

# Birds of Passage: Return Migration, Self-Selection, and Immigration Quotas\*

Zachary Ward<sup>†</sup>

April 2014

## Abstract

Estimating the self-selection of return migrants is key to understanding migrant assimilation and the quality of migrant stock remaining in the United States, especially since temporary migration is common. While return migrants are rarely observed, I use two unique datasets that record characteristics of return migrants in the early 1900s: one observes migrants who actually returned home and the other observes incoming migrants who planned to return home. I estimate the self-selection on quality for both actual and planned return migrants. Actual return migrants negatively self-selected into return migration based on occupation upon departure; planned return migrants positively self-selected into return migration upon arrival. I discuss different interpretations for this result. Finally, I show that migration quotas in the 1920s changed self-selection of actual return migrants from negative to positive as supply shocks made it easier for low-skilled migrants to succeed. Despite evidence showing that migrant quotas increased skills of migrants upon arrival, positive self-selection into return migration led to no effect of immigration quotas on the migrant stock. (*JEL* J61, N30)

Preliminary and Incomplete: Data collection in progress which could change results.

---

\*I would like to thank Ann Carlos for numerous conversations and insightful guidance on this project. Other helpful comments were given by Dustin Frye, Ian Keay, Edward Kosack, and Steven Smith. I would also like to thank participants at the World Cliometrics Conference and University of Colorado Economic History Workshop for helpful comments. All errors are my own.

<sup>†</sup>University of Colorado Boulder. Email: Zach.A.Ward@gmail.com

# 1 Introduction

A substantial fraction of immigrants return home. Bandiera, Rasul, and Viarengo (2013) estimate that between 1900 and 1920, the out-migrant flow was three-fifths the size of the total immigrant flow, suggesting that the average migrant was temporary rather than permanent.<sup>1</sup> While many select into temporary migration, little is known about the characteristics of these temporary migrants. This is surprising since return migration influences the impact of immigration on labor markets, the estimation of the assimilation of immigrants into the United States economy, and the magnitude of a brain-drain from the source country – some of the most important questions in the immigration literature.

In this paper I analyze who selected into temporary rather than permanent migration during the early 20th century. Depending on whether migrants were positively or negatively self-selected into return migration based on skill or quality, return migration could have reduced or increased the quality the remaining migrant stock (Borjas, 1985). The direction of selection depends on the motivation for migration in the first place. Return migrants could have intended to return home prior to arrival or they could have changed their decision from permanent to return after experiencing a negative shock in the United States. The first type of migrant is typically defined as a target saver, accumulating financial capital for investments back home, a model commonly applied to contemporary Mexican migrants (Lindstrom, 1996). This kind of return migrant could be positively or negatively selected on skill, depending on the relative returns to skill and costs of migration (Borjas and Bratsberg, 1996). The second type of return migrant, one who changed his or her decision from permanent to return, is thought to have failed in the labor market as his or her income did not meet expectations, and is likely to be negatively self-selected on skill (Abramitzky, Boustan and Eriksson, 2014). Depending on the combination between these two types, return migrants could have been positively or negatively self-selected on skill. Further, comparing self-selection of actual return migrants to self-selection of planned return migrants can uncover the role of shocks in the self-selection process.

Because of the availability of data, the early 20th century provides a unique setting for the study of return migration. I use two datasets to distinguish between migrants who actually returned home and those who *planned* to return upon arrival. For actual return migrants, the United States started to keep track of migrants leaving the country in 1908 but discontinued these statistics in 1957, creating a unique situation where there is better data historically than contemporarily. The data include out-migrant information observed

---

<sup>1</sup>Following the literature, I use temporary migration, out-migration, return migration, and emigration interchangeably.

*upon departure.* For planned return migrants, I use incoming passenger lists; the United States asked incoming migrants how long they intended to stay within the United States. I randomly sample ships from these records from 1917 to 1924.

These data yield several improvements over other methods of understanding return migration. Some current studies on return migration either track individuals across countries (e.g. Mexican Migration Project) or link migrants across countries in historical censuses, a method available for only a few countries with digitized censuses (e.g. Norway, Sweden, or Canada).<sup>2</sup> United States immigration statistics has data on return migrants to all countries rather than just a specific one; further, linking migrants from census to census cannot distinguish between target savers and failed migrants. In other studies on return migrants, out-migrants are not observed – to capture data on return migrants, these studies use indirect methods where out-migrants are residuals (Lubotsky, 2007; Abramitzky, Boustan, and Eriksson, 2014).<sup>3</sup> This paper improves on residual methods by using data that directly observes migrant’s occupation upon departure, making it possible to know precisely the distribution of skill of return migrants. Further, using data on migrants’ intentions to return provides insight into separate selection of planned and actual out-migrants.

United States immigration statistics also provide the opportunity to study how return migration responded to distinct shifts in migration policy: the 1917 literacy test and 1920s immigration quotas. The United States government started to collect out-migration records during a time of free mobility across the United States border in 1908, and continued after the implementation of the quota system in the 1920s, allowing analysis of the effect of these policies on temporary migration and ultimately the migrant stock.

The first part of this paper estimates the self-selection of return migrants from the foreign-born population in the United States.<sup>4</sup> I find that return migrants were negatively self-selected as emigrants’ occupations were less skilled on average than occupations of the total foreign-born.<sup>5</sup> This finding is consistent with indirect studies of return migrants in a similar time period (Abramitzky, Boustan, and Eriksson, 2014). However, while migrants who

---

<sup>2</sup>The 100% count censuses for these countries can be found at the North Atlantic Population Project. Papers that pursue this method include Abramitzky, Boustan and Eriksson (2012) and Kosack and Ward (2014).

<sup>3</sup>Specifically, these studies compare panel data (only permanent migrants) to repeated cross-sections (both permanent and return migrants) to uncover information on return migrants. For a discussion of using residual methods to back out estimates on return migrants, see Van Hook et al. (2006).

<sup>4</sup>Borjas (1987) initiated the extensive literature on the selection of immigrants from the source country’s population. Further research has investigated what type of individuals migrate temporarily versus permanently (Borjas and Bratsberg, 1996; Constant and Massey, 2003; Duleep, 1994; Jasso and Rosenzweig, 1988).

<sup>5</sup>A limit to this study is that I can only determine how temporary migrants and permanent migrants differ in occupations rather than within occupation. It is possible that return migrants are lower paid within an occupation, but it is impossible to determine whether or not this is true.

actually out-migrated were negatively self-selected on skill, migrants who planned to return home were positively self-selected on skill, suggesting that actual return migrants are mostly those who experienced negative shocks in the United States.<sup>6</sup>

While actual migrants were negatively self-selected overall from 1908-1932, there was heterogeneity in selection patterns both across countries and across time. Most countries (Eastern, Southern and Western Europe) exhibited negative selection as unskilled laborers constituted a large fraction of the return flow; however, other countries (East Asia and South America) exhibited positive selection as many entrepreneurs and merchants returned.<sup>7</sup> I also measure self-selection in 1910, 1920, and 1930, points before and after World War I, the imposition of the literacy requirement in 1917 and migration quotas in 1921. Migrants were negatively selected on skill in 1910, a pattern that intensified in 1920. However, by 1930, actual return migrants were positively selected.<sup>8</sup> As quotas were imposed, more unskilled migrants remained in the United States, causing the pattern of self-selection to be more positive – a result that I explore in detail in the second part of this paper.

The second part of this paper explores how positive self-selection of return migrants in 1930 influenced the stock of migrants remaining in the United States. In particular, I extend Greenwood and Ward’s (2014) finding that quotas decreased the out-migration of the unskilled more drastically than other skill groups to explore the effect on the remaining migrant stock. If enough unskilled migrants remained in the United States, it is possible that any increase in migrant skills caused by quotas was cancelled out by return migration.

First, I show that migrant quotas led to higher skilled incoming migrants, even after controlling for literacy rates and for fallout from World War I. A 60% restriction on a country’s flow increased the average skill level by approximately 3%. However, as measured in the census this effect for the flow does not continue to hold for the recently arrived migrant stock. I show the quotas caused the out-migrant flow’s skills to increase by approximately 3.6%, canceling out the gains from the incoming flow. The findings show that migration policy aimed to improve migrant assimilation worked as migrants stayed longer within the United States; however, the quotas’ attempt to increase skills ultimately failed due to unintended effects on temporary migration.

The remainder of the paper is organized as follows. Section 2 discusses return migration

---

<sup>6</sup>An alternative interpretation is that while planned return migrants were positively self-selected on occupation, they were negatively self-selected in terms of height, a metric that proxies for nutrition, intelligence and productivity (Kosack and Ward, 2014). Migrants were often employed in jobs that were different from those they did at home (Wyman, 1996), so height may be a more relevant metric for how well migrants would do in the United States. Since planned return migrants were negatively self-selected on height, it follows they would be negatively self-selected on United States occupational scores.

<sup>7</sup>Ilahi (1999) also observes that many return migrants in contemporary Pakistan are entrepreneurs.

<sup>8</sup>Intermediately selected means they have on average the same skills as the migration population.

and immigration policy during the early 20th century and Section 3 talks about different ways to model return migrants. Section 4 describes the data used for the project and Section 5 shows self-selection of return migrants on demographic and geographical characteristics. Section 6 estimates the self-selection of migrants on quality. In Section 7, I show how the quotas affect the incoming flow, out-going flow and migrant stock, and then I draw conclusions in Section 8.

## 2 Historical Background

### 2.1 Immigration to the United States

During the early decades of the Age of Mass Migration (1850-1913), rates of return migration were relatively low due to high costs of migration (Bailey, 1912). Before steamships, migrants boarded sailing ships and paid high fares to travel for months across the Atlantic (Cohn, 2009). Not only did the high ticket prices and opportunity costs deter return migration, but the trip itself was dangerous. Mortality rates on sailing ships were high, likely decreasing the incentive to return (Cohn, 1984).

Following the Civil War, shipping technology shifted from sail to steam, dramatically lowering the costs of migration. While fares dropped slightly, the more important effect of steam technology on migration was that travel time fell from two months to two weeks (Cohn, 2009). This decrease in the cost of traveling to the United States altered the selection of migrants arriving in the United States; migrants who were formerly constrained by the costs of migration could now migrate, which would lower the average quality of migrant (Chiquiar and Hanson, 2005). Further, the development of networks in the 19th century also lowered costs for migrants and subsequently the average quality of migrant (McKenzie and Rapoport, 2010; Spitzer, 2014). Indeed, Abramitzky et al. (2014) show that migrant cohort quality decreased in the later 20th century.

At the same time when the costs of migration were falling, the benefits to migration for unskilled labor were rising as the United States offered a premium in real wages over European countries (Williamson, 1995). Accordingly, the composition of migrants changed during the late 19th century, shifting from a “family” movement where individuals moved their households to an influx of young, single men who were more mobile and thus more likely to return home (Baines, 1994). Millions came from Southern and Eastern Europe to find work in relatively more industrialized United States economy, jobs that required skills that were different from the mostly agricultural jobs back home (Wyman, 1996). However, if migrants were not successful in finding good jobs, it was relatively cheap and easy to return

home.

In response to the influx of these new and lower skilled migrants, populist anger led Congress to attempt immigration restrictions, trying to either change the skill composition through a literacy test, or to limit the total number of migrants entering the country (Goldin, 1994). Part of the populist anger was directed at a perceived lack of assimilation of migrants into United States society (Lleras-Muney and Shertzer, 2013), with “birds of passage” remaining for a couple of years without attempting to learn the language, laws or culture of the United States. Some argued that these temporary migrants leaked savings out of the country, lowered living standards for natives through falling wages, and that they exacerbated the business cycle as firms overproduced commodities due to cheap labor (Bailey, 1912).

Instead of simply succumbing to political pressure and passing laws to restrict migration, Congress decided in 1907 to collect data on how migrants were affecting the economy, leading to the formation of the famous Dillingham Commission – the most extensive federal study of immigrants in the early 20th century. At the same time, Congress required the Commissioner General of Immigration to collect passenger lists of boats leaving the country, leading to the first emigration statistics, published in 1908. Interest in return migration gained ground as officials began to ask incoming migrants first whether they had been in the United States previously, and in 1917, asked migrants their intentions to return home.

Using this data on return migrants allows us to understand how important self-selection of out-migrants was to the assimilation process. Typically in the immigration literature, migrants assimilate in the labor market when their wages catch up to natives (Chiswick, 1978). However, since return migrants attrite from their cohort when they move out of the United States, a migrant cohort’s wage profile could mechanically increase just because return migrants were the lowest earning of a cohort; in other words, return migrants are negatively selected (Borjas, 1985). Thus, when a migrant cohort is “assimilated” by having the same wages as natives, it could be because all of the least able migrants left rather than the most able migrants achieving highly. Abramitzky et al. (2014) suggest that this mechanism explains approximately 45 percent of the convergence between natives and immigrants’ occupational scores during the first two decades of the 1900s.

Despite the importance of self-selection of return migrants in the assimilation process and the prevalence of return migration in the early 20th century, there is little research on why some people decide to out-migrate.<sup>9</sup> Biavaschi (2013) takes an important first step by exploring how the age, sex, and occupational composition of out-migrants changes from

---

<sup>9</sup>There are several studies written by historians on the return migration experience of specific ethnicities during the early 20th century. See Wyman (1996) for a bibliography of ethnicity-specific studies of return migration.

1908-1957 and suggests that return migrants became less negatively selected on skill over time.<sup>10</sup> I delve further into this result by both disaggregating return migrants into ethnicities in order to explore heterogeneity of the return migration experience and also use data on incoming migrants' intentions to return to study selection upon arrival. While Biavaschi (2013) explores out-migrant composition, Greenwood and Ward (2014) research how migrant quotas and World War I affect out-migration rates for European countries.<sup>11</sup> They show that unskilled migrants were much more likely to stay following migration quotas compared to higher skilled occupations, arguing that supply shocks in the migrant labor market raised the benefits to staying in the United States. I expand on their result by studying how quotas affected the overall migrant stock that remained in the United States.

## 2.2 Description of Quota Laws

The backlash against high volumes of migration led to the implementation of the literacy test in 1917; not satisfied with this qualitative restriction, Congress employed a second method of limiting immigration: quota restrictions (Zeidel, 2005). Passed in 1921, the Emergency Quota Act limited immigrants from a given country to 3% of the total foreign-born from that country as enumerated the 1910 census. The law applied only to the Eastern Hemisphere (the Western Hemisphere was excluded from quota limits) and was especially restrictive for Eastern and Southern European countries with a small migrant stock in 1910. From July 1921 to June 1922, out of a possible 356,995 quota immigrants, only 243,953 entered as many Western and Northern European countries did not fill their quotas (Report of the Commissioner General of Immigration or RCI, 1922).<sup>12</sup> The act categorized people into "quota" and "non-quota" immigrants, limiting only quota migrants after the numerical limit was reached. Non-quota migrants (children of citizens under the age of 18 and government officials) were not counted against the quotas.<sup>13</sup>

---

<sup>10</sup>This result is consistent with both the observation the incoming migrants became less positively self-selected over the 20th century and Borjas and Bratsberg's (1996) theoretical model that shows the selection of return migrants should be the opposite of the selection of incoming migrants.

<sup>11</sup>Others estimate the extent of the outflow from the population (Gould, 1980; Bandiera et al, 2013), showing high rates of return migration for "new migrants" compared to "old migrants." New migrants refer to individuals from Southern and Eastern Europe (e.g. Spain, Portugal, Italy, Russia, Hungary, etc.), while old migrants refer to individuals from Western and Northern Europe (e.g. England, Ireland, France, Germany, and Scandinavian countries)

<sup>12</sup>The total amount of immigrants admitted during the fiscal year 1922 (July 1921-June 1922) was 309,556, a sharp drop from 805,228 the previous year (Report of the Commissioner General of Immigration or RCI, 1922).

