DISCUSSION PAPERS IN ECONOMICS

Working Paper No. 13-01

Who Crossed the Border? Self-Selection of Mexican Migrants in the Early 20th Century

Edward Kosack
University of Colorado at Boulder

Zachary Ward
University of Colorado at Boulder

August 2013, revised February 2014

Department of Economics

University of Colorado at Boulder
Boulder, Colorado 80309

Who Crossed the Border? Self-Selection of Mexican Migrants in the Early 20th Century

Edward Kosack
Department of Economics
University of Colorado at Boulder
edward.kosack@colorado.edu

Zachary Ward
Department of Economics
University of Colorado at Boulder
zachary.a.ward@colorado.edu

February 2014*

Abstract: We estimate the self-selection of Mexican migrants into and out of the United States in the 1920s. Officials recorded migrant height on border crossing manifests, which we use to proxy migrant quality and to measure self-selection into migration in 1920. Migrants were positively selected on height compared to the Mexican population. We link these migrants to the 1930 U.S. and Mexican censuses to obtain samples of permanent and return migrants and to estimate the selection into return migration. Return migrants were not differentially self-selected on height relative to permanent migrants.

JEL Classification: J61, J24, N0, N36
Keywords: Mexico, early twentieth century migration, self-selection, height

*We would like to thank Ann Carlos for her guidance and encouragement in this project. We would also like to thank Lee Alston, Francisca Antman, Brian Cadena, Dustin Frye, Michael Greenwood, Frank Lewis, Jason Long, Carol Shiue and Steven Smith for helpful comments. We are grateful to Chris Minns and Gill Newton for supplying the Double Metaphone algorithm for linking names. We would also like to thank participants at the annual meetings of the Economics and Business Historical Society, Western Economic Association International and the Canadian Network for Economic History. All errors are our own.

Introduction

Through the beginning of the twentieth century, Europeans dominated migrant flows to the United States, arriving freely with few laws restricting entry. This era of free mass migration ended abruptly in the 1920s with the Immigration Acts of 1921 and 1924 which imposed quotas to curtail European migration to the United States; however, migration from Mexico remained relatively unrestricted.[1] More individuals from Mexico arrived in the United States during the 1920s than did migrants from Ireland, Germany, Greece, Spain and other European countries (see Figure 1). Mexican migrants became an increasingly important source of labor in the United States in the early twentieth century, yet little is known about those who decided to migrate.

This paper asks both who decided to come to the United States from Mexico and who decided to make this move permanent. Specifically, we measure the pattern of selection into migration and then examine whether there is any differential selection into return migration. Because only some individuals are willing to cross borders and leave their native land, the economic consequences of the quality of migrants relative to those who remain behind could affect the home and host economies through multiple channels (Borjas, 1987). For the United States, the specific pattern of selection affects both migrant assimilation (Chiswick, 1978; Borjas, 1985; Ferrie, 1999) and the return to migration (Abramitzky, Boustan, and Eriksson, 2012b). For Mexico, whether those leaving were of higher or lower quality than those staying is important for understanding potential "brain drain" (Gibson and McKenzie, 2011), as well as income inequality (McKenzie and Rapoport, 2007).

[1] The Emergency Immigration Act of 1921 and Immigration Act of 1924 placed annual limits on European migration while imposing no restrictions on Western Hemisphere countries. While Mexican migrants were not limited by quotas, the Immigration Act of 1917 did require all migrants to pass a literacy test and to pay an eight dollar head tax.

If migrants were better (worse) than the Mexican population in terms of productivity, education or health, then they were positively (negatively) self-selected (Chiquiar and Hanson, 2005). In order to determine the pattern of selection, one can compare the wages that Mexican migrants would earn in Mexico to the wages of those in Mexico who do not migrate (Borjas, 1987).[2] However, migrants and their wages are typically only observed in the host country. As prices for skills vary from country to country, comparing wages once migrants have crossed the border does not give the proper counterfactual.[3] Further, in many cases individual wages are not known. Some studies of historical selection use aggregated measure of human capital, such as occupational scores, to compare movers to stayers (Abramitzky, Boustan, and Eriksson, 2012a; Collins and Wanamaker, 2014). However, if occupations reported in the historical immigration statistics were not representative of an individual's place in the skill distribution (i.e., occupational downgrading upon arrival), then these data would systematically underestimate the true quality of a migrant worker. Additionally, occupational scores are not specific to the individual. Using a disaggregated measure of human capital will provide more precise information about the pattern of selection, and will also allow us to look at how migrants differed from the home population within reported skill class.[4]

[2] Borjas (1987) defined selection not only in terms of comparing migrants' wages to the home country's distribution, but also in terms of how they compared to the host country's distribution of wages. We follow the recent direction of the literature comparing migrants only to those in the home country (see Chiquiar and Hanson, 2005).

[3] Techniques used to circumvent this problem to generate the appropriate comparison of migrants and non-migrants include propensity score matching (Chiquiar and Hanson, 2005) and sibling fixed effects (Abramitzky, Boustan, and Eriksson, 2012b).

[4] It is unclear which occupation was reported for migrants; the one that migrants were going to have in the United States or the one they previously had in Mexico. It is possible that migrants listed downgraded occupations on arrival that reflect labor demand in the United States. However, reported occupation is likely correlated with actual occupation, which makes comparing movers to stayers within occupational skill class informative.

We use height as an alternative and improved measure of the historical self-selection of Mexican migrants. A long literature argues that greater stature is correlated with higher earnings, greater intelligence, and increased health; in other words, height is positively correlated with "quality" (see Steckel, 1995; Steckel, 2009).[5] A migrant's height does not change as he crosses the border into the United States, unlike occupation or wages. Further, height gives a measure of human capital that is specific to the individual, important when there is little variation in migrant occupation. Since the vast majority of Mexican migrants claimed laborer or miner as their occupation, we are able to determine if the United States received the better laborers or the better miners by using height data.

We examine not only characteristics of the Mexican flow *into* the United States, but also the characteristics of the Mexican flow *out of* the United States. Much of the migration from Mexico to the United States was circular, and many individuals returned home instead of settling permanently (Gratton and Merchant, 2013). Measuring the selection into migration is not sufficient for understanding the effect of migrants on the labor force in both Mexico and the United States since return migrants might be differentially self-selected (Borjas and Bratsberg, 1996). However, the direction of selection for return migrants is unclear. Return migrants may have been "target earners" that migrated to accumulate savings in order to start a business back home, leaving the quality of these migrants ambiguous (Mesnard, 2004). On the other hand, if return migrants made their decision ex post, they could have been those who failed in the United States labor market and would thus be negatively self-selected (Abramitzky, Boustan and Eriksson, 2012a). We measure the self-selection into return migration to determine the long-run effects of migration on the stock of workers in both Mexico and the United States.

[5] Thus, using height allows one to infer the selection of migrants on a number of different but highly correlated dimensions of human capital including productivity, health, and education.

We utilize newly collected data from individual border manifests for migrants crossing through border towns in Arizona and Texas in 1920. To determine the selection of the migrant population compared to the home population, we compare heights for migrants to samples of heights for soldiers in the military and for those who applied for passports.[6] Having estimated the self-selection of inflows, we estimate the self-selection of outflows. To do this, we link our sample of migrants who crossed the border in 1920 to the 1930 United States Census to create a sample of permanent migrants, and to the 1930 Mexican Census to create a sample of return migrants. We compare the heights of each sample to determine the self-selection of return migrants relative to permanent migrants.

We find that Mexican migrants in 1920 were positively self-selected on height from the Mexican population. They were four to five centimeters taller than soldiers in the military, members of the lower class of Mexican society, and they were only one and a half centimeters shorter than passport holders, members of the higher class of Mexican society (López-Alonso and Condey, 2003). Our result holds within occupational skill class as the United States received the taller laborers, the taller skilled workers, and the taller professionals. We also find that although a substantial proportion of Mexican migrants returned home (between 13 and 44 percent), there is no differential self-selection on height into return migration. Our measured result of positive selection for migrant inflows is a good proxy for the change in the quality of the overall stock of Mexican migrants in the United States in the early twentieth century.

In the next section, we provide some historical context for the state of Mexican migration to the United States in 1920. Then, we explain selection into migration and the benefits of using height to proxy migrant quality. We follow with a description of our hand-collected dataset used

[6] The military and passport height data was collected by López-Alonso and is publicly available at the ICPSR (2003).

to estimate the pattern of selection from Mexico in 1920. Because selection into migration is not sufficient for understanding the total impact of Mexican migration since it does not account for those who return home, we describe how we construct a linked dataset to both the United States and Mexican censuses, which we use to estimate the selection into return migration. We end with a discussion about the overall impact of migrant selection on the stock of workers in both Mexico and the United States.

U.S. – Mexico Migration in 1920

There is an extensive literature on the history of migration between the United States and Mexico (see Cardoso (1980), Ettinger (2009), and Gutierrez (1995) for an overview). Indeed, Mexican migration patterns transformed dramatically during the early twentieth century. The Mexican Revolution pushed migrants out during the 1910s, while the immigration quotas of 1921 and 1924 curtailed unskilled labor from Eastern and Southern Europe in the 1920s and pulled Mexican workers into the United States. We choose 1920 as a benchmark year, falling as it does directly between these two major events, to reveal how the self-selection process operated with limited, confounding institutional factors.

While there were some restrictions to entering the country in 1920, picking a year prior leads to several challenges for our analysis. First, the fighting in the Mexican Revolution was heaviest between 1910 and 1917, making it difficult to separate migrants moving for economic reasons versus those fleeing as refugees. Although some small amount of fighting continued in 1920, it was limited to the North while most of our sample comes from central Mexico. Second, the United States only started to systematically collect immigrant records for individuals crossing the Mexican border in 1907, and the process was not settled by 1909, the year before the Mexican Revolution (Immigration Act of 1907, Sec. 32).

