0.167)
1 hit with exact last name match	-0.628*** (0.149)	-0.608*** (0.144)
1+ hit with exact last name match	-1.355*** (0.123)	-1.272*** (0.119)
One hit	0.242 (0.190)	0.274 (0.185)
Absolute difference in length of last name	-0.299*** (0.0482)	-0.287*** (0.0468)
Exact NYSIIS last name match and year of birth diff==0	0.876*** (0.191)	0.886*** (0.186)
Exact NYSIIS last name match and year of birth diff==1	0.695*** (0.196)	0.676*** (0.191)
Exact NYSIIS last name match and year of birth diff==2	0.462** (0.203)	0.451** (0.197)
Constant	1.316*** (0.288)	0.720*** (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	II	III	IV	V
Start Year	Set to link	End Year	Linked Number	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.



Oxford Economic and Social History Working Papers

are edited by

Mattia Bertazzini (Nuffield College, Oxford, OX1 1NF)

Marco Molteni (Pembroke College, Oxford, OX1 1DW)

Papers may be downloaded from:

[https://www.economics.ox.ac.uk/50?search=working\\_papers&task=search](https://www.economics.ox.ac.uk/50?search=working_papers&task=search)