<sup>13</sup>People entering the country were also split into "immigrant" (categorized into quota and non-quota) and "non-immigrant," which included individuals traveling through the United States to another destination and people visiting the United States for business or pleasure and after the 1924 quota, included people who traveled home temporarily. Non-immigrants were not counted against the quotas. In 1925, of the

The initial quota system of 1921 was temporary, designed to be in place for a year as Congress debated over a permanent system. After extensions in 1922 and 1923 to keep quotas intact, Congress passed the Immigration Act of 1924 on May 24, 1924.<sup>14</sup> Now the quota formula was 2% of the foreign-born population from a given country based on the 1890 Census.<sup>15</sup> This change lowered the annual quota from 357,803 to 164,667, but unequally affected old migrants and new migrants. For example, Italy's quota dramatically dropped from 42,607 to 3,845 and Russia's quota fell from 24,405 to 2,248, while Germany's only decreased from 67,607 to 51,227 (RCI, 1924-1925). The policy achieved its desired effect as migrant composition shifted sharply between 1914 and 1925 from new migrants back to old migrants. Northern and Western Europe's percentage of total immigrant flows increased from 20.8% to 75.7% while Southern and Eastern Europe's dropped from 75.6% to 10.8% (RCI, 1925).

The Immigration Act of 1924 stipulated the quotas until July 1927, after which a board would alter the quotas to reflect the "national origins" of the United States based on the 1920 census. After 1929, the formulas were no longer calculated based on the number of foreign-born in the United States, but now also attempted to reflect "the origins of United States citizens."<sup>16</sup> The main reallocation of quota numbers was between Northern and Western European countries as Germany's quota fell from 51,227 to 25,957 and the United Kingdom's increased from 34,007 to 65,721. Italy's increased slightly from 3,845 to 5,802 and Russia's increased from 2,248 to 2,784 (RCI, 1930). In this study, I take advantage of three quota changes in 1921, 1924, and 1929 to estimate the effects of quotas on temporary migration and the migrant stock (development in Section 8.1).

The quotas had provisions that could affect return migration through two channels: by restricting frequent trips back and forth from the home country, and incentivizing migrants to become citizens. First, in order to travel back home temporarily, migrants had needed to apply for a reentry permit for a fee of \$3.<sup>17</sup> Initially some foreign-born were confused

---

250,912 non-quota immigrants that entered, 7,217 (2.9%) were wives and children of United States citizens, 64,632 (25.8%) were those returning after a visit abroad, and 175,069 (69.8%) were from countries in the Western Hemisphere (RCI, 1925). Other non-quota immigrants were government officials, temporary visitors, professors, ministers, government officials, and students.

<sup>14</sup>Minor adjustments to country's quotas were made between 1921 and 1924 years as officials attempted to accurately estimate the effect of political boundary changes on quotas (RCI, 1922-1924).

<sup>15</sup>The Immigration Act of 1924 formalized a completely new admission process. Now potential migrants needed to apply for immigration visas from consul offices located in their home country. Since these consuls reviewed migrant applications to verify that migrants would satisfy entry requirements, immigration stations such as Ellis Island in New York and Angel Island in San Francisco were no longer needed as screening centers.

<sup>16</sup>Congress delayed changing quota limits for two years until the final National Origins formula was implemented in 1929. The total quota number was now 150,000.

<sup>17</sup>Migrants already in the United States were allowed to leave temporarily and freely come back without restriction from the quota system. While migrants returning from a temporary visit abroad were counted

over whether or not they could return under the quota laws as the 1925 Report noted an increase in applications for temporary visit abroad “due, no doubt, to the fact that the provision of the law which authorizes their issuance is becoming more generally known to the alien population.” (RCI, 1925, pg. 13) By 1930, 114,173 people traveled back home with a reentry permit; the Commissioner General of Immigration called the reentry permit a “popular provision” of the law (RCI, 1930, pg. 25). For migrants who pursued a strategy of moving constantly back and forth across borders, quotas would inhibit mobility due to paperwork and fees, but that migration strategy was still available. Further, when leaving temporarily, the out-migrants were classified as “non-emigrants,” and returned as “non-immigrants,” and thus are not included in the data.

A second effect of the quotas on migration is that due to the law, migrants had incentive to become citizens of the United States in order to improve the likelihood of reuniting with their relatives. Preference in quotas was given to relatives of citizens and those who had applied for citizenship, including children (under age 21), parents, brothers, sisters, and fiancées.<sup>18</sup> Migrants had to file a petition for relatives to receive non-quota visas, or to receive preference under the quota – around 30,000-40,000 petitions were filed yearly between 1925 and 1930 (RCI, 1930).<sup>19</sup>

Naturalization laws required the foreign-born to live in the United States for at least five years before becoming citizens. In addition, migrants had to file a declaration of intent to become a citizen at least two years before they could become naturalized. Furthermore, applicants were required to be able to speak English, but did not have to know how to read and write in English (Naturalization Act, 1906). Rather than return home to bring their family with them like migrants would previous to 1921, migrants now had to remain in the United States in order to be naturalized – possibly a factor for why out-migration rates fell following the implementation of quotas (Greenwood and Ward, 2014).

### 3 Return Migration Theory

Return migration defies simple income maximization models since individuals return to poor countries (Sjastaad, 1962), but even the foundational work of Ravenstein (1885) observed

---

against the quota from 1921 to 1924, they were allowed to enter if the quota was full. This is opposed to the 1924 law which never counted returnees against the quota.

<sup>18</sup>Preference was also given to agricultural workers and their family members, but it is unclear how the preferences influenced visa issuances.

<sup>19</sup>Despite the preference, many family members could not be reunited with citizens in the United States; as such, the General Commissioner of Immigration recommended in 1925 that parents and children under the age of 21 be included as “non-quota immigrants” rather than just receive preference in the quotas (RCI, 1925, pg. 28).

that “counter-currents” of out-going migrants run in opposite direction of incoming migrants.<sup>20</sup> While one way to motivate return migration is to alter the utility function where temporary migrants prefer consumption at home, utility preferences do not predict the self-selection of return migrants with respect to skill or ability.<sup>21</sup>

Another motive for temporary migration is to save wages earned abroad in order to start a business within the home country, ultimately increasing their lifetime income at home (Ilahi, 1999; Mesnard, 2004). However, this observation has led to a debate over the effect of income on return propensity: a positive income shock would lead a migrant aiming at a savings target to return sooner, while other migrants would remain longer to take advantage of higher income (termed life-cycle model) (Yang, 2006). Evidence on which effect dominates for contemporary migrants suggests that migrant behavior is a mixture between the two models (Bijwaard and Wahba, 2014), but Greenwood and Ward (2014) suggest that migrants in the 1920s display behavior consistent with a life-cycle model since they are less likely to return and stay longer following a labor supply shock caused by quotas.<sup>22</sup>

Borjas and Bratsberg (1996) incorporate both savings return motives and the possibility of negative income shocks into the original Borjas (1987) model of migrant selection to predict which migrants would return home. Individuals decide to migrate if the benefits outweigh the costs; however, they will decide to return home if savings or experience accrued abroad increases earnings home. Those that *plan* to migrate temporarily are those whose skills are rewarded more in the destination country, but not a high enough premium to attract them permanently; thus, return migrants can be thought of as a “marginal migrant” who is just on the edge of migrating permanently or not migrating at all. Deviations from planned return migration are due to unexpected shocks in the United States, whether positive (more likely to stay) or negative (more likely to return home).

The Borjas and Bratsberg (1996) model generates a few predictions for return migration in the early 20th century. First, lowering the cost of migration and the cost of returning home leads more people to select into temporary migration. It is likely that return migrants settled in places that were easier to move back home; for example, European return migrants were more likely to settle in New York City near shipping lines rather than Illinois. Quotas

---

<sup>20</sup>See Dustmann and Weiss (2007) for an overview of some other contemporary return migration models.

<sup>21</sup>Some papers find migrants explicitly state the reason for return was because of family reasons or cultural preference (Constant and Massey, 2003; Gibson and McKenzie, 2011). Cerase (1974) suggest that people return because of relative deprivation which yields to negative self-selection, yet Piore (1979) argues that those that stay in the United States are those that have failed because they unwilling to return home a failure.

<sup>22</sup>Abarcar (2014) and Yang (2006) shows the Filipino migrant behavior is more consistent with a life-cycle model while Nekeoi (2013) shows that Mexican migration to the United States is consistent with a target-earnings model.

likely raised the costs of migration, which would lead to fewer people selecting into return migration, consistent with the finding by Greenwood and Ward (2014) that return migration rates decreased. However, depending on how the costs of migration varied across the skill distribution, the type of selection into return migration on skill is unclear (Chiquiar and Hanson, 2005; Fernandez Huertas-Moraga, 2013).

One variable that influences selection into migration and likely temporary migration are networks. Networks lower the costs of migration by covering transportation and housing for incoming migrant, suggesting that both low-skilled and temporary migrants join networks (McKenzie and Rapoport, 2010). At the same time, networks improve job prospects through connections and also provide more information to individuals making the migration decision, suggesting that these migrants were less likely to experience negative shocks after arrival. This would make the overall effect of networks on temporary migration ambiguous.

## 4 Data

I use two separate data sources to compare temporary migrants to the rest of the migrant population: one is based on temporary migrants who actually leave the United States (termed *actual* return migrants), and the other is those who plan to leave the United States upon arrival (termed *planned* return migrants).

### 4.1 Actual Emigrants: Annual Reports

I combine two data sources to compare actual return migrants to the migrant population. First, data on out-migrants are found in the Annual Report of the Commissioner General of Immigration (henceforth RCI) between 1908 and 1932, which records statistics on emigrants upon departure from the United States. When leaving the United States, ship captains had to deliver a passenger list to the port's customs agents, a manifest similar to ones for arriving immigrants.<sup>23</sup> Passenger lists for out-migrants included their name, age, country of intended residence, nationality, occupation, years within the United States, and state of last permanent residence. The Department of Immigration issued a yearly report on the status of immigration, which includes a wealth of tables tabulating data from ship manifests, and serves as the basis for Ferenzci and Willcox's (1929) expansive volume of migration data. Importantly, this data holds only aggregations of those who leave, making analysis of micro-determinants of out-migration impossible. Collection of data on out-migrants ended

---

<sup>23</sup>This collection was mandated by the Immigration Act of February 20th, 1907, the same Act that created the famous Dillingham Commission that produced 41 volumes of data on foreign-born in the United States.

in 1957, so no contemporary data exists directly recording out-migrants' characteristics.<sup>24</sup> Any current study needs to rely on demographic techniques or residual methods to estimate the flow of out-migrants (Van Hook et al., 2006).

In order to determine how these individuals selected into out-migration, I need to compare their characteristics to the migrants living in the United States. Data on the migrant population in the United States is taken from 1% IPUMS samples from 1910-1930. When comparing characteristics such as age, sex, marital status, and location, I use the full sample of foreign-born in the census. However, when comparing migrant skills (i.e., occupational scores), I drop those under the age of 16 and over the age of 65 in order to represent those in their working age.

Reflective of a time period where United States was race obsessed, the RCI records return migrants' occupations by ethnicity rather than by country, requiring a match between ethnicities in administrative data to individuals in the census. A natural method to connect the two data sources would match ethnicity in RCI to a person's country of birth; however, this is problematic as political boundaries for many Central and Eastern European countries were redrawn following World War I. Instead, I match to ethnicity based on language. I use the variable "mother's tongue" to match ethnicity to language, which provides a more detailed measure to properly identify ethnicities that political boundaries are unclear during this time period such as Hebrew, Magyar, Polish, and Slovakian.

Recent research by Bandiera, Rasul, and Viarengo (BRV, 2013) argue that these administrative records severely undercounted both immigrants and emigrants. Measurement error results from improperly compiling ship manifests and ship captains not recording cabin class passengers. Thus, the RCI data can be thought as a sample of total out-migrants, just as the IPUMS is a sample of the entire population – however, the question is whether the RCI data is representative of the return migrant population. The main reason suggested by BRV for undercounting is careless compiling of ship manifests by the Department of Immigration. Which ships were not recorded in the official tally is unclear, and it is plausible that this measurement error is random and does not affect the representativeness of the sample; however, there is no way to verify whether this is true. The other concern is whether or not cabin classes were not included in the official statistics, which would create a negative bias since the recorded out-migrants are less skilled.<sup>25</sup> However, this problem seems to apply to

---

<sup>24</sup>Because reports following 1932 do not contain information on the occupations of emigrants, I only use years 1908-1932.

<sup>25</sup>Willcox (1931) reports that only steamship passengers are included in official statistics before 1904, but changes in classification of immigrants by 1907 solve the problem by including the cabin class. The definition of an immigrant in 1908 change to one whose last permanent residence was outside the country and intends to reside within the United States for at least 12 months at a time, and thus included cabin class passengers who did not wish to be associated with the term "immigrant." (Hutchinson, 1958)

only immigrants and not emigrants because there was a concern cabin classes could avoid the \$4 head tax (RCI, 1912).

## 4.2 Planned Emigrants: Ship Manifests

In order to determine how the selection of planned temporary migrants differed from planned permanent migrants upon arrival, I collect individual data from ships arriving at Ellis Island. When arriving in the United States, ships had to provide a manifest of passengers to the port's custom agents. Originally the first passenger lists recorded sparse information for migrants where officials listed only the migrant's name, age, sex and country of origin; as the decades went and millions of migrants arrived in the United States during the Age of Mass Migration (1850-1913), the government became more interested in who was arriving at the borders. By the 20th century, the passenger lists included information such as the migrant's final destination, city of birth, height, whether they paid for their passage, and the name and relation of any contact they were joining in the United States.

Three important questions were added to these manifests by 1917: whether a migrant had been to the United States previously, whether they intended to return to their country, and if so, how long they intended to remain in the United States.<sup>26</sup> Using this information I can determine whether an individual was a repeat migrant and if they were planning to return migrant.<sup>27</sup> If migrants listed their intention to return as "uncertain" or length of stay "indeterminate," I allocate these individuals to the planned return home category.

It is important to note that these intentions to return home may have never been realized; migrants could have had a better-than-expected experience in the United States and decided to remain permanently. Thus these migrants would not be an *actual* return migrant. It is also possible that those who intended to remain permanently could have experienced a negative wage shock and then decided to return home.

I randomly sample 1% of ships that originated in Europe and arrived in Ellis Island from 1917 to 1924.<sup>28</sup> While the sample is under construction, I currently have a sample of 47 ships covering 20,060 passengers. To maintain consistency with United States definitions of immigrants and emigrants, I drop those who are traveling temporarily through the United States, had previously been in the United States in the same year, and those that were listed but did not actually embark on the trip. I also drop those whose ethnicity were not legible, leading to a sample of 13,860 incoming migrants.

---

<sup>26</sup>The question whether migrants had been to the United States previously was added in 1907 while the other two were added in 1917.

<sup>27</sup>Return and repeat migrants are not mutually exclusive categories.

<sup>28</sup>I randomly sample the ships from the Statue of Liberty Ellis-Island Foundation.

## 5 Descriptive Statistics

### 5.1 Actual Emigrants

Who decided to leave the United States? Out-migrants have similar characteristics to immigrants, where they are mostly male, “new migrants”, and hold an unskilled job (Biavaschi, 2013). While Biavaschi (2013) reports statistics for the overall out-migrant flow, I disaggregate her statistics by ethnicity and also show other attribute of selection to include geographic, marital and length of stay.<sup>29</sup>

Table 1 shows the demographic characteristics of return migrants for years 1908-1932, separated by ethnicity. Most out-migrants were between the ages of 16-45, with few children and old migrants returning. Return migrant flows were heavily male (on average 75%), even more skewed towards men than the immigrant flow. The Irish and out-migrants to the West Indies come close to gender parity (51% male), but other ethnicities, especially Eastern European, have percentage of males over 80%. One deviation from the characteristics of typical return migrants are Chinese return migrants, where only 34% of out-migrants were between 16 and 45. This is likely because many young Chinese immigrants to the United States were restricted by the Chinese Exclusion Act of 1882, which only left older migrants who were in the United States before that policy.