The Mexican Revolution, a multi-sided conflict, raged during the early 1910s although the major fighting subsided by 1920.[7] At the beginning of the Revolution, conflict occurred in both northern and southern parts of Mexico as revolutionaries from different states fought to overthrow President Díaz, with the most intense fighting occurring between 1913 and 1916. Following the creation of a new constitution in 1917, major warfare subsided with only Pancho Villa skirmishing in small battles in the North. By 1920 most fighting halted as Villa surrendered and Álvaro Obregón was elected to the presidency (Knight, 1986).

During the Revolution, thousands of Mexicans temporarily fled to the United States (Report of the Commissioner-General of Immigration, 1914). As a result of refugees fleeing during the revolution the migrant flows became more skilled between 1913 and 1916 during the most intense period of fighting, but by the end of the 1910s the skill mix of the inflow had returned to pre-Revolutionary levels (see Figure 2). Even though thousands crossed the border, the United States absorbed these migrants easily as World War I increased the need for labor (Rockoff, 2004). In fact, in 1917 the United States encouraged temporary Mexican migrants to work in agriculture, railroads, and mining, briefly suspending entry restrictions by allowing contract laborers, discontinuing the head tax, and waiving the literacy requirement (Cardenas, 1975). By 1920 thousands of Mexicans traveled northward yearly to earn higher wages offered by employers in the United State, but many of these same migrants also returned home (Clark, 1908; Report of the Commissioner-General of Immigration, 1920).

As Congress was encouraging migration to the United States from Mexico, they simultaneously passed qualitative restrictions on migration in 1917 by requiring migrants to be able to read and write in their own language. While this policy was aimed at limiting Southern and Eastern Europeans migration, it could also limit a significant portion of the largely illiterate

[7] See Knight (1986) for a review of the Mexican Revolution.

Mexican population. However, the United States did not consistently enforce this law for Mexican migrants. For example, the literacy requirement and head tax were waived in 1917 to encourage migrants to fill the labor shortfall during World War I (Cardenas, 1975). The U.S. government continued to reissue the waiver from time to time until March 1, 1921 to allow individuals to enter for agricultural work, especially sugar beet production (Cardoso, 1976). The waiving of the literacy test is clear when comparing literacy rates to the skill mix of inflows, as we show in Figure 2. Prior to the literacy test in 1917, literacy and migrant skill level were positively correlated, as expected; however, following 1917 the correlation became negative. For instance, the percent of the migrant flow that was unskilled increased from 71% to 79% in the first year after the literacy test while the percent literate of Mexican migrants increased from 85% to 95% for the same year.[8] By the year of our study in 1920, the percent unskilled was even higher at 84%, while the literacy rate increased to 99.4%. Even when agricultural workers were waived from the literacy test and head tax, official statistics probably still recorded them as literate.

The literacy test did not appear to sufficiently restrict migration from Southern Europe and so Congress imposed quantitative restrictions in 1921 and 1924, dramatically reducing migration from Europe (Zeidel, 2004). The quota system, however, placed no limits on migrants coming from the Western Hemisphere, and so Mexican migration was relatively unimpeded. Following the quotas, Mexican immigration increased dramatically as Mexicans acquired jobs due to a labor shortfall (Bloch, 1929). The large increase in numbers would eventually lead to concerns over the racial origins of Mexican migrants (Foerster, 1925), to the creation of the

[8] These numbers are based on authors' calculations from the Reports of the General Commissioner of Immigration (1908-1930). Before 1917 the literacy rate is calculated for individuals 14 years and older, after 1918 it was for 16 years and older.

Mexican Border Patrol and to the criminalization of undocumented entry in the 1920s (Ngai, 2002).

Selection into Migration

Migrants are not a random draw from their home country's population. While many have observed that migrants are somehow different from those who stay in the home country, Borjas (1987) first argued that we can predict the direction of self-selection for migrants based on the relative distribution of wages across economies. If human capital is rewarded more (less) in the United States compared to the home country, then people with more human capital are more (less) likely to migrate. If the migrants who leave from the home country are on average better (e.g., more motivated, more educated, higher wages, etc.) than those who stay, then the self-selection is positive; if migrants are worse along these dimensions than stayers, then self-selection is negative.

Selection is influenced by the variation in expected benefits of migration across the human capital distribution of potential migrants. In the early twentieth century, the benefits of migration were immediate as job opportunities were plentiful for Mexican workers. In the southwestern United States, many farms, railroads and mines hired migrants directly at the border (Report of the Commissioner General of Immigration, 1920). While low-skilled jobs were readily available, high-skilled jobs were not as prevalent. Also, wages were higher for these unskilled laborers than in Mexico, suggesting a significant return to migration for unskilled laborers, which could lead to negative self-selection (Clark, 1908).

Selection is also influenced by the variation in costs of migration across the human capital distribution of potential migrants. While high wages abroad may entice an individual to move, his mobility might be restricted by the costs of moving (e.g., transportation,

psychological, informational or opportunity costs). Even though low-skilled Mexicans could earn a premium in the United States, if the costs of migration are sufficiently high, then the flow of migrants would only consist of those able to afford to migrate. Chiquiar and Hanson (2005) argue that the high costs of migration reconcile the theoretical prediction that contemporary Mexican migrants should be negatively self-selected with the empirical evidence that they are intermediately self-selected. McKenzie and Rapoport (2010) show that networks lower the cost of migration and can alter the pattern of self-selection. Fernandez-Huertas Moraga (2011) suggests that differing costs of migration could determine why we see negative selection into recent migration from urban areas of Mexico where it is easier to travel to the United States and positive self-selection from rural areas of Mexico.

The benefits to migration were clear but the costs may have constrained individuals from traveling. Transportation costs were non-trivial. While improvements in transportation from central Mexico to the United States border, especially the completion of the Mexican railroad in the late nineteenth century, lowered the cost of migration and subsequently spurred large waves of emigration, the cost of a ticket from central Mexico to the United States border was still high for poorer individuals (Clark, 1908; Coatsworth, 1981). Additionally, the 1917 migration legislation required all migrants to pay an eight dollar head tax. Although the enforcement of this law during 1920 is unclear, if low-earning individuals were unable to finance the trip abroad, then self-selection could have been positive.

While a handful of papers analyze the selection of migrants from Europe (Abramitzky et al, 2012b, 2012a; Hatton and Williamson, 2006; Stolz and Baten, 2012), little is known about selection of Mexican migration to the United States during the early twentieth century. Feliciano (2001) is the only paper to our knowledge that explores the historical self-selection of Mexican

migrants, finding that in 1910 Mexicans in the United States had a higher rate of literacy than did individuals in Mexico. We extend her results by incorporating evidence on immigrants following the Mexican Revolution, by using a measure (height) that is constant across borders, and by exploring the self-selection of return migrants which could alter the quality of the stock of Mexican migrants observed in the census.

Height as a Measure of Selection

Multiple metrics of human capital have been used in studies of selection, including income (Chiquiar and Hanson, 2005), skill class (Hatton and Williamson, 2006), occupational scores (Abramitzky et al., 2012a; Collins and Wanamaker, 2014), age-heaping[9] (Stolz and Baten, 2012), years of education and literacy (Feliciano, 2001). While each of these metrics provides insight into the selection of migrants, they are not available for our sample of migrants because we have no data on the wages and education level of Mexican migrants in 1920. Therefore, we employ a different metric, height, to measure the quality of an individual migrant.

When income and wage data are not available, economists must rely on other measures to proxy for standard of living. In particular, height as a measure has been used since it is positively correlated with income and improved health and nutrition (See Steckel, 1995 and 2009 for a review of height studies). Higher living standards with ample food during childhood increase height, while poor nutrition and health can stunt growth.

Not only does the average height of a society indicate overall health and well-being, but taller people also earn more than their shorter counterparts within a country. For example, Schultz (2002) shows that the return to a one centimeter increase in height is comparable to an additional year of schooling. The return to physical strength is especially important in

[9] Age-heaping is based on the observation that many people do not know their actual age but round to the nearest 0 or 5 when reporting it. A'Hearn et al. develop a measure (called ABCC) that is the percent of the population that knows their true age (A'Hearn, Baten, and Crayen, 2009).

developing countries where large sectors of the economy rely on the physical productivity of labor. Height is a determinant of wages in these countries since larger and stronger men (as measured by BMI) are rewarded in the labor market (Thomas and Strauss, 1997). Mexican migrants worked in labor-intensive jobs, such as mining, railroad construction, and farm labor, where improved physiology could lead to higher productivity (Clark, 1908; Report of the Commissioner-General of Immigration, 1920).

Persico et al. (2004) argue that higher wages for taller individuals are due to non-cognitive characteristics (e.g., confidence), while others (Case and Paxson, 2008; Schick and Steckel, 2010) argue that early childhood inputs into health and nutrition can increase the cognitive functioning of an individual later in life. For example, taller individuals are more likely to remember their exact date of birth (Humphries and Leunig, 2009) and taller individuals score higher on early childhood cognitive and non-cognitive tests (Case and Paxson, 2008). While the return to physical strength explains much of the wage premium for stature in developing countries, increased cognitive functioning can perhaps explain why even in developed countries taller individuals earn more (Steckel, 2009). Either way, the evidence suggests that taller individuals, on average, earn higher wages. If the migrants who arrived in the United States were taller than those who remained in Mexico, then this would indicate a pattern of positive selection for Mexican migrants.

Migrant height is well suited for measuring selection in the context of early twentieth century migration from Mexico. First, adult migrant height does not change as the individual crosses the border. Comparing wages across borders can yield invalid predictions since the return to skill can vary by country. Moreover, recorded occupation could underestimate a migrant's true position in the human capital distribution if they downgrade to unskilled jobs

which are plentiful and yield relatively high wages. Second, height is a measure of human capital that is measured at the level of the individual. There is little variation in the occupations of Mexican migrants so occupational measures of self-selection yield little information about the quality of migrants. Additionally, measuring human capital at the individual level allows us to estimate selection overall, as well to compare heights within reported skill class.