While one may expect that a higher percentage of males are reflective of a heavy entry and return of young single males, in reality a significant fraction of return migrants were married.<sup>30</sup> For many ethnicities, over half of those leaving were married, reaching as high as 85% for Romanian emigrants. This could be because married individuals are more likely to return to their families, or because single individuals marry while they are in the United States and leave with their family. To determine which is more likely, if one assumes that each married female leaves with a married male, 70% of married males return by themselves, indicating that many return migrants were married in the first place.<sup>31</sup>

A unique facet of the data are statistics on temporary migrant’s duration of stay. A substantial fraction of migrants stayed a short amount of time, with most returning after less than five years in the United States. Research that treats return migrants as residuals that drop out of the sample between decadal census dates miss temporary migrants that arrive and leave between censuses. Interestingly, some new migrants remain longer in the United States than old migrants. For example, Greeks, Russians, and Italians, those with

---

<sup>29</sup>To make the comparisons between the migrant stock and out-migrant flow, Biavaschi (2013) drops those who have stayed longer than ten years, while I keep those over ten years and reweight the individual data to match the length of stay of each ethnicity.

<sup>30</sup>Married statistics for emigrants are only reported for years after 1918.

<sup>31</sup>Author’s calculations from RCI data.

high return rates, have about 50% of those who stay for less than five years, while French, German, and English migrants have over 65% that stay for less than five years. The highest percentage of those staying less than five years is for countries in close proximity (i.e. Spanish Americans, Cubans, and Mexicans, all over 75%), suggesting that costs of migration had an effect on selection into temporary migration. Conversely, the lowest share of those returning less than five years is the Chinese, where only 25% have stayed less than five years.

Out-migrants mostly returned from northeastern states, with the Midwest being the second highest, except for ethnicities that locate mostly in the South (e.g., Cubans) or locate in the West (e.g. Chinese, Japanese or Mexican). The Southwest and West were relatively more labor scarce than the Northeast, so perhaps migrants found land and wages more attractive to settle than in the East. Also, the further away migrants were from the East coast would raise the cost of returning home; returning home from New York City was relatively easier than returning home from Wyoming.

The demographic information sheds light on the typical out-migrant, but comparing out-migrants to the migrant population living in the United States shows self-selection based on demographics and location. However, comparing all out-migrants to all stayers may provide misleading inferences since a large part of the migrant stock is not similar to temporary migrants. The most striking and consistent difference between out-migrants and stayers is the duration of stay as out-migrants resided a short period of time before returning. Over 50% of out-migrants stayed less than five years, and over 90% stayed less than ten years. I reweight individual IPUMS observations to match the distribution of length of stay for the out-migrant flow in order to control for length of stay. For example, 70.8% of German out-migrants stayed less than five years, but only 10.9% of Germans observed in the census stayed less than five years. I reweight the German IPUMS observations to match the 70.8% for temporary migrants by placing a higher weight on recent migrants and lower weight for long-established migrants.<sup>32</sup>

Table 2 breaks down differences between out-migrants and those in the IPUMS (1910-1930). Positive numbers indicate out-migrant's mean was higher than foreign-born. Out-migrants differed from the migrant population significantly in terms of age gender, marriage, length of stay, and geographical location.

First, out-migrants were much more likely to be male than the foreign-born in the United States. Every ethnicity shows a higher out-going male percentage than those remaining in the United States. Second, evidence on age and marital status disputes the story that out-migrants were more likely to be young and single than the rest of the foreign-born population.

---

<sup>32</sup>Using different weights like re-weighting on sex and years in the United States, or re-weighting on age does not affect results qualitatively.

Out-migrants were 8.5 percentage points less likely to be younger, under the age of 16.<sup>33</sup> This is relatively unsurprising because many temporary migrants were working and migrated while in working age; families tended to remain in the United States. The more interesting result is for older ages: Overall return migrants were only slightly likely to be between ages 16 and 45 (1.2 percentage points), but much more likely to be above the age of 45 (7.3 percentage points). Selection into return migration between ages 16 and 45 was varied across ethnicity – for many, return migrants are concentrated in these ages, such as Hebrew, French and Russians. At the same time, other ethnicities have return migrants have a smaller share of those between 16 and 45, such as Bulgarians, Romanians, and Greeks. Almost every ethnicity has the share of return migrants above the age of 45 higher than the share of those within the United States. This observation is consistent with Cerase’s interviews with Italian return migrants in the 1960s, where he records that migrants would often repeat the phrase “American is not a land for old people.” Unfortunately, I cannot precisely estimate at which age migrants tend to return home because I do not have individual data.

Return migrants also had a higher share of married individuals leaving than those within the United States. This is partly due to return migrants being slightly older than those staying in the United States, but this cannot explain the entire 20.5 percentage point differential in married rates. It is possible that migrants that were married intended to return home the entire time, to bring back savings to their family; however, it is also possible that married migrants moved individually to the States and intended to bring the rest of their family with them if they succeeded. The high rate of those leaving that are married suggests that maybe these are the migrants who failed.

The share of return migrants locating in the Northeast was 10.3 percentage points higher than the foreign-born, suggesting that return rates were highest amongst places like New York City and Boston. Almost every ethnicity had a higher return rate from the Northeast except Koreans and East Indians. This is evidence in support that temporary migrants were more likely to locate where the costs of return migration were cheaper. Ethnicities that had to travel across the Atlantic were much less likely to leave from the West; ethnicities that had to travel across the Pacific were much less likely to leave from the East. However, not every ethnicity conforms to this pattern – Japanese were more likely to leave from the East.

Historical patterns of settlements, such as the Dutch, Finnish and Scandinavians in Michigan, Wisconsin and Minnesota, display patterns of less return migration from these areas. Dutch return migrants had a share from the Midwest that was 16 percentage points fewer than the Dutch foreign-born – Scandinavians 14 percentage points fewer and Finnish 6.3

---

<sup>33</sup>Technically reports from 1908-1917 break into age categories of 0-14, 14-45 and above 45 while after 1917 the reports change to 0-16, 16-45 and above 45. I simplify to use to term under 16 for ease of reading.

percentage points less likely. However other ethnicities that traditionally settled in the Northeast, such as Hebrew, Italian, and Russian, had higher fractions of those leaving the Northeast than in the census. More analysis would be needed to separate the effects of ethnic enclaves and locations close to New York.

## 5.2 Planned Emigrants

Actual return migrants were more likely to be slightly older, male, “new,” married and from the Northeast; however, these characteristics may be different from those planning to return upon arrival. Following Borjas and Bratsberg (1996), I attribute any difference between planned and actual return migrants to unexpected shocks in the United States, perhaps due to income being lower than anticipated or homesickness.<sup>34</sup>

In Table 3, I show the difference in descriptive statistics between return and permanent migrants for the years 1917-1924. There are a total of 1,585 migrants who intended to return and 12,275 that intended to be permanent, yielding a planned return rate of 11.4%, much lower than actual return rates suggested by Bandiera, Rasul and Viarengo (2013). The higher actual return rate than planned return rate suggests that many out-migrants returned home due to negative shocks in the United States.

There are many differences in characteristics between those that planned to return home versus those that actually returned home. First, planned return migrants were less likely to be male than permanent migrants, a stark contrast to data on actual return migrants which shows that males were more likely to return. Two explanations of this difference are that either males experienced worse than expected outcomes or females experienced better than expected outcomes. Planned returners were much less likely to be married, which also contrasts with actual returners who are more likely to be married. It is possible that these two facts are consistent if planned returners get married in the United States and then return married; or, it is also possible that married people traveling alone were more likely to return back home due to homesickness. Planned returners were also less likely to bring children and more likely to be traveling alone, as measured by similar surnames on the manifest.<sup>35</sup>

The correlation between networks and selection into return migration is unclear because networks could lower costs leading to a temporary trip being more profitable; however, maybe networks provide benefits to the individual that entice people to migrate permanently. The share of planned migrants joining a family member upon arrival was less than those that

---

<sup>34</sup>It is also possible that there is a shock to expected income when returning home, such as better than anticipated recovery following World War I or unexpected death of parent.

<sup>35</sup>I group migrants into families if they are traveling with others with the same last name; children are those with the same last name that are under the age of 16.

planned to migrate permanently; networks appear to encourage permanent settlement in the United States. However, precise causality is not given here so it is possible that networks are located in areas that migrants were less likely to return home.

An important aspect of “birds of passage” is that they would continuously move back and forth across the Atlantic; anecdotal evidence talks of some migrants moving back and forth over 15 times (Wyman, 1996). About 14% of migrants who arrived in between 1917 and 1924 said they had previously been in the United States, which is similar to the number reported in the early 1900s Annual Reports. Planned return migrants were more likely to have been in the United States previously than planned permanent migrants, perhaps because they were more familiar with short-term trips to increase savings.

Another difference between actual and planned return migrants is locational residence – actual return migrants were more likely to leave from the Northeast and less likely to leave from the West. Planned return migrants exhibit the opposite direction where planned return migrants were more likely to leave if locating in the West and no more or less likely to leave if locating in the Northeast. Additionally, planned return migrants were more likely like big cities as their destination.<sup>36</sup> Finally, initially it appears that planned migrants report lower-skilled occupations by 4.4% and are slightly shorter than planned permanent migrants. However, this is a result I will explore into further detail in the upcoming sections.

### **5.3 Shocks after Arrival: Return Rates of Actual and Planned Out-Migrants**

We have already seen discrepancies in characteristics between actual and planned out-migrants, especially with respect to age, location choice, ethnicity and marital status. Further, the predicted return rate for migrants arriving between 1917 and 1924 is 11.4%, lower than Bandiera et al.’s suggestion that temporary migration was fairly high. Any difference between actual return migrants and planned return migrants is due to shocks after arrival – somehow the decision was altered. It is possible that planned return migrants decide to stay after arriving due to a positive income shock in the United States; however, it is also possible that planned permanent migrants decide to return after arrival due to a negative income shock. The high rates suggested by Bandiera et al. versus a much lower planned rate suggests that migrants, on average, experienced negative shocks within the United States.

To estimate how often shocks changed decisions after arrival, I compare the actual return rate for those arriving in the United States between fiscal years 1917 and 1925 to the planned

---

<sup>36</sup>Urban areas are defined as those with more than 2,500 residents according to the 1920 census, and big cities are those that have more than 100,000 residents.

return rate.<sup>37</sup> In order to do this, I need to estimate the return rate for those who arrival between 1917 and 1925. Information in the RCI gives how many leave the United States in year, but does not precisely give which year the migrant entered the United States. For example, of the 5,715 Germans that left the United States in the fiscal year 1922, it is unclear how many of these emigrants arrived in 1920 or arrived in 1910. However, the reports do list how many years these emigrants stayed within the United States – of the 5,715 German emigrants, 837 stayed less than five years (implying they arrived between 1917-1921) while 3,034 stayed between five and ten years (implying they arrived between 1912-1916). This information could be used to help proxy the return rate between 1917-1921.

I use information from the annual reports to estimate two different return rates for the cohort that entered between fiscal years 1917-1925. The first return rate is the best approximation for the true return rate; the second return rate is a lower bound estimation on the first rate. Both return rates follow the basic premise of adding up all the return migrants who entered the United States between 1917 and 1924 and dividing them by the number of migrants who entered between fiscal years 1917 and 1925.

First, the lower bound rate makes the most conservative estimates of which emigrants entered between 1917 and 1925, specifically aiming to estimate the lowest possible return rate. To do this, I sum all of the emigrants who claimed less than five years of stay from 1922 to 1926 because they definitely entered the United States between 1917 and 1925. To this number I add all emigrants who claimed staying five to ten years from the 1927 to 1931 reports, because these individuals also definitely entered the United States between 1917 and 1925. Equation 1 shows the calculation where  $e$  is an ethnicity:

$$\text{Lower Bound Return Rate}_{e,1917-1925} = \frac{\text{Emigrants Less 5}_{e,1922-1926} + \text{Emigrants 5 to 10}_{e,1927-1931}}{\text{Immigrants}_{e,1917-1925}} \quad (1)$$

This underestimates the true return rate since it does not excludes many emigrants who likely arrived between 1917 and 1925. For example, probably a significant fraction who left in 1921 and claimed staying less than five years actually arrived between 1917 and 1920; however, this rate assumes that all of these migrants who left in 1921 arrived in 1916. This is a strong assumption that I relax in order to get a calculation likely closer to the true return rate. Specifically, I assume that the average emigrant who claimed staying less than five years actually only stayed two years; those who claimed to stay between 5 and 10 years actually stayed seven years.<sup>38</sup> This expands the number of emigrant I include in the numerator. Now

---

<sup>37</sup>Fiscal years were from June-July, so data on calendar years 1917-1924 would be for fiscal years 1917-1925.

<sup>38</sup>The main reason for assuming two and seven years is because I have data on reports until 1932 – any

I sum up those who claimed staying less than five years for report years 1919 to 1927 and those staying five to ten years for reports 1924 to 1932.

$$ActualReturnRate_{e,1917-1925} = \frac{EmigrantsLess5_{e,1919-1927} + Emigrants5to10_{e,1924-1932}}{Immigrants_{e,1917-1925}} \quad (2)$$

Equation (2) is my estimate of the actual emigration rate. It should be noted that both of these estimated return rates are likely downwardly biased due to Bandiera, Rasul, and Viarengo’s (2013) evidence that out-migrants were undercounted.

I compare the actual return rate and lower-bound return rate to the planned return rate from migrants upon arrival to determine how much shocks influenced return migration. The results are presented in Table 4, split by ethnicity. The table is ordered by planned return rates from lowest to highest. The data shows the ethnicities that had the lowest planned return rates, including Hebrews, where only 1.2% planned to return, and Poles, where only 2.1% planned to return. On the opposite end were “old” migrants, such as Irish, who had a 29.9% planned return rate, and French who had a 38.2% return rate. Once again, this goes against conventional wisdom the new migrants had higher return rates.

Column (2) shows that planned return rates deviated from actual return rates – Column (3) shows the difference. Hebrews had low planned return rates (1.2%) and also low actual return rates (0.7%), suggesting that Jewish migrants experienced outcomes that aligned with expectations – perhaps even slightly better than expected. On the other hand, Bohemian migrants expected a return rate of 1.7% but had an actual return rate of 31.7%, a large discrepancy. It is likely that Bohemians did not experience good outcomes in the United States, and quickly decided to return home. This is a pattern common across “new” migrants, where actual return rates were higher than expected return rates. The 1920s was a time when many in the United States were xenophobic and discriminated against certain migrants following World War I, leading to higher return rates than expected.

Old migrants seemed to have experienced better than expected outcomes in the United States. Scandinavians, French, and Dutch all expected to return at higher rates, but had low actual rates of return. One may expect that the discrepancy between expected and actual return rates would narrow as networks expanded and expectations are more closely aligned to reality, but the data does not suggest this. However, it is also possible that the estimated return rates do not approximate the true migration rate well, because of undercounting of migrants.

The difference between actual and expected rates are likely due to shocks to income after  


---

 assumption that is higher than two and seven go into data that I do not have.

arrival – now I explore in detail the self-selection of migrants based on quality.

## 6 Self-Selection of Return Migrants on Quality

### 6.1 Measures of Migrant Quality: Occupational Scores and Height

In order to determine whether return migrants were “better” or “worse” than permanent migrants, I use two separate metrics to capture quality. First, I use occupational scores to measure the self-selection of planned and actual migrants, which is standard amongst other historical self-selection papers (Abramitzky, Boustan, and Eriksson, 2012, 2014; Collins and Wanamaker, 2014). Ideally, one would compare wages instead of occupations, but administrative data only records the occupations of returnees.<sup>39</sup> Lacking individual-specific wages, I assign an occupational score to each occupation to reflect its earnings, where all individuals claiming an occupation receive the same score. Accordingly, self-selection estimates are based on how temporary and permanent migrants differ on the occupational ladder. One may suggest that actual return migrants have lower wages within occupation, but this is impossible to determine with the given data.

Second, for the planned migrant micro-data, I am able to use an alternative individual-level measure of quality that is recorded on incoming manifests: the migrant’s height. Height is positively correlated with wages, nutrition, intelligence and strength, all of which are important measures of quality (Steckel, 2009). Also, height has favorable attributes for measuring selection of migrants such as being constant across borders and having a standard measurement, unlike occupation, education or wages (Kosack and Ward, 2014). Further, I am able to explore whether planned return migrants have lower heights, even when controlling for occupation, unlike occupational scores.

### 6.2 Details on Occupational Scores

Occupational scores should reflect earnings of the foreign-born between 1908 and 1932, but there is no single data source that provides sufficient detail across occupations during this time period. Rather than using the IPUMS variable “occscore,” which reflects occupational earnings of the entire population (native and foreign-born) in 1950 and is commonly used in other historical self-selection papers, I create an occupational score based on wages reported in the 1940 census. There was significant wage compression between the early and mid-

---

<sup>39</sup>Furthermore, wages are likewise not recorded in the census for stayers. 1940 was the first year that the census recorded wages, but they did not record the year of arrival of migrants, which makes comparing return and permanent migrants impossible.

twentieth century (Goldin and Margo, 1992), so using measures further from the early 20th century would positively bias self-selection estimates as a large number of out-migrants are laborers.

Similar in spirit to Collins and Wanamaker (2014), I use income and wage data in the 1% 1940 IPUMS sample to assign each occupation the mean wage based purely on migrants’ earnings in 1940.<sup>40</sup> This has the advantage of being closer in time to period of study (1908-1932), and also reflects migrants’ earnings rather than natives. Furthermore, I vary occupational scores by new or old migrant, assigned each the mean wage within an occupation for that group.<sup>41</sup> A disadvantage to this method is that self-employed earnings are not reported, so I drop those with zero reported income when assigning occupational earnings. This is particularly important due to farmer’s earnings being unreported. However, farmers make up a small portion of return migrant occupations (less than 4%), so estimates are not strongly sensitive to this restriction.

One of the goals of this paper is to determine the indirect method of estimating self-selection is valid; in other words, comparing my estimates of self-selection to results from Abramitzky, Boustan, and Eriksson (2014). They use occupational scores based on the IPUMS variable “occscore” partly because they also compare migrant to native earnings – I provide results using this measure in the appendix.

## 7 Estimation of Self-Selection

### 7.1 Self-Selection of Actual Emigrants on Skill

I estimate the self-selection of out-migrants by comparing return and permanent migrants mean occupational score for a given ethnicity. If the measure is positive, return migrants were positively selected and held higher paying occupations than those staying in the United States. The measure is given in the following equation

$$Selection_{jt} = \overline{\log(ReturnMigOccScore_{jt})} - \overline{\log(ForeignBornOccScore_{jt})} \quad (3)$$

where  $j$  is the ethnicity of the migrant and  $t$  is the year of observation. In the following section I estimate the self-selection of migrants across the entire 1908-1932 time period, but also in

---

<sup>40</sup>Collins and Wanamaker (2014) create another occupational score based on the Historical Statistics of the United States (Margo, 1996) and Lebergott’s (1964) data on earnings between 1900 and 1928, a more suitable time period for my study. However, this data is based on very broad industry categories which cannot be applied to the RCI data since industry is not recorded for laborers.

<sup>41</sup>Generally, I assign based on new migrants (Western and Northern Europe) and old migrants (Eastern and Southern Europe, South America, Asia and Africa). More specific earnings could be based on country of birth, but the 1940s 1% sample does not have enough observations to fill occupation-country cells.

1910, 1920, and 1930. An advantage of estimating the self-selection of return migrants is that occupations are compared within a single economy (i.e., the United States). Other selection papers observe movers in the host economy and stayers in the home economy and can only compare the two by predicting earnings of movers across economies, essentially attempting to correct for skill price differences. My measure of self-selection is more straightforward since both returnees and foreign-born are observed within the United States economy where prices for skills are the same.

The data lacks individual-level observations for out-migrants but only records the total number of occupations leaving for an ethnicity. Therefore, I cannot estimate the self-selection of out-migrants by individually controlling for age, sex or location. While it is informative to make comparisons after controlling for these characteristics, the self-selection of all return migrants still gives the overall effect of which migrants leave and is arguably what policy makers care the most about. One this I do control for is length of stay within the United States, which I do so by reweighting observations in IPUMS to match RCI – the same method I use for Table 2.

Estimates for the self-selection of migrants on occupation are presented in Figure 2. There is no consistent pattern of negative selection where return migrants are always those that fail in the labor market. English, Spanish, Finnish and Mexican are a few of the ethnicities where temporary migrants have higher skills than those remaining in the United States. However, a majority of ethnicities have temporary migrants returning with lower skills than stayers, especially ethnicities from Eastern and Southern Europe. Eastern Europeans are particularly negatively self-selected on skill, where Russian, Greeks, Romanians, Polish, Magyar and Bulgarian ethnicities are strongly negatively selected, earning about 5-15% less than permanent migrants. An overwhelming majority of temporary migrants from these ethnicities were unskilled, where most of the returnees are farm laborers, laborers or miners, while those in the census are machine operatives and semi-skilled workers.

## 7.2 Self-Selection of Planned Emigrants on Skill and Height

Did planned emigrants exhibit the same self-selection patterns on skill as actual emigrants? I estimate how migrants who planned to leave compare to migrants who planned to stay on both skill and height. I estimate the correlation of individual attributes on out-migration using a linear probability model where the dependent variable is one if a migrant plants to return and zero if the migrant plans to stay permanently.

$$PlannedReturn_i = \beta_0 + \beta_1 Log(OccupationalScore)_i + \beta_2 Height_i + X_i' \delta + \epsilon_i \quad (4)$$

Now that I have individual data, I am able to control for age, sex, and other characteristics that provide more insight into how migrants self-select into return migration. In addition to ethnicity fixed effects, I control for time trends using fiscal year fixed effects.<sup>42</sup> I also include a control for quotas because selection processes likely differ before and after new migration restrictions are put into place (Greenwood and Ward, 2014).

The results of Equation 4 are presented in Table 6. The first three columns only include controls for ethnicity and year fixed effects without any individual characteristics in order to mimic results for actual migrants. Before continuing on to discuss self-selection into planned return migration, I should note that the coefficient for the effect of quotas on self-selection into return migration is negative, corroborating the results in Greenwood and Ward (2014) that return migration decreased following migration quotas

The first column estimates the effect of a higher occupational score on planning to return home. The results show that a 1% increase in occupation score (.01 increase in log occupational score) results in a .01% higher likelihood of returning, showing that migrants *positively* self-selected into planned return migration. This is a statistically significant effect but is questionably economically significant; yet it is informative that migrants did not negatively self-select into planned migration based on occupational score. This is at odds with the results that actual return migrants were overall negatively self-selected on occupation, perhaps because of negative shocks after arrival for the lower skilled.

An alternative explanation is that the occupations reported upon arrival are different than those reported upon departure because migrants are referring to different economies. Occupations reported upon arrival should reflect occupations in the home country, while occupations upon departure are occupations in the United States. It is possible that those leaving the United States did not have skills that translated well across economies; nevertheless, it suggests that temporary migrants were drawn from the upper end of the home country's skill distribution.

The story changes when we alter the metric of quality from occupational scores to height. While migrants were positively self-selected into planned return migration based on occupational score, they were simultaneously negatively selected on height. It is possible that height more truly reflects the migrant's underlying ability to perform in the United States, and thus negative self-selection on height should correlate more strongly with negative self-selection of actual migrants. Column (3) adds occupational scores and height in the same regression and the results still hold. One possible interpretation that migrants were the lower skilled of those within an occupation – while laborers were less likely to be selected into return migration

---

<sup>42</sup>Fiscal years (July-June) fixed effects are used to coincide with the Reports of the Commissioner General of Immigration.

upon arrival, the lower quality laborers (in terms of height) were more likely to self-select into return migration. Column (4) adds in all individual controls and the results remain the same.

Table 7 shows results of the same regression specification in Equation 4, except I replace the dependent variable with the number years return migrants planned to stay. Just as the quotas caused less people to select into return migration, there is evidence that quotas led to people to plan longer stays in the United States (Greenwood and Ward, 2014). The results for quality show that those with higher occupational scores select into a shorter migration period, where a 1% increase in occupational score leads to 0.04 less years in the United States. Height has no correlation with planned length of stay. When all controls are added in Column (4), then there is no differential self-selection on either occupational score or height into planned length of stay.

### 7.3 Self-Selection in 1910, 1920 and 1930

How did the self-selection of return migrants change from 1910 to 1930? I use the RCI data from 1911, 1921 and 1931 fiscal years (for example, the 1911 fiscal year is out-migrants leaving from July 1910 to June 1911) to compare return migrants to those immediately previously recorded in the 1910, 1920 and 1930 census. Migrants returning in the fiscal year following the decadal census are drawn precisely from the distribution just observed in the census, increasing the precision of self-selection estimates (Moraga, 2011). Further, it allows analysis how self-selection changes after World War I, the 1917 literacy test and the imposition of immigration quotas in 1921, 1924 and 1929. However, a drawback is that self-selection is measured at different points in the business cycle, which could bias results if occupations have different propensities to return over the business cycle (Jerome, 1926). Similar to prior results, the estimation is done by matching the demographics of out-migrants to individuals in IPUMS.

I report estimates of the self-selection of return migrants for years 1910, 1920 and 1930 in Table 5. The average (weighted by number of out-migrants) of the self-selection of return migrants was negative in 1910, where out-migrants earned about 2% less than stayers. Negative self-selection intensified in 1920 as out-migrants earned 6.4% less than permanent migrants. However, by 1930 out-migrants were *positively* self-selected on occupational score. All ethnicities self-selection are listed in Table 5.

A particularly interesting pattern is the self-selection over time for “old” and “new” migrants, displayed in Figure 3. For both 1910 and 1920, newer migrants were more negatively self-selected than old migrants. The pattern of negative self-selection intensified from 1910

to 1920, which could be because those migrants who intended to leave following World War I were unable to, and these were lower skilled migrants. While the 1920 result is partially contaminated by World War I, it is interesting that both old and new migrants were similarly affected. The pattern of new migrants were more negatively self-selected than old migrants completely reverses by 1930 where both old and new migrants were positively self-selected and new migrants were more positively self-selected than old migrants. While this result is a natural corollary to Greenwood and Ward's (2014) result that quotas caused the lowest skilled's return rates to fall the most, it is a new finding that the self-selection for both migrants is now *positive*. If the self-selection is positive enough, it is possible that return migration actually lowered the skills of the migrant stock remaining in the United States, the opposite direction than is commonly assumed and found by the literature (Abramitzky, Boustan and Eriksson, 2014; Borjas, 1985; Lubotsky, 2007). To uncover the overall effects of quotas and self-selection of return migrants on the migrant stock, I turn to analyze the effect of quotas on those in the census.

## 8 The Effect of Quotas on the Migrant Stock

In this section, I attempt to compare two different observations to discover how quotas affected the migrant stock. First, quotas appear to have increased the skills of incoming migrants, although this has not yet been verified for all migrants but only those traveling to the United States through Canada (Massey, 2012). Hatton and Williamson (2005) argue that the unskilled were less able to enter the United States after the quota laws, but only show how broad occupational classifications change from 1920 to 1930. However, Hatton and Williamson do not control for literacy rates, something that could change occupational skills. I extend both of these papers results to include all migrants in the United States and to also control for the effects of the literacy test, making sure to separate the effects of two different migration policies.

The second observation is a combination of Greenwood and Ward's (2014) result that out-migration decreased most amongst unskilled groups and the finding of the first part of this paper that the self-selection of out-migrants in 1930 was positive. If the effect of quotas on out-migration was strong enough and the positive self-selection of return migrations was large enough, the skills of migrant stock remaining would fall.

This effect on the migrant stock is exactly what appears to have occurred in Figure 4, which shows the skills of the recent migrant stock ( $\leq 10$  years) over time separated by old and new migrants. Roughly speaking, old migrants were slightly restricted by quotas while new migrants were strongly affected. Skills of all migrants increased over time as

expected by economic growth; further, old and new migrant’s occupational scores trended together over time. The difference between the two scores is mostly stable between 1900 and 1920 where new migrants earn approximately 3% less than new migrants, but by 1930 the difference turns sharply downward where new migrants earn 8% less than old migrants. Old migrants’ occupational scores sharply increase from 1920 to 1930 while new migrants’ scores slightly increase, exactly the opposite effect if quotas indeed increased the skills of entrants as suggested by Hatton and Williamson (2005) and Massey (2012).

I use a regression framework to motivate the association between quota implementation of the change in migrant stock’s skills. I pool 1900-1930 IPUMS samples and run a difference-in-difference regression where one difference is before and after the quota policy change, and the other difference is between new and old migrants.

$$\text{Log}(\text{OccScore}_{ijt}) = \beta_0 + \beta_1 \text{New}_{ij} \times \text{Year1930}_t + \delta' X_i + \nu_j + \varphi_t + \epsilon_{ijt} \quad (5)$$

where  $i$  is the individual,  $j$  is the country of birth, and  $t$  is the census year of observation.  $\beta_1$  is the parameter of interest, which estimates the difference in earnings between new and old immigrants in 1930. Country ( $\nu_j$ ) and year ( $\varphi_t$ ) fixed effects are included.  $X_i$  includes individual characteristics such as literacy, and years in the United States fixed effects. Standard errors are clustered by the country of birth. This regression is descriptive rather than causal, showing how changes in the migrant stock from 1900-1930 were associated with the quotas. Further evidence will be given in later sections to argue causality of  $\beta_1$ .

The results of Equation (5) are presented in Table 8. The main point of the table is the *lack of* increase in migrant’s skills by 1930, which seems to contradict evidence presented by Hatton and Williamson (2005) and Massey (2012). The first column is a barebones regression only controlling for decade, years in the United States, and ethnicity. There appears to be no effect of quotas on new migrant’s occupational scores. The next column includes literacy controls to account for changes after the 1920 law and there is still no effect of the quotas on increasing the skills of the migrant stock. Finally, it is possible that occupational scores do not increase because of an increase in the fraction of females following quotas, which could bring down the skills of the overall migrant stock. In Column (3) I restrict the sample to only males and find that despite changes in the sex composition, the skills of male migrants still do not increase following migrant quotas.

## 8.1 Measuring Changes in Quotas

To estimate the effect of quotas on the migrant stock, I use the continuous measure of quota restriction that was pioneered by Greenwood and Ward (2014). The measure is based on the variation that the acts of 1921, 1924, and 1929 cause in the degree of the restrictiveness of the quotas on different countries. The Emergency Quota Act of 1921 assigned immigration quotas to be 3% of the number of people from a country based on the 1910 Census. The 1924 Immigration Act further reduced the quota 2% of the foreign-born population based on the 1890 census, substantially reducing immigrants from countries that did not start coming en masse to the United States until after 1890. Finally, the quotas slightly changed in 1929 to more favor English immigrants in order to reflect the “national origins” of the country. The measure is describe as:

$$QuotaRestriction_{j,t} = \frac{AvgImmigrants_{j,1908-1914} - Quota_{j,t}}{AvgImmigrants_{j,1908-1914}} \quad (6)$$

where  $Quota_{j,t}$  is the number that the quota stipulates for country  $j$  in year  $t$ , and  $AvgImmigrants_{j,1908-1914}$  is the average number of immigrants that arrived yearly between 1908-1914 for country  $j$ . For the years prior to the implementation of the quota, the value of  $QuotaRestriction_{j,t}$  is 0. Note that the variable  $QuotaRestriction_{j,t}$  varies both across ethnicity and across time, and ranges between 0 and 1, where a higher number indicates a larger restriction. Ethnicities that are not subject to quotas remain at zero for the time period, while other ethnicities, such as Italian, reach highs of 0.98, or a 98% restriction on the total number of migrants. The average value for quota restriction after 1922 is around 0.60 or 60%.

## 8.2 Effects on Immigrants upon Arrival and Immigrants in Census

Other have found that quotas increased the skills of migrants upon arrival (Hatton and Williamson, 2005; Massey, 2012), which I verify and expand by taking into account more ethnicities from the Eastern Hemisphere. The increase in skills is commonly interpreted as due to an increase in costs of migration following quotas. Potential migrants now had to apply for a visa in order to migrate and its possible that individuals in countries with very lower quota numbers had to bribe officials in order to make it to the United States.

Since I am comparing RCI data to census data, I collapse the census data into ethnicity/year cells. However, since the census does not capture all ethnicities by year (for example, Koreans in 1916) and thus is an unbalanced panel. To make the data comparable to RCI, I drop ethnicities that do not have observations in every ethnicity/year cell and create a

balanced panel of 464 observations (21 ethnicities by 22 years). The ethnicities dropped are mostly Asian and small European ethnicities.