Data

Border Crossing Manifests

In order to understand exactly who migrated to the United States from Mexico in 1920, we construct a unique dataset from the manifest lists for those crossing at the border towns of Ajo Arizona, Douglas Arizona, Brownsville Texas, and El Paso Texas in 1920.[10] In Figure 3 we show the geographical coverage of our sample. While some studies use ship records arriving from the Atlantic to study immigration to the United States, we are unaware of any other study that uses these border manifests (Ferrie, 1999; Cohn, 2009; Bandiera, Rasul and Viarengo, 2013). Height was recorded on each manifest by border officials and was often rounded to the nearest quarter inch. In addition to height, a wealth of information about migrants upon arrival was recorded on the manifest, including demographic (age, sex, marital status), geographic (place of birth, place of last residence, intended destination), economic (occupation, savings), and network (join a friend, relative or employer) data. We collect all available data for each

[10] We accessed the manifest lists for land border crossings from Mexico from both the Rocky Mountain Regional Office of the National Archives in Denver and the genealogy website, Ancestry.com. As the map in Figure 3 shows, these towns are well-distributed along the border. Furthermore, there is no systematic difference in the outcome of interest (height) across border towns after controlling for state of birth and decadal fixed effects, suggesting that heights were consistently measured across border stations.

adult male (18 years or older) classified as an immigrant.[11] In total, we have microdata for 3,671 male migrants who crossed the border in 1920.

To determine the representativeness of our sample, we compare the characteristics of our migrants with those of similar migrants recorded in the 1920 United States Census. We use the 1% 1920 IPUMS sample to identify migrants who arrived in the previous year, who were literate, over the age of 18, and male (Ruggles et al., 2010). Our sample is representative of the distribution of skills for migrants recorded in the census with no statistical difference in occupational mix between the two samples.[12] There is also no difference in marital status, although our sample is about two years younger and overrepresented by people moving to Texas.[13]

Our dataset is constructed from information collected by border officials as individuals crossed the border from Mexico into the United States. Therefore, our sample of immigrants contains only those who were documented by crossing at an official border crossing station rather than those migrants who entered the country unobserved.[14] Although Bloch (1929), in a comparison of census numbers with net migration flows, estimates that undocumented entries could have been substantial for the decade from 1910 to 1920, he also admits that there is a lack of reliable information to make study of this population feasible.[15] To be precise, our results

[11] We employ a systematic approach to the collection of these data. An observation was collected if and only if the individual's intended length of stay was listed as permanent or indefinite, the last permanent residence was outside of the United States, the place of birth was outside of the United States, and the final destination was within the United States. These are similar to the criteria used by the United States officials to classify each individual as a measureable immigrant on the form.

[12] Results for the representativeness of the sample are available from the authors upon request.

[13] The fact that our sample is overrepresented by people headed to Texas is an artifact of the majority of it being recorded from the El Paso and Brownsville border stations.

[14] Migrating to the United States was not technically illegal until later in the 1920s, when the United States government created the Border Patrol in an attempt to stop Mexicans and other European ethnicities that tried to enter the United States through the south (Foerster, 1925, Ngai, 2002).

[15] Bloch (1929) compares the change in the number of Mexican individuals from the 1910 to the 1920 Census with net migration flows (i.e., immigration flows minus emigration flows) to estimate that over

apply to those migrants who crossed through official border crossing stations, and not necessarily to all migrants (i.e., those who crossed without going through an official crossing station).[16]

The border-crossing data allow us to create a profile for the "typical," documented, male migrant who crossed the border from Mexico to the United States in 1920, shown in Table 1. Male migrants to the United States were, on average, 29 years old, equally likely to be married as single, and almost universally recorded as literate. Immigrants came most often from central and northern Mexico, with very few coming from the southern states.[17] A large portion of our migrants were born in the Mexican states of Chihuahua, Guanajuato, and Jalisco, which are still high-sending states today, and most reported a final destination of Texas. Only 14% of migrants in the sample reported meeting someone (friend, relative or employer) upon entry, much lower than Europeans in 1920 with 83% of Germans, 96% of Italians, and 97% of Greeks joining a network upon arrival.[18] Also, networks varied across Mexican birth state. Fewer than 10% of people from the high-sending states of Guanajuato, Jalisco, and Zacatecas reported joining a network upon arrival, while those states that bordered the United States, such as Sonora and

---

111,000 entered the country without documentation from Mexico over the course of the decade. In his own study of Mexican migration, however, he admits that reliable figures for these individuals do not exist and he must concentrate his own analysis on those who crossed through official stations such that they are included in the official immigration statistics.

[16] We recognize that those crossing as undocumented migrants could differ from those crossing at official border stations and that the self-selection for these individuals could be different. In particular, they might be negatively selected. We conduct a robustness check to test the sensitivity of our results to undocumented migration and find that the result holds, even for extremely negative selection of undocumented migrants.

[17] We follow the same region of birth classification as López-Alonso and Condey (2003) to maintain consistency across samples. Region of birth is split into North, Bajio, Center, and South. North includes Baja California Sur, Baja California, Sonora, Sinaloa, Chihuahua, Coahuila, Nuevo León, and Tamaulipas. Bajío includes Jalisco, Colima, Michoacán, Nayarit, San Luis Potosí, Durango, Zacatecas, Aguascalientes, Querétaro, and Guanajuato. Center includes Distrito Federal, México State, Morelos, Tlaxcala, Puebla, Veracruz, and Hidalgo. The South includes Guerrero, Oaxaca, Tabasco, Campeche, Yucatan, Quintana Roo, and Chiapas. It is well noted that the construction of the Mexican railroad helped transport Mexicans to the United States. However, the railroad did not reach the southern states below Veracruz by 1920, which explains why few of our observations are from the southern Mexican states.

[18] Based on authors' calculations from the Report of the Commissioner General of Immigration (1920).

Tamaulipas, had over 40% of people meeting someone across the border. On average each Mexican migrant brought $39 cash with him across the border, lower than non-Mexican migrants arriving through the port of New Orleans in the same year who brought over $200.[19] Migrants from Mexico, however, probably did not need much cash on hand to cover their costs when they arrived, as most came from and settled in areas quite close to the border.

We classify migrants as unskilled, skilled or professional workers based on their reported occupation.[20] The majority (about 87%) of immigrants in the sample were unskilled. It is because of this lack of variation in skill class that occupational rankings yield little information in determining self-selection. Height allows us to examine whether migrants, within a given occupational class, were better or worse than non-migrants remaining in Mexico. Of course, to determine whether migrants from Mexico were positively or negatively selected requires reference groups of individuals remaining in Mexico.

Comparison Samples: Military and Passport Data

To make an inference about the selection of migrants from Mexico we need to compare the heights of migrants to those living within Mexico. Here we use two distinct samples—military soldiers and passport holders. Bodenhorn et al. (2013) warn that samples of historical heights are likely selected, which could lead to incorrect inferences about the underlying population. We acknowledge that both of these samples are not representative – the military sample is from the lower part of the height distribution of Mexico, and passport records are from the upper part of the height distribution of Mexico. However, by comparing migrants to both samples and determining which sample migrants most closely resemble we can infer whether migrants were positively or negatively selected.

[19] This figure comes from similar work that we have done with passenger lists for boats arriving at New Orleans in 1920.
[20] We follow López-Alonso's (2000) occupational classification.

The *Secretaría Nacional de la Defensa* houses federal military records in the *Archivo de Concentración,* recording deceased soldiers in the *Sección de Personal Extinto* and deserters in the *Sección de Cancelados* (López-Alonso and Condey, 2003).[21] Since the military did not have required service until 1939, only those who made the choice to join the military appear in the data, thus it is not representative of the entire Mexican population. Indeed, López-Alonso (2007) argues that the federal military represents the lower middle class of a highly unequal Mexican society.

Characteristics of the military sample are also listed in Table 1. It shows that 77% of military males were in unskilled occupations and that individuals were well-represented across different regions of Mexico. At first glance, the military sample appears to be higher skilled than the migrant group, since 87% of migrants were unskilled compared to 77% of individuals in the military, implying negative self-selection. However, migrants may have reported intended occupation rather than previous occupation, leaving their true position in the skill distribution of Mexico unclear. Importantly, a comparison of average height reveals that migrants were approximately five centimeters taller than those in the military. We illustrate this comparison in Figure 4 by showing that the estimated height distribution for the migrant sample lies well to the right of the estimated height distribution for the military sample.

We also compare migrants to a sample of passport applications from Mexico collected by López-Alonso (2003) from the *Archivo de Pasaportes*. Those holding passports did so for business and leisure and this group reflects an underlying population with the funds to afford

[21] Birth records did not become widely available until the 1930s, so the military kept track of members (who might potentially desert) by recording their height, place of birth, age and occupation. The *Sección de Cancelados* contains information on members of the military who deserted the army before their service time ended, and the *Sección de Personal Extinto* contains individuals who died in service or retired and then died afterwards (López-Alonso, 2012). The majority of the military data is for individuals who joined the Mexican Army between 1915 and 1935.

such travel. Unfortunately, this sample only includes age and does not give region of birth or skill classifications. Height was not measured for passports but was self-reported, possibly creating an upward bias since height tends to be over-reported (Spencer et al., 2002). If Mexican migrants were positively selected, the tendency to over-report on passport applications would bias us against finding this result. Summary statistics in Table 1 show that passport applicants were only about one and a half centimeters taller than those immigrating to the United States. In Figure 4 we show that the estimated height distribution for the migrant sample lies very close to the estimated height distribution for the passport sample. While the average migrant was over four centimeters taller than the average member of the military, he was similar in height to the average passport applicant.

Estimating Self-Selection into Migration

We utilize a linear regression model to explore the pattern of selection among Mexican migrants in 1920 as measured by migrant height. Although the analysis of the estimated densities in Figure 4 suggests a pattern of positive selection, it is possible that greater stature is simply correlated with other characteristics that are more prevalent in the migrant sample, such as a particular region of birth. Thus, we estimate Equation (1) to control for many of these additional characteristics that could confound our positive selection result.

$$Height_i = \beta_0 + \beta_1 Migrant_i + \delta' X_i + \epsilon_i \quad (1)$$

An individual's height is regressed on a constant, an indicator variable for whether or not the individual is from the migrant sample, and a vector of controls. In the vector of controls, we include dummy variables for age bins of 18 to 20 years and 21-23 years in order to account for

patterns in human growth rates.[22] We also include controls for decade of birth to account for any conditions that may have affected the height of all those born in Mexico during those times.[23] Furthermore, we include geographic controls to account for any spatial pattern in Mexican heights.[24] Finally, we include controls for occupational skill class which allows us to describe how migrants differed from others within skill class.