Figure 5 shows the evolution of the skills of immigrants between 1908-1930 after removing literacy rates and year and ethnicity fixed effects. Prior to the 1917, new and old immigrant countries have similar trends for occupational scores. The shaded area are times during World War I, when migration was interrupted and costs of traveling increased. The vertical lines following World War I show the imposition of the 1921 and 1924 quotas. Subsequent to the quotas, the new migrant occupational scores increase while old migrants decrease, suggesting that quotas caused an increase in skills of incoming migrants.

I proceed to use a regression framework to estimate the effect of the quotas on immigrant’s occupation scores *upon arrival*. Using a panel of ethnicities across years 1908-1930, I use a difference in difference framework where treatment ( $Treat_j$ ) are different indicators for the 1921 quota, 1924 quota, and the quota restrictiveness measure developed by Equation (6).

$$\text{Log}(\text{OccScore}_{jt}) = \alpha_0 + \alpha_1 \text{New}_j \times \text{Treat}_t + \delta' X_{jt} + \nu_j + \varphi_t + \epsilon_{jt} \quad (7)$$

Equation (7) estimates the effects of the quota on immigrant occupational scores. In addition to ethnicity ( $\nu_j$ ) and country fixed effects ( $\varphi_t$ ), I include controls for World War I death rates the literacy rates of immigrant migrants in  $X_{jt}$ . I verify that the empirical strategy is valid in a subsequent section by testing for pre-treatment trend differences. Further, I take advantage of the three changes in the “dose” of quotas, which more properly identify the effects of quotas than a blunt dummy variable for treated and non-treated.

The results of Equation 7 are presented in Panel A of Table 9. The results are consistent with the previous literature where quotas increased immigrant skills upon arrival, as the increase was 2.2% after the 1921 quota, or 5.0% after the 1924 quotas. When using both treatment variables, it is apparent that both quotas increased occupation skills, with the more restrictive 1924 quotas having a larger magnitude than the 1921 quota. Column (4) shows that restricting migrant flows by 60% (average quota restriction) increased occupational scores of immigrants upon arrival by 3.6%. While a priori it is unclear that a quota would increase skills, the 1920s quotas clearly did. Importantly, these results hold after controlling for literacy rates of the incoming migrant cohort. The coefficient on literacy rate is approximately .40, suggesting that a 1% increase in literacy rate (0.01 increase) increase occupational scores by .4%.

Migrant skills upon arrival clearly increased, but what happened to a cohort’s skill when it is observed years later? Differences between the two reflect both assimilation rates in the United States and attrition of return migrants. Using census information on year of arrival,

I estimate the effects of the quota on cohort quality for the cohort that survives until 1930.<sup>43</sup> Since 1930 is years after cohort’s initial arrival, many migrants have returned home, creating an attrition bias. If return migrants have lower skills than the rest of the cohort (negatively selected), skills observed at 1930 should increase; if return migrants are positively selected, then the surviving cohort will have lower skills.

Figure 6 shows how cohort quality evolves over time from 1908-1930 (year of cohort arrival on x-axis), after correcting for place of birth, literacy and years in the United States. Note that this figure is based on pooling 1910, 1920, and 1930 samples. The trends for new and old migrants are similar prior to the literacy exam, but then high quota line drops sharply following the implementation of 1921 and 1924 quotas.<sup>44</sup>

I estimate the effect of quotas on new and old immigrants using the following estimating equation, which is very similar to Equation (7) but using IPUMS micro-data. However, to make the regressions between migrant cohort upon arrival and at census similar, I collapse IPUMS by year of arrival and ethnicity.

$$\text{Log}(\text{OccScore}_{jt}) = \alpha_0 + \alpha_1 \text{New}_j \times \text{Treat}_t + \delta' X_{jt} + \nu_j + \varphi_t + \epsilon_{jt} \quad (8)$$

The results are in Panel B for Table 9. There is no evidence that quotas increased the skills of the migrant stock remaining in the United States. Column (1) shows that there was no effect of quotas on skill when measured after the 1921 quota, similar to Column (2). Column (3) includes both quotas and while the direction on the coefficient is similar to the direction for incoming migrants where the 1921 quota coefficient is negative and the 1924 coefficient is positive, both coefficients are statistically insignificant. Finally, Column (4) shows that using the more precise measure of quota restriction has zero effect on a cohort’s occupational score. The differences in results from Panel A and Panel B suggest that new and old migrants are experiencing differential changes between time of arrival and time at census during the 1920s. I will argue in the next two sections that temporary migration changes between new and old migrants can explain much of the difference.

### 8.3 Effects on Actual Emigrants

Why did quotas increase cohort skills upon arrival but not by the 1930 census? One explanation for the lack of increase in skills over time is a differential change in return migration behavior. If unskilled new migrants were much more likely to remain following quotas, then

---

<sup>43</sup>The census asks how many years one has been in the United States, which I use to construct Year of Arrival = Year – Years in the U.S.

<sup>44</sup>I remove those staying zero years from the graph because the full cohort is not observed by the time of the census.

the cohort skills will remain low over time rather than (artificially) increasing as return migrants leave the cohort. This is the opposite direction of Borjas's (1985) argument that negatively selected return migration positively bias wages for a given cohort; I argue that return migration was more positively self-selected after the quotas (more positive for new migrants relative to old migrants), which keeps a cohort's skill low over time.

There is evidence that quotas led to fewer low-skilled return migrants. Greenwood and Ward (2014) show that the out-migration rate of the unskilled decreases for selected European countries while out-migration rates for semi-skilled and professional workers are constant. However the magnitude of the effects are unclear, and the result needs to be extended to more countries affected by quotas rather than 12 European countries. Figure 7 plots out-migration scores across time for new and old migrants (correcting for fraction of male out-migrants). Occupational scores mostly trend together until after the implementation of the 1921 quotas, where the high quota line separates and is higher than the low quota countries.

I use the same method to estimate the effect of quotas on out-migrants found in Equation (7) but using out-migrant occupational scores as the dependent variable. Table 10 shows the results of the regression using out-migrants' occupational scores. Unlike immigrants where cohort skills increase after the 1924 quota, the effect on out-migrant occurs immediately after 1921 where high quota's out-migrant occupational scores increase by 3.6%. This effect does not hold past the 1924 quota. However, using the more precise quota restriction measure, the overall effect of restricting a country's immigrant flow by 60% is an increase of out-migrant scores by 3.6%.

Importantly, the effect of quotas on out-migration occupational scores was *larger* than the effect on incoming immigrant's occupational scores. Indeed, the coefficient on quota restriction for out-migrant occupational scores is 0.0609, while the effect on immigrant upon arrival's occupational scores is 0.0510. If one subtracts the two, you get -0.010, which is a very rough approximation of the effect on the census. The coefficient for immigrants remaining in the census is -0.009, very close to the rough approximation.

One caveat is that measures for out-migrants are based on year of departure, not for the year of a cohort. Out-migrants leaving a given year are a mix of temporary migrants across cohorts, but with the data it is impossible to nail down the effect of quotas on temporary migrants for a given cohort.

## 8.4 Increase Fraction of Permanent Migrants

While an increase in out-migration occupational scores would lead to a drop in cohort’s skills, it is unclear that the number of those remaining is large enough to bring down the occupational score of the migrant stock. Return migration rates need to be large in order to have significant effects on the entire cohort. If temporary migrants make a small fraction of the incoming migrant cohort, the positive return selection would not strongly bias the overall effect.

In this section I demonstrate that quotas led to a large increase in the number who remain in the United States, a big enough effect to cause cohort skills to drop. Greenwood and Ward (2014) and Biavaschi (2013) show that the quotas lead migrants to stay longer within the United States, and Greenwood and Ward (2014) show the out-migration rates fall after the quota is implemented, but this is just for a handful of European countries. I combine census and RCI data to estimate the magnitude of those who remain in the United States.

Equation (9) gives a measure to estimate how many migrants remain from a given ethnicity’s cohort until the 1930 census. The measure is the ratio of those who “survive” until the census over those initially counted as incoming from the RCI data.

$$FractionStay_{jt} = \frac{\sum_i ArrivedinYear(Census)_{ijt}}{ArrivedinYear(RCI)_{jt}} \quad (9)$$

where individuals in the census as summed over ethnicity  $j$  (matched to country), and cohort’s year of arrival  $t$ .<sup>45</sup> Fraction Stay is top-coded at one, and cohorts where no immigrant is measured in a given year (e.g. mostly World War I years when immigration flows were low) are dropped. Only males are included for the numerator and denominator.

Figure 8 shows the evolution of the difference in Fraction Stay between new and old migrants from 1908-1930. Prior to the implementation of migration restrictions, new migrants had a higher return rate than old migrants (as demonstrated by negative values), consistent with the literature (Gould, 1980; Bandiera et al., 2013). The end of World War I shows a large fraction of new migrants staying within the United States until the aftermath of the War clears; however, this short change in return migration behavior reverts by the year 1920. There is not strong break between new and old migrants for the 1921 law, but rather the difference only increases slightly. There is clear jump after the 1924 quota laws where new migrants remained at higher rates than old migrants.

I estimate the magnitude of the effect using Equation (10), which regresses  $\text{Log}(FractionStay_{jt})$  on different treatment measures for quota restriction ( $Treat_t$ ), using year ( $\varphi_t$ ) and ethnicity

---

<sup>45</sup>Individuals are weighted by person weights to get an appropriate sum from the IPUMS samples.

( $\nu_j$ ) fixed effects. I use cohorts year of arrival 1908-1929 for the regression.<sup>46</sup>

$$\text{Log}(\text{FractionStay}_{jt}) = \beta_0 + \beta_1 \text{New}_j \times \text{Treat}_t + \nu_j + \varphi_t + \epsilon_{jt} \quad (10)$$

Table 11 shows the results of the above regression. Consistent with Figure 8, there was no significant difference between new and old migrant countries after the 1921 quota, but the effect appears after the severe restrictions in 1924. The coefficient is both statistically and economically significant, where new migrants are 0.235 log points more likely to stay. The third column separates 1921 and 1924 quota effects to determine which is more important. The fourth column uses the Quota Restriction measure created in Equation (6). Multiplying the average quota restriction (0.60) by the coefficient yields that reducing immigrant flows by 60% leads to 14% more individuals remaining.

The quotas caused enough migrants to remain in the United States until the 1930 census to cause the cohort skills to drop. Those that out-migrated were on average higher skilled, mostly because the unskilled remained in the United States. Selective attrition of the highest skilled migrants led to a drop in cohort’s skills measured in the 1930 census. The quotas initial effect of increasing the skill of incoming migrants was completely reverse due to unintended consequence of keeping the least skilled within the United States.

## 8.5 Robustness Checks

### 8.5.1 Validity of Empirical Strategy

To convincingly show that a difference-in-difference in strategy is causal, there must be no unobservable differences that vary over time between new and old migrants. One way to show this is by verifying that before quotas are implemented, new migrants and old migrant’s skills trend together. While visual evidence from Figures 5, 6, and 7 suggests that new migrants and old migrants’ skills trend closely together prior to migration restrictions, I continue to use placebo treatment effects in a regression framework to argue that the empirical strategy is valid. More specifically, I use data on pre-treatment years (before the first 1921 quota) and vary the treatment year from 1909-1921 to determine if an effect of “quota” would show up prior to 1921. If my empirical strategy is valid, there should not be an effect of “placebo” quotas on skills of migrants. I run placebo tests in three separate samples: one using RCI data for immigrant skills upon arrival, one using IPUMS data for immigrants observed years later at census, and one using RCI data for emigrants skills upon departure.

The results of placebo tests on migrant skills are shown in Table 12. Each number in

---

<sup>46</sup>Year 1930 is not included because the full cohort has not arrived by census enumeration.

the columns represent a separate regression where the treatment variable is the year in the leftmost column. The empirical strategy is valid as there is no statistically significant difference between new and old migrants, even going through World War I.

Interestingly, there is a strong correlation between migrant scores upon arrival and migrant scores measure years later. One may suggest that occupational downgrading upon arrival would lead to new migrants having lower occupational scores, but there is no evidence that there are large differences in occupational downgrading between new and old migrants since cohorts measured upon arrival and measured years later have similar differences. Also, it could be true that after the implementation of quotas, lower quality new migrants, in unobservable terms, arrived in the United States, which lower their occupational upgrading over the 1920s. This is possible, but there is no evidence that there are differences in occupational upgrading between new and old migrants prior to the quotas.

### 8.5.2 World War I

Immigration policy was only one major factor that shocked immigration flows in the early 20th century; World War I and its subsequent fallout could have altered return migration behavior. Greenwood and Ward (2014) show that larger population loss leads to fewer migrants returning back home, likely as migrants avoided the damage and fallout to economies during the late 1910s and early 1920s. While the return rates for an average migrants fell, it is unclear how this may affect the composition of out-migrant flows and thus the selection of return migrants. All previous regressions include controls for World War I death rates, so this should have no effect on the estimation of the effect of quotas.

## 9 Conclusion

Of the millions that decided to immigrate to the United States in the early 20th century, many returned home. Reasons for returning home vary across different parts of the population, but to paint the complete picture of how immigration impacts the economy, one needs to take into account return migration.

Using data that directly observes out-migrants upon departure, return migrants were more likely to be male, older, married, “new” and from the Northeast than the migrant population. However, patterns slightly differ across ethnicities. These results are in sharp contrast to migrants who planned to leave upon arrival; planned return migrants were more likely to be female, single, “old” and settling in the West. Differences between actual and planned return migrants are most likely due to shocks to income following arrival.

I also show that temporary migrants typically were lower skilled than the rest of the migrant population. However, those that plan to return home were positively self-selected on occupational scores; it is possible that the least skilled migrants are more likely to experience negative shocks in the United States. Another possibility is that reported occupations do not truly represent a return migrant's skills – planned return migrants were negatively self-selected on height, which is positively correlated with health, intelligence and productivity.

This negative self-selection of out-migrant on occupations remained between 1910 and 1920, but by 1930 return migrants no longer were negatively self-selected. The increase the return migrant occupational scores was especially strong for new migrants where it turned positive. I explore whether migration quotas was one of the causes for the increase in self-selection.

Quota laws had unintended consequences on the migrant stock due to effects on temporary migrants. Temporary migrants arrived for a couple years and would leave the United States quickly, where those that remained were of higher skill. However, following the quotas, low-skilled migrants who arrived no longer went back home, but stayed in the United States. This is consistent with a large supply shock in unskilled markets, driving up wages and employment opportunities for low-skilled migrants. The same jobs and wages in northern industrial centers that attracted blacks from the South (Collins, 1997) kept new migrants remaining in the United States – migrants are not solely a deterrent to black labor, but also to other migrants.

Migration policy aimed at affecting the characteristics of the migrant stock needs to account not just for effects on the incoming migrant flow but also the out-going migrant flow. It is possible that entry restrictions change the behavior of those already in the United States or those lucky enough to cross the border, which in the end alter the effects of policy of the overall migrant stock.

## References

- ABRAMITZKY, R., L. P. BOUSTAN, AND K. ERIKSSON (2012): “Europes Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration,” *American Economic Review*, 102(5), 1832–1856.
- (2014): “A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration,” *Journal of Political Economy*, p. Forthcoming.
- BAILEY, W. B. (1912): “The Bird of Passage,” *American Journal of Sociology*, 18(3), pp. 391–397.
- BAINES, D. (1995): *Emigration from Europe 1815-1930*, vol. 11. Cambridge University Press.
- BANDIERA, O., I. RASUL, AND M. VIARENGO (2013): “The Making of Modern America: Migratory Flows in the Age of Mass Migration,” *Journal of Development Economics*, 102, 23–47.
- BIAVASCHI, C. (2013): “Fifty Years of Compositional Changes in US Out-migration, 1908–1957,” Discussion paper, Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit.
- BIJWAARD, G. E., AND J. WAHBA (2014): “Do High-Income or Low-Income Immigrants Leave Faster?,” *Journal of Development Economics*.
- BORJAS, G. J. (1985): “Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants,” *Journal of labor Economics*, pp. 463–489.
- BORJAS, G. J. (1987): “Self-Selection and the Earnings of Immigrants,” *The American Economic Review*, 77(4), pp. 531–553.
- BORJAS, G. J., AND B. BRATSBERG (1996): “Who Leaves? The Outmigration of the Foreign-Born,” *The Review of Economics and Statistics*, 78(1), pp. 165–176.
- CHIQUIAR, D., AND G. H. HANSON (2005): “International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States,” *Journal of Political Economy*, 113(2).
- CHISWICK, B. R. (1978): “The Effect of Americanization on the Earnings of Foreign-Born Men,” *The Journal of Political Economy*, pp. 897–921.
- COHN, R. L. (1984): “Mortality on Immigrant Voyages to New York, 1836-1853,” *The Journal of Economic History*, 44(2), pp. 289–300.
- COHN, R. L. (2009): *Mass Migration Under Sail: European Immigration to the Antebellum United States*. Cambridge University Press.
- COLLINS, W. J. (1997): “When the Tide Turned: Immigration and the Delay of the Great Black Migration,” *Journal of Economic History*, 57, 607–632.

- COLLINS, W. J., AND M. H. WANAMAKER (2014): “Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data,” *American Economic Journal: Applied Economics*, 6(1), 220–252.
- CONSTANT, A., AND D. S. MASSEY (2003): “Self-selection, Earnings, and Out-migration: A Longitudinal Study of Immigrants to Germany,” *Journal of Population Economics*, 16(4), 631–653.
- DULEEP, H. O. (1994): “Social Security and the Emigration of Immigrants,” *Soc. Sec. Bull.*, 57, 37.
- DUSTMANN, C., AND Y. WEISS (2007): “Return Migration: Theory and Empirical Evidence from the UK,” *British Journal of Industrial Relations*, 45(2), 236–256.
- FERENCZI, I., AND W. WILLCOX (1929): “International Migration Statistics,” in *International Migrations, Volume I: Statistics*, pp. 47–76. NBER.
- FERNÁNDEZ-HUERTAS MORAGA, J. (2013): “Understanding different migrant selection patterns in rural and urban Mexico,” *Journal of Development Economics*, 103, 182–201.
- FRIEDBERG, R. M., AND J. HUNT (1995): “The Impact of Immigrants on Host Country Wages, Employment and Growth,” *The Journal of Economic Perspectives*, 9(2), 23–44.
- GIBSON, J., AND D. MCKENZIE (2011): “The Microeconomic Determinants of Emigration and Return Migration of the Best and Brightest: Evidence from the Pacific,” *Journal of Development Economics*, 95(1), 18–29.
- GOLDIN, C. (1994): “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. University of Chicago Press.
- GOLDIN, C., AND R. A. MARGO (1992): “The Great Compression: The Wage Structure in the United States at Mid-Century,” *The Quarterly Journal of Economics*, 107(1), 1–34.
- GOULD, J. D. (1980): “European Inter-Continental Emigration. The Road Home: Return Migration from the USA,” *Journal of European Economic History*, 9(1), 41–112.
- GREENWOOD, M. J., AND ZACHARY WARD (2014): “Immigration Quotas, World War I, and Emigrant Flows from the United States in the Early 20th Century,” .
- HATTON, T. J., AND J. G. WILLIAMSON (2005): *Global Migration and the World Economy: Two Centuries of Policy and Performance*. Cambridge Univ Press.
- HUTCHINSON, E. P. (1958): “Notes on immigration statistics of the United States,” *Journal of the American Statistical Association*, 53(284), 963–1025.
- ILAH, N. (1999): “Return Migration and Occupational Change,” *Review of Development Economics*, 3(2), 170–186.
- JEROME, H. (1926): “Migration and Business Cycles,” .
- KOSACK, E., AND Z. WARD (2014): “Who Crossed the Border? The Selection of Mexican

- Migrants in the Early 20th Century,” .
- LAFORTUNE, J., AND J. TESSADA (2012): “Smooth (er) Landing? The Dynamic Role of Networks in the Location and Occupational Choice of Immigrants,” .
- LINDSTROM, D. P. (1996): “Economic Opportunity in Mexico and Return Migration from the United States,” *Demography*, 33(3), 357–374.
- LUBOTSKY, D. (2007): “Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings,” *Journal of Political Economy*, 115(5), 820–867.
- MASSEY, C. (2012): “Immigration Quotas and Immigrant Skill Composition: Evidence from the Pacific Northwest,” .
- MESNARD, A. (2004): “Temporary Migration and Capital Market Imperfections,” *Oxford Economic Papers*, 56(2), 242–262.
- MORAGA, J. F.-H. (2011): “New Evidence on Emigrant Selection,” *The Review of Economics and Statistics*, 93(1), 72–96.
- SJAASTAD, L. A. (1962): “The Costs and Returns of Human Migration,” *The Journal of Political Economy*, pp. 80–93.
- SPITZER, Y. (2014): “Pogroms, Networks, and Migration: The Jewish Migration from the Russian Empire to the United States 1881-1914,” .
- STECKEL, R. H. (2009): “Heights and Human Welfare: Recent Developments and New Directions,” *Explorations in Economic History*, 46(1), 1–23.
- UNITED STATES BUREAU OF IMMIGRATION (1908–1930): *Annual Report of the Commissioner General of Immigration*. U.S. Government Printing Office.
- VAN HOOK, J., W. ZHANG, F. D. BEAN, AND J. S. PASSEL (2006): “Foreign-born Emigration: A New Approach and Estimates Based on Matched CPS files,” *Demography*, 43(2), 361–382.
- WILLCOX, W. F. (1931): “Appendices to” International Migrations, Volume II: Interpretations,” in *International Migrations, Volume II: Interpretations*, pp. 637–658. NBER.
- WILLIAMSON, J. G. (1995): “The Evolution of Global Labor Markets Since 1830: Background Evidence and Hypotheses,” *Explorations in Economic History*, 32(2), 141–196.
- WYMAN, M. (1996): *Round-trip to America: the immigrants return to Europe, 1880-1930*. Cornell University Press.
- YANG, D. (2006): “Why do Migrants Return to Poor Countries? Evidence from Philippine Migrants’ Responses to Exchange Rate Shocks,” *The Review of Economics and Statistics*, 88(4), 715–735.
- ZEIDEL, R. F. (2004): *Immigrants, Progressives, and Exclusion Politics: The Dillingham Commission, 1900-1927*. Northern Illinois University Press.

Figure 1: Wage Profile of a Cohort

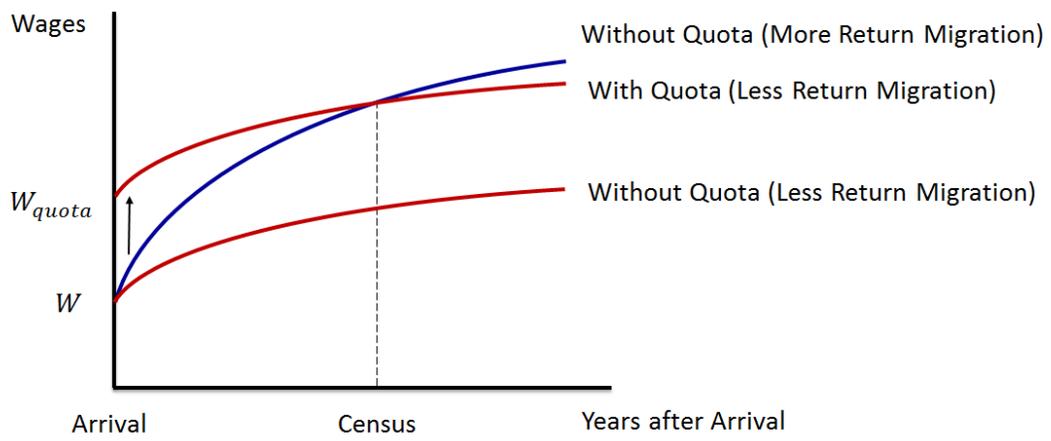
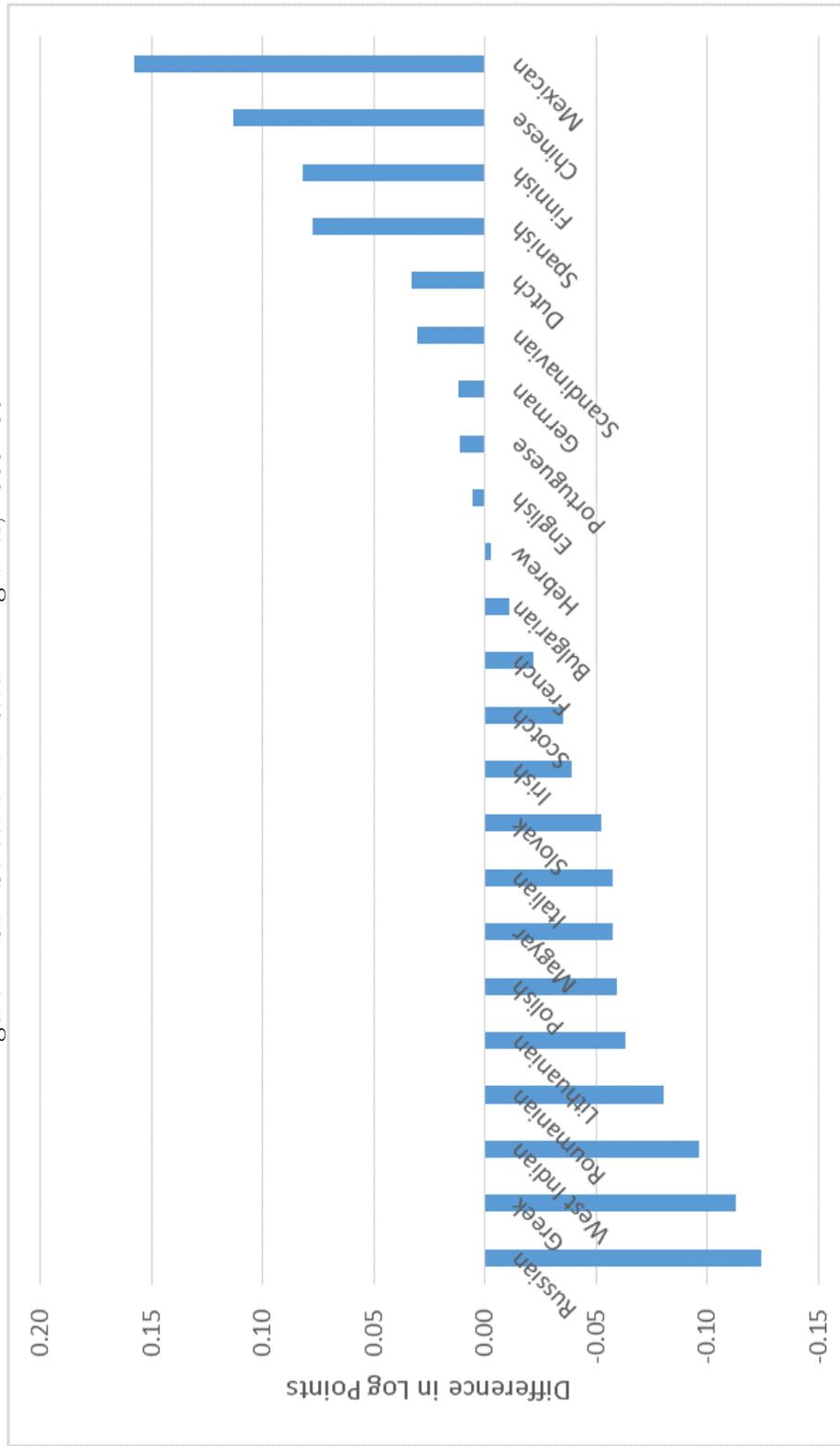
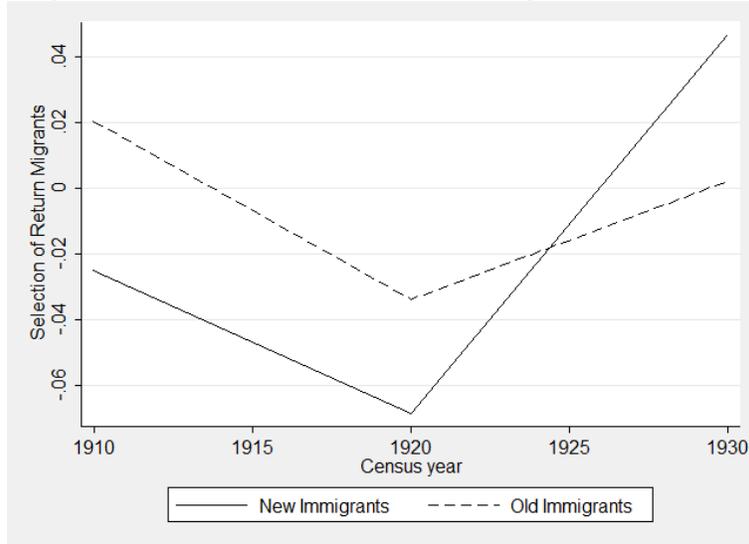


Figure 2: Self-Selection of Return Migrants, 1908-1932



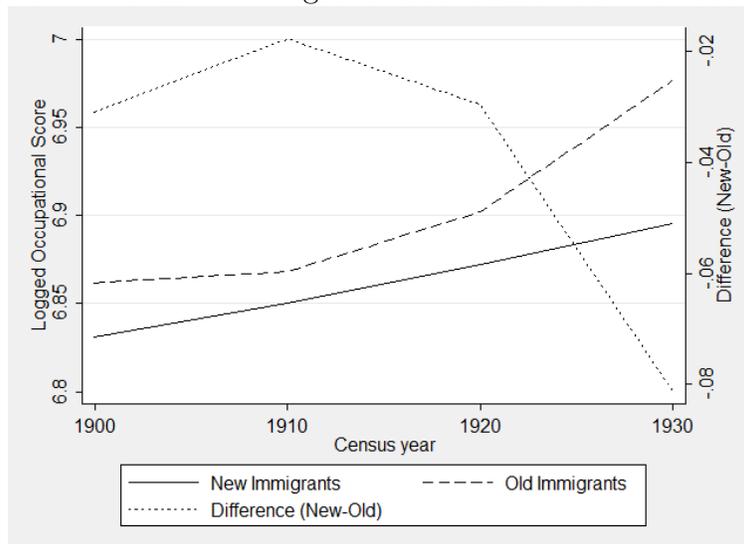
Notes: Data is from the Annual Report of the Commissioner General of Immigration (1908-1930), and IPUMS (1910-1930). The difference is between the logged occupational scores, a positive value indicated return migrants have higher occupational scores than foreign born. The foreign born occupational score is weighted to match gender and length of stay of return migrant population.

Figure 3: Self-Selection of Return Migrants, 1910-1930



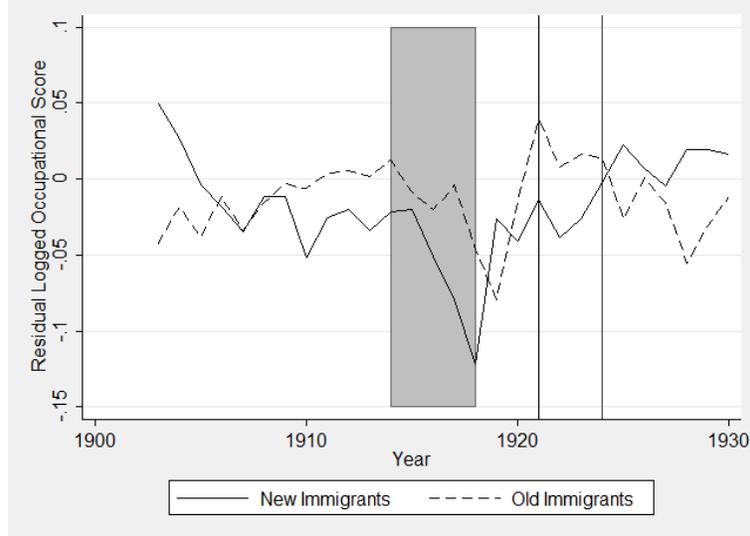
Notes: Data is from the Annual Report of the Commissioner General of Immigration (1911, 1921, and 1931) and IPUMS (1910, 1920, and 1930).