Results of the selection regressions comparing the sample of male migrants to the sample of males in the Mexican military are presented in Table 2. Column (1) is a basic comparison of means and Columns (2) through (4) systematically add controls to the regression. First, our regression model reveals expected patterns in heights. For example, adults in the 18 to 20 year age bin were shorter than adults over 24 years old, while those in the 21 to 23 year age bin were only slightly shorter and the difference loses statistical significance. This is consistent with the growth pattern of humans where heights increase at a decreasing rate up to around age 24. Also, those in the skilled class were taller than those in the unskilled class, while those in the professional class were taller than individuals in either of the other two occupational skill classes, supporting the claim that height is correlated with income, productivity and cognitive ability.[25] Second, the result of positive selection as measured by height holds in each of these specifications, with the migrant sample measuring four to five centimeters taller than those individuals in the military sample. Migrants were taller than those in the military even though they reported lower skilled occupations. Finally, in Column (4) we show that migrants were

[22] Final adult height may not be reached until 24 years of age and so individuals who are between 18 and 24 years might still be growing. The results are qualitatively similar in regressions that exclude those under 24 years of age.
[23] Results are robust to the inclusion of birth year fixed effects.
[24] For example, those born in the North region are significantly taller than those in other regions, consistent with a diet richer in protein, which leads to taller individuals (Steckel, 2009).
[25] Even though reported occupation might not reflect true occupation for some migrants, height is still positively correlated with reported skill class. This suggests that reported skill class and actual skill class are correlated.

taller than those in the military within occupational skill class. Although the descriptive statistics show that those who chose to migrate tended to come from lower skilled occupations, we find that within skill class the individuals who migrated tended to be taller than those who were in the military. As we show in Table 3, the height premium was largest for professional workers, who were over five centimeters taller than those in the military, while unskilled and skilled workers were about four centimeters taller than their military counterparts.

We also present in Table 2 the results of selection regressions comparing the sample of male migrants to the sample of males applying for Mexican passports. Column (5) again shows a simple comparison of means between migrants and passport applicants, while Column (6) includes controls for ages less than 24 years and decade of birth.[26] We confirm the result from the simple comparison of height distributions in Figure 4. Those in the migrant sample were, on average, just under a centimeter and a half shorter than those in the passport sample. Given that the difference in height is quite small and the fact that those holding passports probably came from the upper end of the distribution in Mexican society, this is additional evidence consistent with a pattern of positive selection into Mexican migration in 1920.

Robustness of the Results

We present alternative specifications in Table 3 to address concerns about the Mexican Revolution and the effect of the 1917 literacy test requirement. It is possible that the self-selection result is not due to economic forces but rather because of refugees fleeing the Mexican Revolution. We test for differences in the pattern of selection by region of birth to determine whether positive selection was strongest in the North where fighting continued, and show the results in Table 3. Migrants born in the southern parts of Mexico had the largest height premium,

[26] Fewer controls are included in this model because the passport sample lacks much of the detail found in the military sample.

while those born in the Bajio region had the smallest. This could be because the costs of travel were relatively higher from the South than from the Bajio region. Similar to the result found for the whole sample in the main specification, migrants born in the northern parts of Mexico were just over four centimeters taller than non-migrants and did not exhibit an abnormal or extraordinary pattern of selection that would give cause for concern.

It is possible that Mexicans who crossed the border were taller than those who remained in Mexico since migrants needed to pass a literacy test. While the degree of enforcement of the literacy test for Mexican migrants in 1920 is ambiguous as discussed earlier, we compare our sample of migrants to a subsample of 3,884 military deserters for whom we have literacy data, recognizing that there is a difference in how literacy is determined in the migrant and military samples. The literacy test required the migrant to read and write a paragraph of twenty five words in a language of his choosing (Goldin, 1994), while literacy in the military sample was determined by whether or not the soldier could sign his name (López-Alonso and Condey, 2003). Our finding that migrants were positively selected still holds when comparing literate samples, and literate migrants were over three centimeters taller than their counterparts in the military.

Our results indicate that documented migrants to the United States in 1920 were positively self-selected from the home distribution but does not account for undocumented entry. Bloch (1929) estimates that roughly 111,000 undocumented individuals entered the United States over the decade ending in 1920.[27] Using this number in combination with the official statistics

[27] This estimate in Bloch (1929) is based on estimates of the Mexican-born population in the United States found in official statistics. Gratton and Gutmann (2000) revise official statistics for the population of Mexican origin in the United States, but these estimates include generations beyond those born in Mexico. Gutman et al. (2000) show that the official statistics counting individuals born in Mexico are very close to estimates made from IPUMS microdata, and they say that the small discrepancies are probably the result of sampling error. Gratton and Merchant (2013) also include estimates of the Mexican-born population from IPUMS microdata, and these estimates are very similar to the ones used in Bloch (1929).

for migration in the 1910s, a back of the envelope calculation suggests that the average undocumented migrant would need to have been 154.29 centimeters tall (nine and a half centimeters shorter than the average male in the military and fourteen and a half centimeter shorter than the average documented migrant) to erase the height advantage over the military.[28] This means that even though institutional constraints could cause negatively self-selected individuals to migrate unofficially, it is unlikely that undocumented migration would cause a reversal of our positive selection result.

Accounting for Return Migration

Selection into Return Migration

Simply measuring the selection into migration is not sufficient to understand its long-term impact, especially when return migration was prevalent as in the case of Mexico. Even though migrants were self-selected from the Mexican population, return migrants could be differentially self-selected from the overall set of migrants, changing the quality of the stock of migrants that remained in the United States permanently and the quality of the stock of labor in Mexico (Borjas, 1985; Borjas and Bratsberg, 1996).

Whether return migrants were positively or negatively self-selected from the migrant population is ambiguous. One possibility is that most migrants were "target earners" and returned when enough was saved to invest in capital back home (Mesnard, 2004; Angelucci,

[28] The official migration statistics for the United States show that 219,004 individuals entered the country legally from Mexico from 1911 to 1920. Thus, the total flow from Mexico for the decade was 330,004, with undocumented entrants accounting for 33.6% and documented entrants accounting for 66.4% of that flow. Our simple comparison of means between the migrant sample and the military sample shows that legal migrants were, on average, 4.83 centimeters taller than those in the Mexican military. If we use a weighted average of documented and undocumented migrants to measure selection, we can calculate how short the average undocumented migrant would need to be to erase the 4.83 centimeter advantage over the military.

2012).[29] It is possible that these target earners were the most entrepreneurial and productive of migrants, leading to positive self-selection of return migrants. An alternative to the target-earnings model is that the decision to return was made ex post when outcomes in the United States were worse than expected (Borjas and Bratsberg, 1996). If return migrants were those who failed in the labor market, then return migrants would have been negatively self-selected.

For the early twentieth century, Abramitzky, Boustan and Eriksson (2012a) show that European migrants to the United States who returned home were negatively self-selected, suggesting that these individuals were those who failed in the labor market. However, Mexican return migrants could have been different from European return migrants, especially considering the proximity of the United States to Mexico.[30] The proximity could have lowered return costs for migrants and increased the feasibility of short and repeated trips to the United States, which may be why present-day evidence finds that return migrant behavior for Mexicans is consistent with a target-earnings model (Lindstrom, 1996; Angelucci, 2012; Nekeoi, 2013).

Linked Sample

Migrants can be split into two different categories: those who stay permanently and those who return home. To estimate the self-selection of return migrants, we link our sample of 3,671 migrants forward to the 1930 United States Census for a sample of permanent migrants, and forward to the 1930 Mexican Census to get sample of return migrants. The link to the 1930 United States Census is based on four characteristics: first name, last name, year of birth and

[29] If the entire migrant population were target earners, then the self-selection of those who return *first* would be positive since the most productive hit their earnings quicker. However, we measure self-selection of return migrants over a ten-year span, which suggests that most target earners, whether highly or lowly productive, have already returned.

[30] Ward (2013) uses administrative records from the United States and finds that Mexican return migrants in the early twentieth century were positively self-selected on occupation, suggesting that return migrants were more likely to be target earners. He points out, however, that administrative records could be overrepresented by ship crossings (containing relatively higher quality return migrants) compared to border crossings (containing relatively lower quality return migrants).

country of birth (Mexico). We also link our sample to the 1930 Mexican census based on the same four characteristics, but are able to match on state of birth in Mexico. We follow the iterative matching procedure similar to Abramitzky, Boustan, and Eriksson (2012a) where we match based on names and a two-year birth window.[31] In order to limit bias from transcription errors, we also standardize names using the Double Metaphone algorithm.[32]

Our linking strategy produces a set of migrants who are either uniquely linked to the United States Census or to the Mexican Census, linked multiple times to the same census, not linked to either census, or linked to both censuses.[33] Failure to link to either census is most likely due to death, name change or transcription error while linking to both censuses or multiple times to the same census is likely due to extremely common names; all of these groups are dropped from the sample.[34] From the original 3,671 migrants, we have a sample of 632 individuals uniquely linked to the 1930 Mexican Census, and 798 uniquely linked to the 1930 United States Census.[35] The match rates of 17.2% and 21.7% for the Mexican and United States

[31] A detailed description of the matching procedure as well as the linking matrix can be found in the appendix.

[32] We standardize names using the Double Metaphone algorithm, graciously provided by Chris Minns and Gill Newton. We use Ancestry.com to perform the linking process.

[33] Migrants that are linked to both censuses present a peculiar problem. On the one hand, they could represent extremely common names that are likely to show up both in the United States and in Mexico and thus should be dropped since their location is indeterminate. On the other hand, these cross-links could actually be in the United States but the household in Mexico enumerates them to Mexican census takers since they consider them "Mexican" and part of the household. The 1930 Mexican Census enumerated all individuals that "usually" live in the household. We present robustness checks to determine whether the assigning the cross-linked group to the U.S. Census or dropping them changes our results.

[34] It is possible that Mexican migrants could have migrated to a third country, but this is unlikely during this time period.

[35] In addition to these matches, there are 1,765 migrants who are unlinked and 261 matched to both censuses. The enumeration date for the United States census was April 1st, 1930, and the enumeration date for the Mexican Census was May 15th, 1930, so it is possible that migrants left in between dates to be counted in both countries. However, given that migrants were already in the United States for ten years, this is unlikely.

censuses are similar to those found for studies of European migrants during the same time period (see Table 1 in Abramitzky, Boustan and Eriksson (2012a)).

There are a total of 632 return migrants out of a total 1,430 uniquely linked migrants, which yields a 44.2% return rate after ten years of stay. This rate is likely an upper bound for the true rate of return since transcription error and name changes are more likely to occur in the United States. We can compare our estimate with official United States records that recorded the number of return migrants to Mexico during this time period. According to administrative data, the return migration rate for males in the decade from 1920 to 1930 was approximately 13.3%.[36] However, these administrative records probably undercounted out-migrants and so this would give a lower bound on out-migration rates (Bandiera, Rasul, and Viarengo, 2013).

Despite the fact that permanent migrants and return migrants end up in different countries, their characteristics upon arrival, as shown in Table 4, were remarkably similar. Return migrants and permanent migrants were statistically indistinguishable in terms of age, marital status, and cash on hand at arrival. Perhaps surprisingly, there was no difference in network connections, which could have supported migrants or provided job referrals and altered return behavior. Importantly, there was also no statistically significant difference in heights, which suggests that return migrants were not differentially selected from the migrant population. We illustrate this in Figure 5. The estimated height distributions for both the return and permanent migrant samples lie nearly on top of each other.

[36] This is calculated as the total number of emigrants from 1921-1930 over the at-risk population to return home during the 1920s, given in proxy by the numbers of immigrants from 1916-1925 (since most migrants leave from the past five years). This rate formula is similar to the repatriation rate in Gould (1980) but does not correct for non-immigrants and non-emigrants due to our migrant sample only containing immigrants rather than non-immigrants. The rate is much lower than the 1910s decade rate of 47.6%, which could provide a more accurate representation of out-migration since measurement error is most severe following World War I (Bandiera, Rasul and Viarengo, 2013).

Return migrants and permanent migrants were not similar in every way. Migrants born in the Center region were more likely to be permanent migrants, while those born in the Bajio region were more likely to be return migrants. The Mexican states with the highest number of returns in the 1920s, such as Jalisco, Michoacán, and Guanajuato, are the same states today that have the highest migration and return rates. A difference between historical and current migration from these regions is that today there are robust networks in these states, while the data from 1920 reports that less than 10% of migrants from high-sending states were joining someone upon arrival.

Those who listed their intended destination as California were least likely to become return migrants. The further distance between sending states and California likely increased the costs of returning, lowering return rates (Borjas and Bratsberg, 1996). In addition, return migrants were slightly more likely to be unskilled, although the magnitude of this difference is small and only marginally significant. While the difference in occupational class suggests that return migrants were negatively self-selected on occupation, it is unknown whether occupation was intended or previous occupation, leaving their true position in the skill distribution unclear.

Estimating Selection into Return Migration

Given the prevalence of temporary and circular migration among Mexicans, we must determine the selection into return migration in order to know whether the selection into migration is a good proxy for the quality impact on the migrant stock in the United States. We return to the observation from Table 4 and Figure 5 that permanent and return migrants had similar heights (168.7 centimeters), and test whether this result holds when controlling for age and region of birth. Specifically, we pool the return and permanent migrant samples, and regress height on an indicator for whether the migrant was a return migrant. We include controls for age

bins, decade of birth, and regions of birth, equivalent to those in the specifications testing for migrant selection, and also test for whether return selection was differential within occupation.

The results of the regression of height on return migration status are presented in Panel A of Table 5. A simple correlation in the first column shows that return migrants were 0.006 centimeters shorter than permanent migrants, a statistically insignificant result. After adding age and region of birth fixed effects, return migrants' heights continue to be statistically indistinguishable. Although occupational structures upon arrival were slightly different between return and permanent migrants, once controlling for occupational structure there was still no differential self-selection of return migrants. Panel B shows alternative sample specifications for samples including only unskilled, skilled, or professional workers, and also including only people born in the North, Bajio or Center region. All regressions show no significant differences between return migrants and permanent migrants in terms of height. Overall our linked sample suggests that return migrants and permanent migrants had similar levels of human capital.

Robustness of Results for the Linked Sample

Linked samples may not be representative of their underlying populations because the links are not made randomly. Specifically, a migrant is more likely to be connected if he has a unique name, and he will not be linked if there was a death, name change, or transcription error.[37] It is likely that transcription errors and names changes were more prevalent in the United States than in Mexico. While transcription error was likely random with respect to height, name changes could have occurred more often for migrants intending to reside permanently in the United States. If those migrants were more adept at English or at acquiring United States

[37] It is worth noting that when we compare permanent migrants to return migrants, we are comparing a linked sample to a linked sample so any bias in estimation from nonrandom links would result from errors differentially occurring in Mexico and the United States

specific human capital, then our linked sample would underestimate the quality of permanent migrants.

Another concern is that mortality might bias results if taller individuals are healthier and likely live longer. We restrict the sample to migrants arriving under the age of 40 who are less likely to die within ten years. The results, shown in Column (5) of Panel A of Table 5, indicate that even with the restricted sample we continue to find no differential self-selection into return migration.

Lastly, it is possible that households in Mexico report migrants in the United States as members of the household to enumerators. This error would imply that links to both the United States and Mexican censuses were actually people who resided in the United States. We include 169 migrants who were uniquely linked to both the United States Census and the Mexican Census in our sample of permanent migrants and regress height on return migrant status. The result reported in Column (6) in Panel A of Table 5 confirms that return migrants defined in this manner were not differentially self-selected from the population.

Conclusions

Despite the prominence of self-selection in the migration literature, the literature is relatively silent as to whether Mexican migrants in the early twentieth century had more or less human capital than those they left behind. Further, little is known about individual migrants who returned back to Mexico. Using hand-collected data from border crossing manifests in 1920, we link our sample to the 1930 United States Census and 1930 Mexican Census to determine the characteristics of permanent and return migrants. The border crossing manifests include the migrant's height that we use as a metric to compare the quality of individuals because, unlike wages or occupation, it did not change when a migrant crossed the border. Furthermore, height

provides useful variation within occupation that allows us to go further than other historical selection studies by comparing heights within an occupational class.

In the early twentieth century, the United States labor market drew the taller workers from Mexico. Mexican migrants were over four centimeters taller than members of the Mexican military and only one and a half centimeters shorter than passport holders. Since the military drew from lower classes of Mexican society and the passport holders were elite, the fact that migrants were taller than those in the military and nearly as tall as passport holders implies that they were positively selected on height from the Mexican population. Within skill class, the United States drew the taller workers, an especially important result since there is little variation in occupation for Mexican migrants. This positive self-selection represents a "quality drain" or "productivity drain" from Mexico to the United States. This result holds through a number of different robustness checks.

From our linked sample we find that return migrants were not differentially self-selected. Since return migrants were not differentially self-selected from the migrant population, the measured quality of the inflow is a good proxy for the quality of the migrant stock in the United States in the early twentieth century. Taller individuals migrated from Mexico, and migrants observed years later in the United States were just as tall. Time and return migration did not impact the quality drain on Mexico or the quality gain to the United States that resulted from positive self-selection into migration from Mexico. However, return rates were high during this time period with between 13 and 44 percent of migrants returning home after ten years.

This paper adds to the growing literature on the self-selection of migrants historically, shifting the focus from Europe to Mexico, an increasingly important source of labor in the early twentieth century. While Europeans faced heavy institutional constraints following the

imposition of quota laws in the early 1920s, Mexican labor migrated relatively freely across the border, uninhibited by a limit on the total number of migrants. As a result, migration results represent self-selection in an environment with relatively few barriers to entry. The fact that Mexican migrants were positively self-selected is consistent with Borjas (1987) and Chiquiar and Hanson (2005) where migrants had high costs of travel or faced credit constraints, limiting the ability of lower quality Mexicans to migrate.

A pattern of positive self-selection of Mexican migrants affects both Mexico and the United States in a variety of ways. The United States received the most productive Mexican workers, and these workers would assimilate into the labor market more quickly than negatively selected migrants. For Mexico, the taller laborers, the taller miners, and the taller farmers left Mexico to work in the United States, draining Mexico of human capital and lowering the productivity of the average Mexican worker. However, the total effect on Mexican development is unclear as migration not only affects labor markets, but also can influence home country savings and investment by increasing remittances. It further affects the home country by changing political institutions if migrants return back home, by increasing technological diffusion with the transmission of techniques or capital goods across borders, or by influencing future migration with the strengthening of networks. This study provides a first step to better understanding the various effects of historical migration from Mexico the United States on the economies of both nations.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." NBER Working Paper No. 18011, 2012b.

———. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102, no. 5 (2012a): 1832-1856.

A'Hearn, Brian, Jörg Baten, and Dorothee Crayen. "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital." *The Journal of Economic History* 69, no. 3 (2009): 783-808.

Angelucci, Manuela. "US Border Enforcement and the Net Flow of Mexican Illegal Migration." *Economic Development and Cultural Change* 60, no. 2 (2012): 311-357.

Bandiera, O., I. Rasul, and M. Viarengo. "The Making of Modern America: Migratory Flows in the Age of Mass Migration." *Journal of Development Economics* 102, (2013): 23-27.

Bloch, Louis. "Facts About Mexican Immigration Before and Since the Quota Restriction Laws." *Journal of the American Statistical Association* 24, no. 165 (1929): 50-60.

Bodenhorn, Howard, Timothy Guinnane, and Thomas Mroz. "Problems of Sample-Selection Bias in the Historical Heights Literature: A Theoretical and Econometric Analysis." Yale Economics Department Working Paper No. 114, 2013.