Figure 4: Trends in Immigrant Recent Stock's Skills 1900-1930



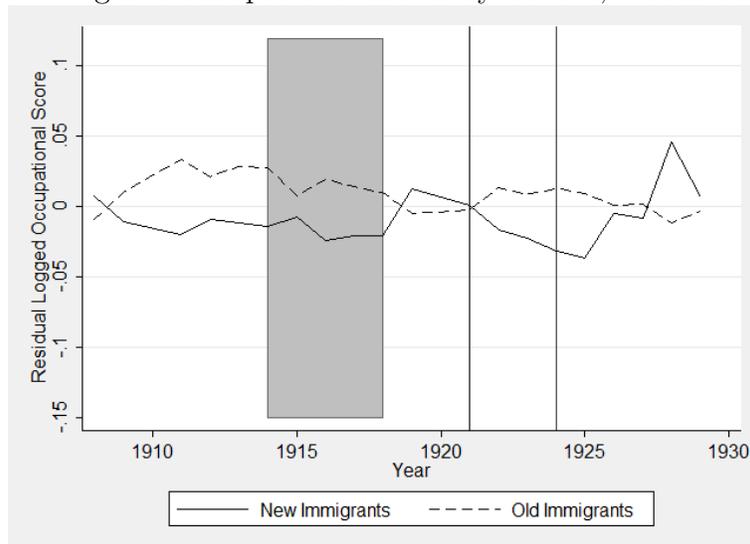
Notes: Data is from IPUMS (1900-1930). Data is for the immigrant stock of those arriving in the past ten years. New migrants are from Eastern and Southern Europe and Asia, while old migrants are from Northern and Western Europe, and North America.

Figure 5: Immigrant Occupational Scores by Cohort, Measured Upon Arrival



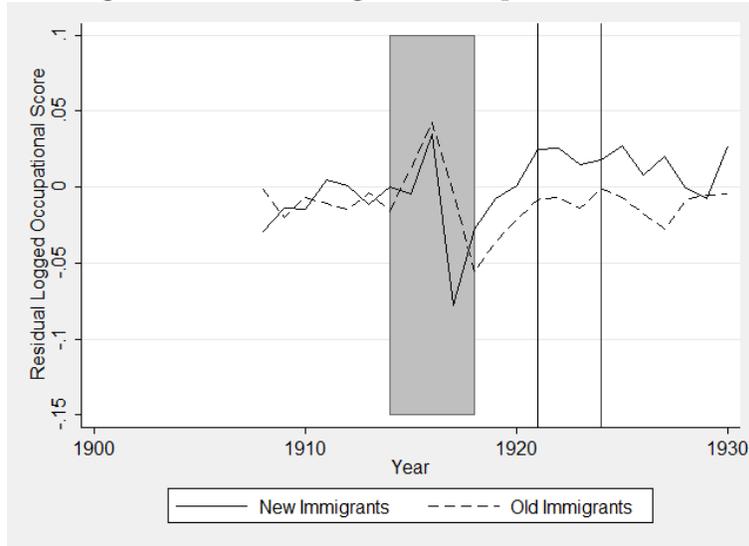
Notes: Data is from the Annual Report of the Commissioner General of Immigration (1908-1930). The shaded area is during World War I, and the two vertical lines coincide with 1921 and 1924 quotas.

Figure 6: Immigrant Occupational Scores by Cohort, Measured at Census



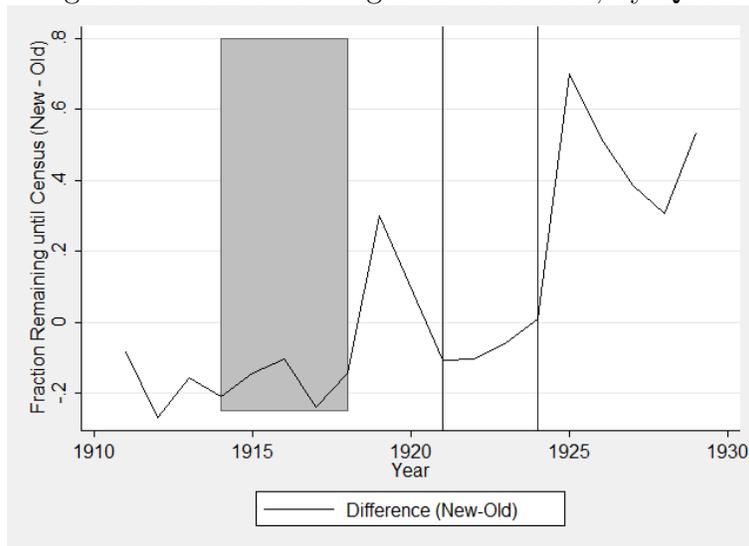
Notes: Data is from recent migrant stock of past ten years, observed IPUMS (1910-1930). Year is the year the cohort arrived, but they are measured later in 1910, 1920, and 1930 census. The shaded area is during World War I, and the two vertical lines coincide with 1921 and 1924 quotas.

Figure 7: Return Migrant Occupational Scores



Notes: Data is from the Annual Report of the Commissioner General of Immigration (1908-1930). The shaded area is during World War I, and the two vertical lines coincide with 1921 and 1924 quotas.

Figure 8: Fraction of Migrants 1900-1930, by Quota



Notes: Data is from the Annual Report of the Commissioner General of Immigration (1908-1932) and IPUMS (1900-1930). Fraction stay is the ratio of number arrived in a year observed in the census over the number arrived in a year observed in RCI.

Table 1: Actual Return Migrants Summary Statistics, by Ethnicity

Ethnicity	Male	Age 16-45	Married	Length (Years)							State of Last Residence					Total
				0-5	5-10	10-15	15-20	20+	NE	MW	S	W				
Overall	79.6	79.4	68.5	61.8	27.0	6.7	2.4	2.2	63.7	20.7	5.8	9.7	3,893,293			
Armenian	94.3	83.6	61.5	57.9	32.6	6.5	2.1	0.9	66.4	26.5	2.1	5.0	8,407			
Chinese	95.8	42.4	74.0	24.5	20.9	11.3	10.1	33.2	21.2	7.5	2.3	68.9	78,779			
Bulgarian	91.8	79.4	78.8	46.7	42.2	7.8	2.4	0.9	33.1	54.5	3.0	9.4	102,882			
Croatian	89.6	85.2	79.5	65.5	27.8	4.7	1.3	0.7	48.7	38.8	3.5	8.9	116,018			
Cuban	66.7	73.1	44.6	81.0	12.5	3.5	1.4	1.6	36.7	1.3	61.4	0.6	34,509			
Dutch	66.4	72.6	58.4	64.2	23.8	7.0	2.6	2.3	47.0	41.6	4.1	7.4	11,664			
East Indian	95.8	76.3	44.9	43.1	36.0	14.2	4.1	2.4	20.4	6.9	4.2	68.4	33,427			
English	56.9	72.0	53.1	69.4	19.2	6.1	2.5	2.9	65.2	17.7	4.6	12.5	2,900			
Finnish	71.9	84.8	55.8	54.7	30.2	8.9	3.6	2.7	44.8	41.1	1.4	12.7	196,033			
French	59.0	74.0	50.6	67.3	20.1	6.3	2.8	3.4	76.5	8.9	4.2	10.4	28,701			
German	60.6	78.9	50.4	70.6	19.8	4.9	2.2	2.6	62.7	27.9	4.1	5.4	76,282			
Greek	95.2	81.8	68.1	55.9	34.3	6.9	2.0	0.8	59.8	26.7	4.7	8.8	156,559			
Hebrew	75.2	80.2	56.1	77.7	17.3	3.3	0.9	0.8	87.7	9.3	1.5	1.5	197,104			
Irish	52.6	79.7	35.6	55.1	25.7	9.6	4.0	5.6	77.7	13.4	1.9	7.0	54,658			
Italian	88.6	82.2	74.4	63.4	28.1	5.9	1.7	1.0	77.9	14.4	3.0	4.8	58,503			
Japanese	77.5	77.0	66.1	43.2	26.5	17.3	8.2	4.8	9.4	1.7	1.3	87.7	1,232,916			
Korean	85.5	75.1	47.8	40.9	23.8	22.9	7.5	4.9	7.2	5.7	2.0	85.0	60,427			
Lithuanian	75.0	80.1	71.6	53.5	28.0	12.1	4.2	2.2	68.8	28.8	2.0	0.3	1,224			
Magyar	75.7	78.9	85.1	62.8	29.5	5.3	1.7	0.8	62.9	33.9	2.5	0.7	37,960			
Mexican	63.3	71.6	50.6	58.9	23.6	9.4	5.0	3.1	5.4	4.2	44.6	45.9	155,282			
Polish	78.4	81.3	84.4	58.5	30.1	7.8	2.4	1.2	70.3	27.2	1.8	0.7	140,583			
Portuguese	76.8	81.5	65.8	61.1	27.3	8.3	2.1	1.3	88.3	0.5	0.7	10.5	338,641			
Romanian	84.6	72.5	84.9	42.2	47.2	7.9	2.0	0.7	36.1	60.4	2.8	0.6	57,420			
Russian	87.0	86.0	69.6	74.1	20.0	4.2	1.3	0.5	66.8	23.5	3.7	6.1	70,810			
Scandinavian	65.3	82.2	39.6	60.1	26.3	7.6	2.9	3.1	51.3	33.7	2.7	12.3	102,506			
Scotch	59.3	74.4	50.1	73.3	18.0	4.8	1.8	2.1	57.6	27.8	3.3	11.2	109,818			
Slovak	78.4	79.3	80.6	57.0	31.2	7.8	2.6	1.4	68.2	27.6	2.8	1.4	53,418			
Spanish	88.1	83.5	53.6	75.3	17.9	4.6	1.3	0.9	70.2	3.2	14.7	11.8	158,722			
Spanish-American	65.2	79.7	34.7	81.6	13.2	3.3	1.0	0.8	67.0	3.8	15.5	13.7	84,626			
Syrian	79.7	78.0	64.0	44.5	39.5	10.9	3.2	1.9	70.8	18.9	7.8	2.4	27,275			
Turkish	95.8	89.2	62.5	66.0	26.9	5.0	1.5	0.7	64.1	28.6	3.2	4.2	15,892			
Welsh	68.6	75.9	49.0	71.7	18.4	5.0	1.8	3.0	52.4	29.4	3.9	14.4	12,275			
West Indian	54.6	80.5	54.3	69.6	21.0	6.0	1.7	1.6	64.0	1.6	33.8	0.7	4,075			

Notes: Data is from the Annual Report of the Commissioner General of Immigration (1908-1932).

Table 2: Selection of Actual Return Migrants, 1908-1932

Ethnicity	Male	Age					Region of Last Residence				
		Under 16	16-45	Over 45	Married	Northeast	Midwest	South	West		
Overall	17.5	-8.5	1.2	7.3	20.5	10.3	-3.1	-0.4	-6.7		
Armenian	35.0	-4.9	3.8	1.1	8.0	11.9	-0.5	1.5	-12.9		
Bulgarian	4.0	-3.5	-11.2	14.7	21.7	21.8	4.2	0.8	-26.8		
Chinese	3.4	0.1	-20.1	20.0	29.9	-1.7	0.0	-1.7	3.5		
Croatian	16.4	-5.7	-1.2	6.8	30.9	14.2	-6.8	-1.2	-6.2		
Dutch	8.9	-10.5	2.6	7.9	9.9	20.5	-16.6	2.1	-6.1		
East Indian	3.6	-3.3	-14.6	17.9	-5.7	14.3	-1.7	1.2	-13.9		
English	6.1	-11.3	4.9	6.4	8.6	11.4	-5.1	-0.9	-5.4		
Finnish	18.8	-6.6	2.0	4.6	10.4	10.5	-6.2	1.1	-5.3		
French	13.8	-17.0	11.2	5.8	11.6	18.1	-13.5	1.2	-5.9		
German	9.4	-7.7	2.5	5.2	2.3	17.2	-10.8	-0.6	-5.8		
Greek	14.2	-2.5	-7.8	10.3	33.0	14.8	-0.5	-2.8	-11.5		
Hebrew	25.3	-17.7	13.3	4.4	11.9	5.8	-4.5	-1.3	0.0		
Irish	14.5	-1.3	-3.7	5.0	4.2	-0.3	0.6	-0.3	0.1		
Italian	27.1	-11.9	5.8	6.1	23.8	6.6	-0.4	-1.0	-5.2		
Japanese	10.7	1.2	-11.9	10.7	10.2	7.1	0.1	1.2	-8.4		
Korean	4.9	-0.3	-9.6	10.0	-1.2	-1.9	-5.8	1.1	6.6		
Lithuanian	11.0	-2.0	-7.1	9.1	20.5	4.6	-2.6	-1.0	-1.0		
Magyar	16.5	-8.4	-1.2	9.6	28.2	10.2	-7.2	-1.4	-1.5		
Mexican	4.5	-8.8	5.4	3.4	7.5	4.6	-1.2	-2.2	-1.1		
Polish	17.6	-6.5	-2.8	9.3	28.0	7.4	-6.0	-0.7	-0.7		
Portuguese	16.6	-8.0	5.4	2.6	13.2	19.3	0.0	-0.9	-18.4		
Roumanian	14.4	-5.4	-9.5	14.9	33.3	1.7	2.8	-0.2	-4.2		
Russian	27.5	-11.4	8.3	3.2	22.3	2.5	1.9	-1.1	-3.3		
Scandinavian	9.7	-6.4	1.1	5.3	3.0	24.4	-14.1	0.6	-10.8		
Scotch	10.8	-9.7	5.4	4.4	6.1	3.0	-0.1	0.5	-3.4		
Slovak	18.4	-8.1	-1.5	9.6	22.2	17.5	-13.5	-2.2	-1.8		
Spanish	18.5	-8.4	2.2	6.1	17.4	23.0	-2.4	-2.0	-18.6		
Spanish-American	6.5	-15.6	12.1	3.6	6.6	15.5	-2.2	3.7	-17.0		
Syrian	23.1	-4.4	-2.2	6.5	15.6	14.8	-8.1	-2.0	-4.6		
Turkish	19.4	-9.1	12.6	-3.5	17.9	7.2	0.7	-2.3	-5.5		
Welsh	9.3	-12.6	7.0	5.7	6.3	-8.0	4.5	3.8	-0.3		

Notes: Data is from IPUMS (1910-1930) and the Annual Report of the Commissioner General of Immigration (1908-1932). Numbers are the mean of the out-migrants minus mean of the foreign born in census.

Table 3: Planned Return and Planned Permanent Summary Statistics

	Planned Return	Planned Permanent	Difference
Age	27.97 (11.45)	27.83 (11.84)	0.135
Male	0.504 (0.500)	0.544 (0.498)	-0.0396***
Number of Children	0.122 (0.472)	0.345 (0.861)	-0.224***
Ever Married	0.325 (0.468)	0.402 (0.490)	-0.0768***
Join Family	0.769 (0.422)	0.868 (0.338)	-0.0993***
Travel Alone	0.814 (0.389)	0.660 (0.474)	0.154***
Log(Occupational Score+1)	5.000 (2.935)	4.774 (3.082)	0.133*
Log(Occupational Score), if have job	6.694 (0.434)	6.738 (0.418)	-0.0441***
Height (cm)	164.1 (13.20)	163.7 (13.52)	0.420
Height (cm), if have job	165.3 (12.19)	166.1 (10.58)	-0.828**
Eastern or Southern European (New)	0.401 (0.490)	0.478 (0.500)	-0.0770***
Repeat Migrant	0.172 (0.377)	0.130 (0.336)	0.0420***
Northeast	0.625 (0.484)	0.612 (0.487)	0.0130
Midwest	0.241 (0.428)	0.290 (0.454)	-0.0496***
West	0.102 (0.302)	0.0587 (0.235)	0.0429***
South	0.0298 (0.170)	0.0381 (0.192)	-0.00829
Urban (2,500 residents)	0.827 (0.379)	0.824 (0.381)	0.00297
Big City (100,000 residents)	0.616 (0.487)	0.594 (0.491)	0.0219*
Planned Length of Stay (Years)	4.352 (3.096)		
Observations	1,585	12,275	

Notes: Comparison of planned returns from incoming passenger manifests and actual returns from RCI.