Borjas, George J. "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics* 3, no. 4 (1985): 463-489.

———. "Self-Selection and the Earnings of Immigrants." *The American Economic Review* 77, no. 4 (1987): 531-553.

Borjas, George J. and Bernt Bratsberg. "Who Leaves? The Out-Migration of the Foreign-Born." *Review of Economic and Statistics* 78, no. 1 (1996): 165-176.

Cardenas, Gilberto. "United States Immigration Policy Toward Mexico: An Historical Perspective." *Chicano Law Review* 2 (1975): 66-91.

Cardoso, Lawrence A. "Labor Emigration to the Southwest, 1916 to 1920: Mexican Attitudes and Policy." *The Southwestern Historical Quarterly* 79, no. 4 (1976): 400-416.

———. *Mexican Emigration to the United States, 1897-1931: Socio-Economic Patterns*. Tuscon: University of Arizona Press, 1980.

Case, Anne and Christina Paxson. "Stature and Status: Height, Ability, and Labor Market Outcomes." *Journal of Political Economy* 116, no. 3 (2008): 499-532.

Chiquiar, Daniel and Gordon H. Hanson. "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." *Journal of Political Economy* 113, no. 2 (2005): 239-281.

Chiswick, Barry R. "The Effect of Americanization on the Earnings of Foreign-Born Men." *The Journal of Political Economy* 86, no. 5 (1978): 897-921.

Clark, Victor S. "Department of Commerce and Labor." *Bulletin of the Bureau of Labor. Volume XVI* (1908).

Coatsworth, John H. *Growth Against Development: The Economic Impact of Railroads in Porfirian Mexico.* DeKalb: Northern Illinois University Press, 1981.

Cohn, Raymond L. *Mass Migration Under Sail: European Immigration to the Antebellum United States.* Cambridge: Cambridge University Press, 2009.

Collins, William J. and Marianne H. Wanamaker. "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data." *American Economic Journal: Applied Economics* 6, no. 1 (2014): 220-252.

Ettinger, Patrick W. *Imaginary Lines: Border Enforcement and the Origins of Undocumented Immigration*. Austin: University of Texas Press, 2009.

Feliciano, Zadia M. "The Skill and Economic Performance of Mexican Immigrants from 1910 to 1990." *Explorations in Economic History* 38, no. 3 (2001): 386-409.

Ferrie, Joseph P. "A New Sample of Males Linked from the Public use Microdata Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census Manuscript Schedules." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 29, no. 4 (1996): 141-156.

———. *Yankeys Now: Immigrants in the Antebellum US 1840-1860*. Oxford: Oxford University Press, 1999.

Foerster, Robert F. and United States. *The Racial Problems Involved in Immigration from Latin America and the West Indies to the United States: A Report Submitted to the Secretary of Labor*. Washington: Government Printing Office, 1925.

Gartner, Scott Sigmund, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright. *Historical Statistics of the United States: Earliest Times to the Present*. Edited by Susan B. Carter. New York: Cambridge University Press, 2006.

Gibson, John and David McKenzie. "Eight Questions about Brain Drain." *The Journal of Economic Perspectives* 25, no. 3 (2011): 107-128.

Goldin, Claudia. "The Political Economy of Immigration Restriction in the United States, 1890 to 1921." In *The Regulated Economy: A Historical Approach to Political Economy*, pp. 223-258. Chicago: University of Chicago Press, 1994.

Gould, John D. "European Inter-Continental Emigration. the Road Home: Return Migration from the USA." *Journal of European Economic History* 9, no. 1 (1980): 41-112.

Gratton, Brian, and Emily Merchant. "Immigration, Repatriation, and Deportation: The Mexican-Origin Population in the United States, 1920–1950." *International Migration Review* 47, no. 4 (2013): 944-975.

Gratton, Brian, and Myron P. Gutmann. "Hispanics in the United States, 1850–1990: Estimates of Population Size and National Origin." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 33, no. 3 (2000): 137-153.

Gutiérrez, David G. *Walls and Mirrors: Mexican Americans, Mexican Immigrants, and the Politics of Ethnicity* University of California Press, 1995.

Gutmann, Myron P., Robert McCaa, Rodolfo Gutiérrez Montes, and Brian Gratton. "Los efectos demográficos de la Revolución Mexicana en Estados Unidos." *Historia Mexicana* 50, no. 1 (2000): 145-165.

Hatton, Timothy J. and Jeffrey G. Williamson. "International Migration in the Long Run: Positive Selection, Negative Selection, and Policy." In *Labor Mobility and the World Economy*, pp.1-31. Springer Berlin Heidelberg, 2006.

Humphries, Jane and Timothy Leunig. "Was Dick Whittington Taller than those He Left Behind? Anthropometric Measures, Migration and the Quality of Life in Early Nineteenth Century London?" *Explorations in Economic History* 46, no. 1 (2009): 120-131.

Knight, Alan. *The Mexican Revolution*. Cambridge Latin American Studies. Vol. 54-55. Cambridge; New York: Cambridge University Press, 1986.

Lindstrom, David P. 1996. "Economic Opportunity in Mexico and Return Migration from the United States." *Demography* 33 (3): 357-374.

López-Alonso, Moramay. "Growth with Inequality: Living Standards in Mexico, 1850–1950." *Journal of Latin American Studies* 39, no. 1 (2007): 81-105.

———. Height of Mexican Men and Women from Military Passport Records, 1870-1950. [Computer File]. ICPSR version. Mexico City, Mexico: Secretaria de la Defensa (SDN) [producer]. Ann Arbor: Inter-University Consortium for Political and Social Research [distributor], 2003.

———. *Measuring Up: A History of Living Standards in Mexico, 1850-1950*. Stanford University Press, 2012.

———. “The Military Option: Health, Nutrition and Living Conditions of Mexican Soldiers.” Unpublished Manuscript, 2000.

López-Alonso, Moramay and Raúl Porras Condey. "The Ups and Downs of Mexican Economic Growth: The Biological Standard of Living and Inequality, 1870–1950." *Economics & Human Biology* 1, no. 2 (2003): 169-186.

McKenzie, David and Hillel Rapoport. "Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico." *Journal of Development Economics* 84, no. 1 (2007): 1-24.

———. "Self-Selection Patterns in Mexico-US Migration: The Role of Migration Networks." *The Review of Economics and Statistics* 92, no. 4 (2010): 811-821.

Mesnard, Alice. "Temporary Migration and Capital Market Imperfections." *Oxford Economic Papers* 56, no. 2 (2004): 242-262.

Moraga, Jesus Fernandez-Huertas. "New Evidence on Emigrant Selection." *The Review of Economics and Statistics* 93, no. 1 (2011): 72-96.

Nekeoi, Arash. "Immigrants' Labor Supply and Exchange Rate Volatility." *American Economic Journal: Applied Economics* 5, no. 4 (2013): 144-164.

Ngai, Mae M. "The Strange Career of the Illegal Alien: Immigration Restriction and Deportation Policy in the United States, 1924-1965." *Law and History Review* 21, no. 1 (2002): 69-108.

Persico, Nicola, Andrew Postlewaite, and Dan Silverman. "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height." *Journal of Political Economy* 112, no. 5 (2004): 1019-1053.

Rockoff, Hugh. "Until it's Over, Over there: The US Economy in World War I." NBER Working Paper No. w10580, 2004.

Ruggles, Steven, Trent Alexander, Katie Genadek, Ronald Goeken, M. Schroeder, and Matthew Sobek. *Integrated Public Use Microdata Series (IPUMS): Version 5.0* [Machine-Readable Database]. Minneapolis: University of Minnesota, 2010.

Schick, Andreas and Richard Steckel. "Height as a Proxy for Cognitive and Non-Cognitive Ability." NBER Working Paper No. 16570, 2010.

Schultz, T. Paul. "Wage Gains Associated with Height as a Form of Health Human Capital." *American Economic Review* 92, no. 2 (2002): 349-353.

Spencer, Elizabeth A., Paul N. Appleby, Gwyneth K. Davey, and Timothy J. Key. "Validity of Self-Reported Height and Weight in 4808 EPIC–Oxford Participants." *Public Health Nutrition* 5, no. 4 (2002): 561-565.

Steckel, Richard H. "Heights and Human Welfare: Recent Developments and New Directions." *Explorations in Economic History* 46, no. 1 (2009): 1-23.

———. "Stature and the Standard of Living." *Journal of Economic Literature* 33, no. 4 (1995): 1903-1940.

Stolz, Yvonne and Joerg Baten. "Brain Drain in the Age of Mass Migration: Does Relative Inequality Explain Migrant Selectivity?" *Explorations in Economic History* 49, no. 2 (2012): 205-220.

Thomas, Duncan and John Strauss. "Health and Wages: Evidence on Men and Women in Urban Brazil." *Journal of Econometrics* 77, no. 1 (1997): 159-185.

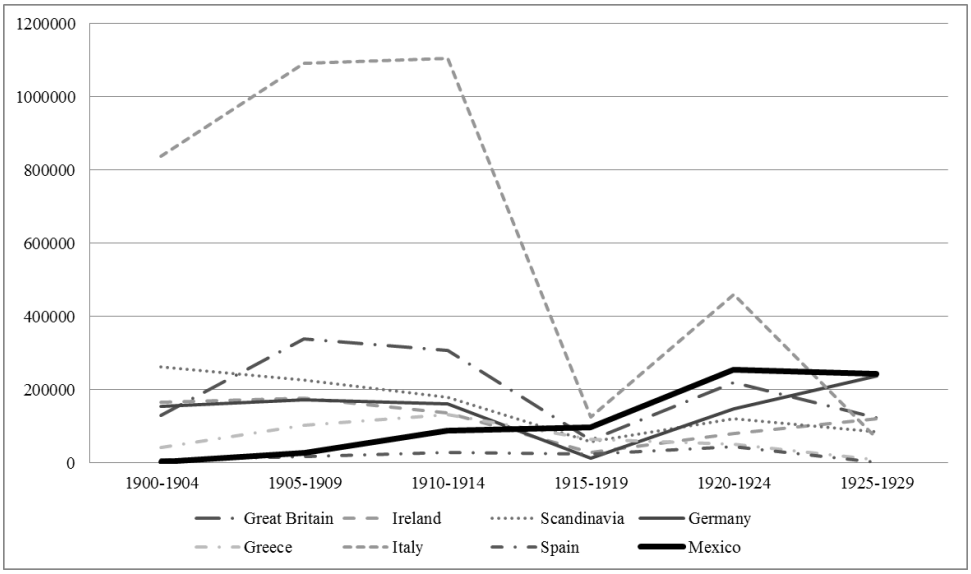
United States Bureau of Immigration. "Annual Report of the Commissioner General of Immigration to the Secretary of Labor." Multiple Years, 1908-1930.

Ward, Zachary. "Birds of Passage: Return Migrants, Self-Selection and Immigration Quotas." Working Paper, 2013.

Zeidel, Robert F. *Immigrants, Progressives, and Exclusion Politics: The Dillingham Commission, 1900-1927*. DeKalb: Northern Illinois University Press, 2004.

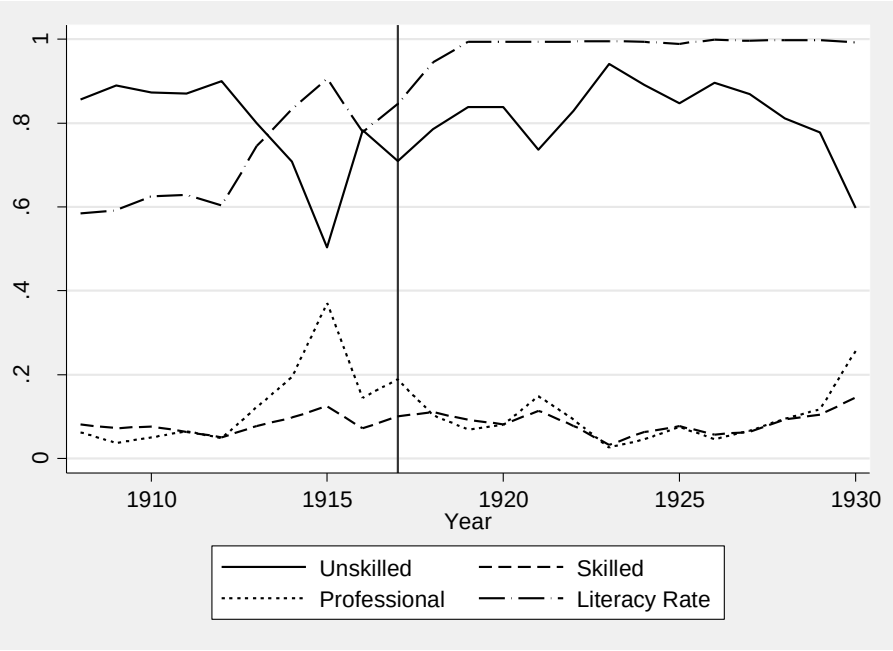
Figures and Tables

Figure 1 – Immigrant Flows to the United States, 1900-1929



Notes: Immigrant flows are aggregated in five year bins.
Source: Historical Statistics of the United States (Carter et al., 2006)

Figure 2 – Skill Composition and Literacy Rate of Mexican Migrants, 1908-1930



Notes: Skill classifications according to López-Alonso (2000). The vertical line at 1917 represents the year of the literacy requirement.
Source: Annual Reports of the Commissioner General of Immigration 1908-1930.

Figure 3 – Location of Border Stations

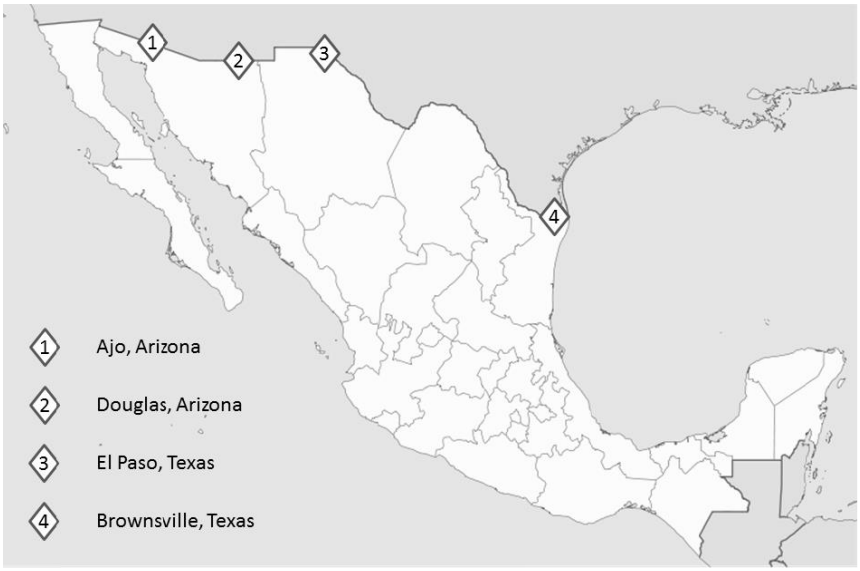
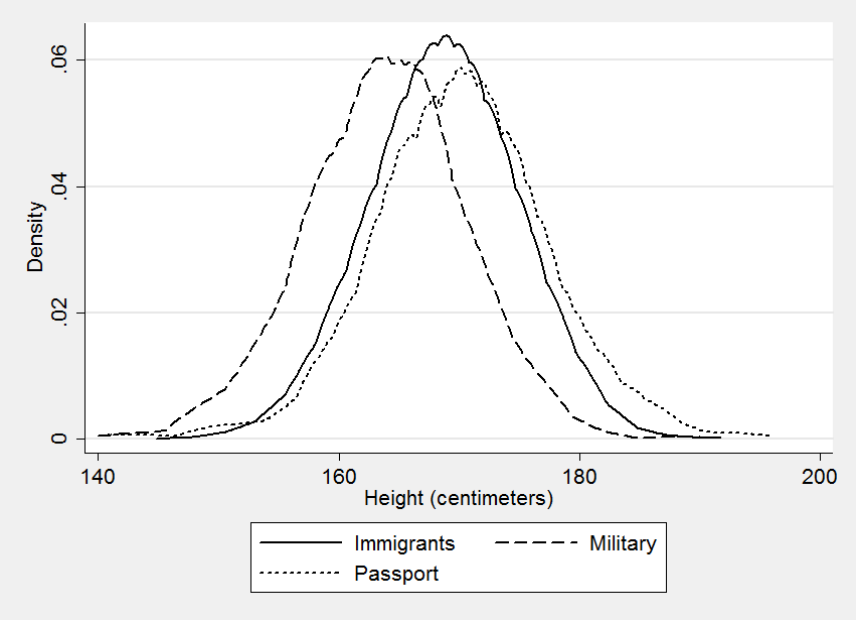
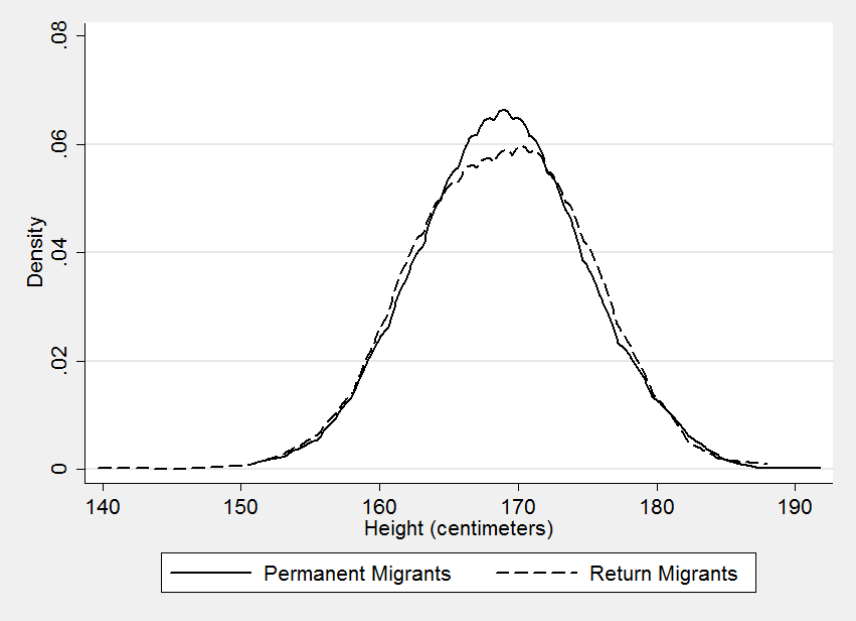


Figure 4 – Heights: Immigrants, Soldiers and Passport Applicants



Notes: Observations below 140 cm in height are dropped, although results are unchanged if they are included.
Source: Migrant heights from borders crossing manifests; Soldier and passport applicant heights from López-Alonso(2003).

Figure 5 – Heights: Permanent and Return Migrants



Notes: Observations below 140 cm in height are dropped, although results are unchanged if they are included. Permanent migrants are those migrants linked to the 1930 US Census and return migrants are those linked to the 1930 Mexican Census.
Source: Migrant heights from border crossing manifests.

Table 1 – Summary Statistics for Migrant, Military, and Passport Samples

Variable	Migrant Sample	Military Sample	Passport Sample
Height (centimeters)	168.66	163.83	170.15
	(6.09)	(6.72)	(7.30)
Age at Arrival (years)	27.86	28.37	38.63
	(9.63)	(7.64)	(10.14)
Unskilled	0.87	0.77	
	(0.33)	(0.42)	
Skilled	0.10	0.21	
	(0.29)	(0.41)	
Professional	0.03	0.02	
	(0.17)	(0.13)	
Literate	0.99		
	(0.07)		
Married	0.49		
	(0.50)		
Single	0.48		
	(0.50)		
Widowed	0.02		
	(0.15)		
Headed to California	0.07		
	(0.26)		
Headed to Texas	0.81		
	(0.39)		
Headed to Arizona	0.08		
	(0.27)		
North	0.22	0.19	
	(0.41)	(0.39)	
Bajio	0.75	0.30	
	(0.43)	(0.46)	
Center	0.03	0.40	
	(0.16)	(0.49)	
South	0.00	0.11	
	(0.04)	(0.32)	
Meeting No One	0.86		
	(0.34)		
Meeting Friend	0.01		
	(0.10)		
Meeting Relative	0.13		
	(0.33)		
Cash on Hand ($)	38.73		
	(300.00)		
Observations	3,671	3,884	1,249

Notes: Standard deviations in parentheses. Proportions unless otherwise noted.
Source: Migrant data from border crossing manifests; Military and passport data from López-Alonso (2003).

Table 2 – Selection Regressions Comparing Migrants to the Military and Passport Samples

	(1)	(2)	(3)	(4)	(5)	(6)
Comparison Sample:	Military	Military	Military	Military	Passport	Passport
Migrant	4.831***	5.062***	4.118***	4.160***	-1.484***	-1.432***
	(0.147)	(0.157)	(0.191)	(0.192)	(0.230)	(0.268)
Age, 18-20 years		-2.682***	-2.593***	-2.529***		-0.568
		(0.291)	(0.284)	(0.284)		(0.553)
Age, 21-23 years		-0.337	-0.278	-0.240		0.200
		(0.211)	(0.210)	(0.209)		(0.286)
Decade of Birth, 1850		0.352	-0.445	-0.511		-0.857
		(0.917)	(0.886)	(0.830)		(1.288)
Decade of Birth, 1860		-1.368**	-1.329**	-1.498**		-0.225
		(0.628)	(0.616)	(0.609)		(0.702)
Decade of Birth, 1870		-0.912**	-0.866**	-0.920***		-0.158
		(0.354)	(0.346)	(0.347)		(0.547)
Decade of Birth, 1880		-0.545*	-0.361	-0.381		0.184
		(0.279)	(0.272)	(0.271)		(0.512)
Decade of Birth, 1890		-0.533**	-0.426*	-0.439*		0.491
		(0.231)	(0.225)	(0.225)		(0.487)
Born, Center Region			1.033***	0.967***		
			(0.346)	(0.343)		
Born, Bajio Region			2.608***	2.652***		
			(0.354)	(0.352)		
Born, North Region			4.261***	4.300***		
			(0.365)	(0.364)		
Skilled				0.924***		
				(0.204)		
Professional				1.830***		
				(0.429)		
Constant	163.8***	164.7***	162.6***	162.3***	170.1***	169.9***
	(0.108)	(0.211)	(0.365)	(0.367)	(0.206)	(0.417)
Observations	7,555	7,555	7,555	7,555	4,920	4,920
R-squared	0.124	0.138	0.165	0.169	0.010	0.014

*Notes: Robust standard errors in parentheses; ***p<0.01, **p<0.05, *p<0.10.*
Source: Migrant data collected from border crossing manifests; Military and passport data from López-Alonso (2003).

Table 3 – Alternative Sample Specifications

Sample Specification	Migrant	Sample Specification	Migrant
Baseline	4.160***	Only Literate	3.356***
	(0.192)		(0.379)
Only North Region	4.265***	Only Unskilled	4.015***
	(0.335)		(0.213)
Only Bajio Region	3.740***	Only Skilled	4.294***
	(0.255)		(0.458)
Only Center Region	5.916***	Only Professional	5.274***
	(0.606)		(0.970)
Only South Region	7.638**		
	(3.638)		

*Notes: Robust standard errors in parentheses; *** p<0.01, **p<0.05, * p<0.10. The dependent variable in each regression is height. Each regression includes the full set of controls for age, location of birth, and occupation, but only the coefficient on migrant is reported. The comparison group is the military sample.*
Source: Migrant data collected from border crossing manifests; Military data from López-Alonso (2003).

Table 4 – Summary Statistics for Permanent and Return Migrants

Variables	(1) Permanent Migrants	(2) Return Migrants	(3) Difference
Height (centimeters)	168.7	168.7	-0.00623
	(5.943)	(6.126)	
Age at Arrival (years)	27.77	27.88	0.112
	(8.720)	(8.916)	
Unskilled	0.855	0.888	0.0330*
	(0.353)	(0.316)	
Skilled	0.113	0.0854	-0.0273*
	(0.317)	(0.280)	
Professional	0.0326	0.0269	-0.00568
	(0.178)	(0.162)	
Literate	0.994	0.997	0.00310
	(0.0790)	(0.0562)	
Married	0.439	0.472	0.0329
	(0.497)	(0.500)	
Single	0.543	0.509	-0.0331
	(0.498)	(0.500)	
Widowed	0.0188	0.0190	0.000190
	(0.136)	(0.137)	
Headed to California	0.0877	0.0633	-0.0244*
	(0.283)	(0.244)	
Headed to Texas	0.799	0.831	0.0312
	(0.401)	(0.375)	
Headed to Arizona	0.0739	0.0633	-0.0106
	(0.262)	(0.244)	
North	0.256	0.225	-0.0310
	(0.436)	(0.418)	
Bajio	0.707	0.764	0.0575**
	(0.456)	(0.425)	
Center	0.0376	0.00949	-0.0281***
	(0.190)	(0.0970)	
South	0.000	0.00158	0.00158
	(0.000)	(0.0398)	
Meeting No One	0.846	0.860	0.0124
	(0.361)	(0.348)	
Meeting Friend	0.00922	0.0100	0.00078
	(0.0957)	(0.0997)	
Meeting Relative	0.144	0.129	-0.0149
	(0.351)	(0.335)	
Cash on Hand ($)	29.24	34.11	4.87
	(92.26)	(183.3)	
Observations	798	632	

Notes: Standard deviations in parentheses. Proportions unless otherwise noted. Permanent migrants are those migrants linked to the 1930 US Census and return migrants are those linked to the 1930 Mexican Census.
Source: Border crossing manifests.

Table 5 – Regression Results for Return Selection

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A:						
Sample Specification	*All*	*All*	*All*	*All*	*Age at Arrival<40*	*Cross-Link*
Return Migrant	-0.00623	-0.0267	0.0586	0.0744	-0.0886	0.00594
	(0.322)	(0.323)	(0.321)	(0.321)	(0.339)	(0.303)
Decade of Birth		X	X	X	X	X
Age Bins		X	X	X	X	X
Region of Birth			X	X	X	X
Occupational Class				X	X	X
Observations	1,430	1,430	1,430	1430	1,268	1,599
Panel B:						
Sample Specification	*Unskilled*	*Skilled*	*Professional*	*North*	*Bajio*	*Center*
Return Migrant	0.0949	-0.211	0.193	0.260	0.0433	-2.046
	(0.340)	(1.128)	(1.808)	(0.673)	(0.370)	(3.864)
Decade of Birth	X	X	X	X	X	X
Age Bins	X	X	X	X	X	X
Region of Birth	X	X	X	N/A	N/A	N/A
Occupational Class	N/A	N/A	N/A	X	X	X
Observations	1,243	144	43	346	1,047	36

*Notes: Robust standard errors in parentheses; *** p<0.01, **p<0.05, * p<0.10. The dependent variable in each regression is height. Each regression has different sample specifications. Permanent migrants are those migrants linked to the 1930 US Census and return migrants are those linked to the 1930 Mexican Census.*
Source: Border crossing manifests.

Appendix

In order to create a sample of return and permanent migrants, we match our initial sample of 3,671 migrants to both the 1930 Mexican Census (MC) and 1930 United States Census (USC). We use slightly different matching criteria for each census. The match to the MC is based on first name/last name, year of birth plus/minus two years, and state of birth, while that to the USC is based on first name/last name, year of birth plus/minus two years, and country of birth. We cannot produce more precise matches based on state of birth in Mexico to the USC because it does not list the state of birth in Mexico.

We follow an iterative procedure for matching, similar to Abramitzky, Boustan, and Eriksson (2012b, web Appendix) and Ferrie (1996). The matching procedure is given here in detail.

(1) We search forward to the 1930 MC and USC using name and exact year of birth and place of birth using Ancestry.com, and collect the top three closest matches.

(2) We standardize the names of potential matches sample by using the Double Metaphone system, an algorithm that corrects for common transcription errors for foreign names.[38]

(3) If the person is linked to one individual, then we consider the individual as a *unique* within census match and stop here. If the migrant is linked to two or more individuals, we consider the individual as a *duplicate* within census match and stop here.

(4) If the individual is not matched, we expand the birth year window to plus or minus one year and repeat steps (2) and (3). If this does not yield a match, we expand the window to plus or minus two years. Any individual that is not matched within a two-year window is termed *unlinked*.

[38] Double Metaphone is intended to be an improvement upon the Soundex algorithm by correcting for more broad phonetic coding errors. We graciously thank Chris Minns and Gill Newton for providing the code to standardize names.

(5) The above process creates 429 *unique* links within both the MC and USC. However, they could be uniquely matched on different windows around the birth year (exact, plus/minus one, or plus/minus two). We allocate the more favorable link (smaller birth window) to that specific census, which moved 149 links to the USC and 258 links to the MC. This leads to 169 migrants that are *uniquely* linked to both censuses with the same name and birth year, which we term as *cross-links*. There are also 92 other cross-links that are either a duplicate match to the MC, USC, or both.

Table A1 shows the results of the matching process displayed in a matrix of unlinked, unique, and duplicate to each census. We use the 632 matched uniquely and only to the MC as our sample of "return migrants" and the 798 matched uniquely and only to the USC as our sample of "permanent migrants." In addition, there are 1,765 unlinked and 261 cross-links. The rest of the 3,671 are matched to duplicates either in the MC or the USC. The forward matching rate to unique links only within one census is 21.7% for the USC and 17.2% for the MC, similar to other countries from Abramitzky, Boustan and Eriksson (2012a).

Table A1 – Matching Matrix.

		1930 Mexican Census			
		Unlinked	Unique	Duplicate	
1930	Unlinked	1,765	**632**	96	2,493
U.S.	Unique	**798**	169	28	995
Census	Duplicate	119	53	11	183
	Total	2,682	854	135	3,671