Table 4: Planned and Actual Return Rates, 1917-1924

Ethnicity	(1) Planned Return Rate	(2) Actual Return Rate	(3) Difference (2)-(1)	(4) Actual Return Rate Lower Bound	(5) Difference (4)-(1)
Hebrew	1.2	0.7	-0.6	0.4	-0.8
Bohemian	1.7	31.7	30.0	22.7	20.9
Polish	2.1	14.7	12.6	8.9	6.8
Croatian	2.6	6.8	4.2	4.5	1.8
Scotch	2.8	6.7	3.9	4.1	1.3
Magyar	3.8	10.9	7.0	6.9	3.0
German	4.9	7.0	2.1	4.3	-0.5
Slovak	6.0	6.1	0.1	3.3	-2.7
Greek	7.1	43.9	36.8	19.3	12.2
Russian	7.5	16.8	9.3	6.7	-0.8
English	13.0	14.4	1.4	8.5	-4.4
Finnish	15.6	13.7	-2.0	6.7	-8.9
Italian	16.4	25.4	9.0	17.2	0.8
Dutch	22.8	15.8	-6.9	9.2	-13.6
Welsh	23.9	6.7	-17.2	3.8	-20.1
Scandinavian	25.0	12.7	-12.3	6.8	-18.2
Irish	29.9	4.4	-25.5	2.4	-27.6
French	38.2	8.9	-29.3	3.3	-34.9

Notes: Comparison of planned returns from incoming passenger manifests and actual returns from RCI.

Table 5: Self-Selection of Return Migrants on Occupation Score from 1910-1930

Ethnicity	1910	1920	1930
Mean	-0.020	-0.064	0.031
Old Migrants	0.020	-0.034	0.002
New Migrants	-0.025	-0.069	0.047
Armenian	-0.015	-0.044	0.017
Bulgarian	0.093	-0.114	-0.188
Chinese	0.094	0.165	0.023
Croatian	-0.035	-0.085	-0.075
Dutch	0.032	-0.041	0.063
East Indian	0.604	0.240	-0.178
English	0.021	0.017	0.004
Finnish	0.122	0.045	0.161
French	-0.024	-0.042	-0.121
German	0.049	-0.026	0.029
Greek	-0.072	-0.157	-0.191
Hebrew	-0.002	-0.002	0.045
Irish	0.002	-0.129	-0.028
Italian	-0.040	-0.077	-0.069
Japanese	0.058	0.216	0.325
Korean	0.293	0.326	-0.519
Lithuanian	-0.035	-0.078	-0.070
Magyar	-0.024	-0.099	-0.062
Mexican	0.348	0.189	0.126
Polish	-0.030	-0.114	-0.066
Portuguese	0.064	-0.004	-0.081
Roumanian	-0.023	-0.114	-0.104
Russian	-0.101	-0.144	0.117
Scandinavian	0.046	-0.031	0.007
Scotch	-0.111	-0.216	0.042
Slovak	-0.046	-0.048	-0.044
Spanish	0.085	0.090	0.089
Spanish-American	0.318	0.045	0.083
Syrian	0.020	-0.161	0.065
Turkish	0.022	-0.124	0.111
Welsh	-0.003	0.055	-0.079

Notes: Data is from IPUMS (1910-1930) and the Annual Report of the Commissioner General of Immigration (1911, 1921, and 1931). Numbers are the mean of the out-migrants' occupational score minus mean of the foreign born's occupational score in census.

Table 6: Self-Selection into Planned Migration on Quality, 1917-1924

	(1)	(2)	(3)	(4)
Log (Occupational Score)	0.0171** (0.00806)		0.0200** (0.00816)	0.0178** (0.00845)
Height (cm)		-0.000854*** (0.000326)	-0.000911*** (0.000331)	-0.000780** (0.000348)
New x Post 1921	-0.0848*** (0.0199)	-0.0885*** (0.0199)	-0.0886*** (0.0199)	-0.0842*** (0.0199)
Year FE	X	X	X	X
Ethnicity FE	X	X	X	X
Individual Controls				X
Observations	9,960	9,858	9,858	9,853
R-squared	0.136	0.136	0.134	0.144

Notes: Data is from incoming passenger manifests, 1917-1924. The dependent variable is whether or not an individual plans to return home.

Table 7: Planned Length of Stay, 1917-1924

	(1)	(2)	(3)	(4)
Log(Occupational Score)	-0.406* (0.217)		-0.438* (0.223)	-0.326 (0.251)
Height (cm)		0.0113 (0.00869)	0.0131 (0.00844)	0.0114 (0.00900)
New x Post 1921	1.647** (0.788)	1.512* (0.825)	1.424* (0.810)	1.187 (0.785)
Year FE	X	X	X	X
Ethnicity FE	X	X	X	X
Individual Controls				X
Observations	743	726	726	725
R-squared	0.331	0.340	0.343	0.357

Notes: Data is from incoming passenger manifests, 1917-1924. The dependent variable is whether or not an individual plans to return home.

Table 8: Trends in Migrant Stock's Occupation Scores 1900-1930

	(1)	(2)	(3)
New Migrant x 1930	-0.0113 (0.0212)	-0.0197 (0.0218)	-0.0147 (0.0318)
Literacy		X	X
Years in United States FE	X	X	X
Country FE	X	X	X
Year FE	X	X	X
Sample Specification	All	All	Males
Observations	80,133	80,133	64,375
R-squared	0.087	0.093	0.112

Notes: Data is from 1% IPUMS (1900-1930). Standard errors are clustered for place of birth. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Dependent variable is logged occupational score.

Table 9: Effect of Quotas on Immigrant's Occupational Score

	(1)	(2)	(3)	(4)
<i>Panel A: Immigrants upon Entry</i>				
New x Post 1921	0.0222*		-.0225	
	(0.0233)		(0.0150)	
New x Post 1924		0.0497***	0.0671***	
		(0.0222)	(0.0197)	
Quota Restriction				0.0510***
				(0.0232)
<i>Panel B: Immigrants at Census</i>				
New x Post 1921	0.0026		-0.0125	
	(0.0170)		(0.0156)	
New x Post 1924		0.0134	0.0237	
		(0.0189)	(0.0193)	
Quota Restriction				-0.00890
				(0.0223)
Ethnicity FE	X	X	X	X
Year FE	X	X	X	X
Literacy Rates and World War I	X	X	X	X

Notes: Data is from Annual Report of the Commissioner General of Immigration (1908-1929) for Panel A and IPUMS (1910-1930) for Panel B. Both panels are country by year panel data with 462 observations. Standard errors clustered by year for both panels. Dependent variable is the logged occupational score. Additional controls include literacy and World War I death rates.

Table 10: Effect of Quotas on Out-Migrant's Occupational Score

Variables	(1)	(2)	(3)	(4)
New x Post 1921	0.0357** (0.0118)		0.0484*** (0.0117)	
New x Post 1924		0.0172 (0.0107)	-0.0191* (0.00721)	
Quota Restriction				0.0609** (0.0150)
Ethnicity FE	X	X	X	X
Year FE	X	X	X	X
Literacy Rates and World War I	X	X	X	X
Observations	462	462	462	462
$R^2$	0.795	0.792	0.796	0.799

Notes: Data is from Annual Report of the Commissioner General of Immigration (1908-1929). Robust standard errors in parenthesis. Demographics control for the fraction of immigrants that are male and the literacy rate for immigrants.

Table 11: Fraction Remain in United States

Variables	(1)	(2)	(3)	(4)
New x Post 1921	0.164** (0.0655)		0.00708 (0.0672)	
New x Post 1924		0.241*** (0.0643)	0.235*** (0.0592)	
Quota Restriction				0.246** (0.0914)
Year FE	X	X	X	X
Ethnicity FE	X	X	X	X
Observations	462	462	462	462
$R^2$	0.576	0.595	0.595	0.588

Notes: Data is from IPUMS (1910-1930) and RCI (1908-1929). The dependent variable is the logged fraction of those that remain until the census enumeration. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$

Table 12: Placebo Tests on Immigrants and Emigrant Skills

Year	(1) Immigrants Upon Arrival	(2) Immigrants at Census	(3) Emigrants
1909	0.0137	-0.00792	0.0145
1910	0.00675	-0.0171	-0.00157
1911	0.00325	-0.0192	0.00710
1912	0.00529	-0.0162	-0.00446
1913	0.0105	-0.0179	-0.00224
1914	0.0181	-0.0113	0.00268
1915	0.0296	-0.00691	-0.00139
1916	0.0371	-0.00281	0.00377
1917	0.0347	0.00114	0.0278
1918	0.0338	-0.00468	0.0383**
1919	0.0181	0.0353**	0.0269
1920	-0.0269	0.0205	0.0321
Ethnicity FE	X	X	X
Year FE	X	X	X
Literacy and WWI	X	X	X
Observations	294	294	294

Notes: Data is from IPUMS (1910-1930) and RCI (1908-1921). The dependent variable is the logged occupational score. Each estimate is from a separate regression with a treatment variable for that year. \*p<0.10, \*\*p<0.05, \*\*\*p<0.01

## A BRV Estimates

In Table A1, I compare official records to estimates for selected countries reported by BRV for the decade 1910-1920 to determine the severity of undercount (see BRV (2013), Table 4).<sup>47</sup> There is a striking variation in mis-measurement of out-migrants across countries, where official statistics record 31.7% of BRV’s estimate of Italian out-migrants and 3.2% of total Great Britain emigrants. Undercounting of “old migrants” is more severe than “new migrants.” Interestingly, BRV’s estimates undercount out-migrants to Canada, where official statistics record 190% of BRV’s number. This over count is likely because official statistics record the intended destination while BRV assume that out-migrants return to their home country. As Biavaschi (2013) notes, BRV’s demographic techniques predict impossible negative emigration rates for almost 40% of nationalities, suggesting some degree of error in their estimates.

## B Occupational Scores

The first metric is the IPUMS variable “occscore,” which reflects the median earnings of individuals in an occupation in 1950. This is the measure used by Abramitzky, Boustan, and Eriksson (2014) for their indirect estimation of the self-selection of return migrants at the turn of the 20th century. The IPUMS variable assigns earnings to three-digit occupational codes, which in the dataset of foreign-born results in 240 occupations. Administrative data categorizes return migrants into 77 different occupations, which I match to IPUMS occupation codes and scores in a similar manner to Lafortune and Tessada (2013). While this variable is easily utilized, it measures earnings in 1950, decades after the time period under study. There was significant wage compression between the early and mid-twentieth century (Goldin and Margo, 1992), which would positively bias self-selection estimates as a large number of out-migrants are laborers. As an alternative metric, I use wages from the Cost of Living Survey (CLS) taken by the Bureau of Labor Statistics in 1901, recorded in Preston and Haines (1991) and used by Abramitzky, Boustan, and Eriksson (2012) to measure self-selection of Norwegian migrants. However, these earnings are based on urban, married households which may not be representative of migrant earnings.<sup>48</sup>

Estimates using the two alternative metrics (IPUMS 1950 and CLS 1901) are presented

---

<sup>47</sup>I use fiscal years 1911-1920 for the 1910-1920 decade, and use the definition “emigrant” for the count. It is impossible to compare out-migrant estimates for years 1901-1910 since the Department of Immigration only started to record out-migrants in 1908.

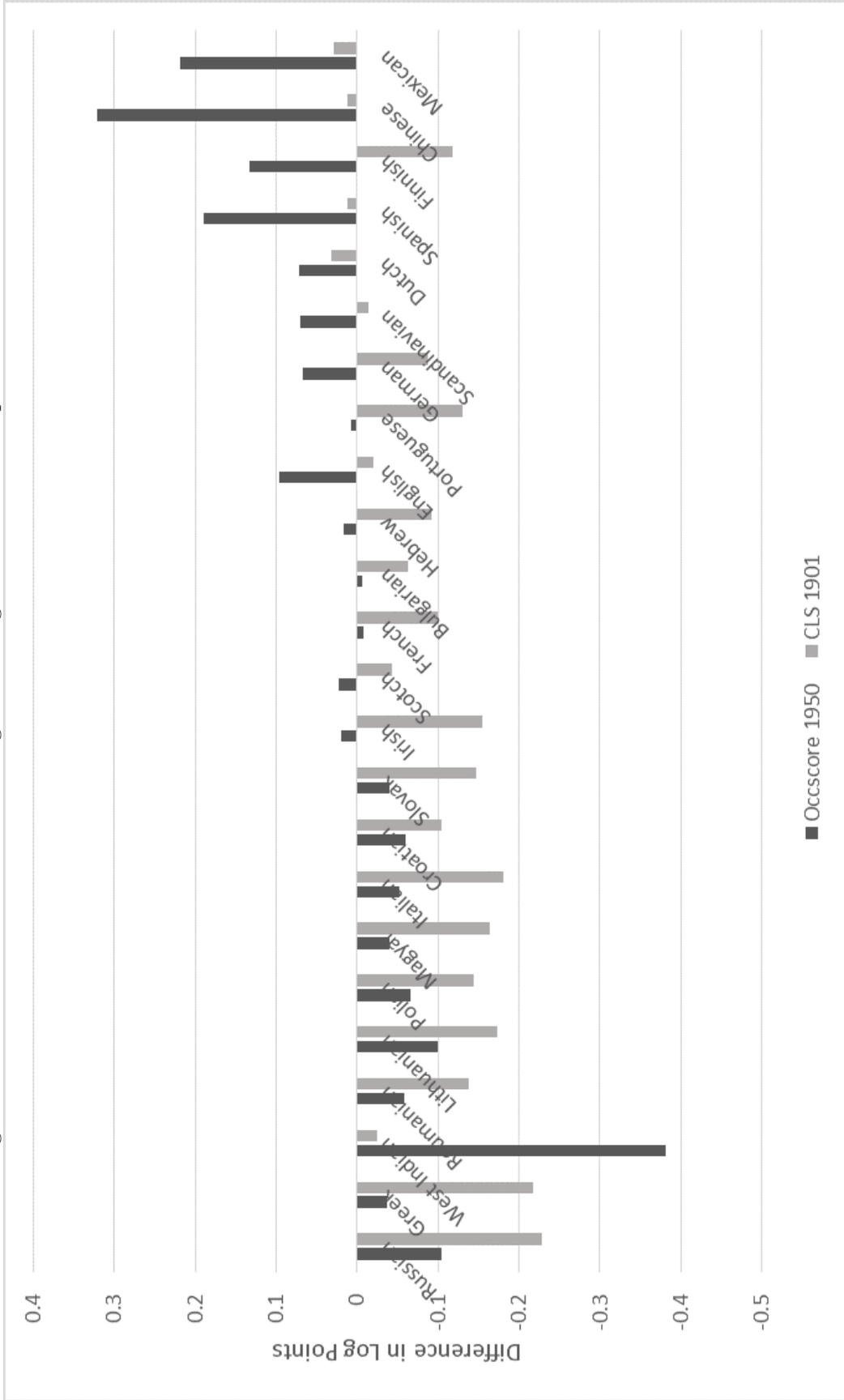
<sup>48</sup>I use the Abramitzky, Boustan and Eriksson’s (2012) calculation for farmer’s earnings reported in their appendix.

in Figure A1. Assigning occupations earnings from 1950 yields more positive estimates of self-selection than assigning earnings from 1940 or 1901.<sup>49</sup> When using the IPUMS 1950 occupational score, positive self-selection results for twelve of the thirty ethnicities reported, higher than eight positive for 1940 earnings and four positive for 1901 earnings. Wage compression from the early to mid-twentieth century positively biases self-selection positively as wages for laborers, a majority of out-migrants, are relatively higher during the 1950s. Indeed, using earnings from 1901 shows mostly negative selection of migrants, consistent with the story that temporary migrants are those that fail in the labor market.

---

<sup>49</sup>The IPUMS 1950 metric is the same measure used by Abramitzky, Boustan, and Eriksson (2014) in their indirect estimation of return migrant self-selection, where they find almost all countries exhibit negative self-selection. Although there are numerous differences for the method of estimation and sample of individuals, it appears that the indirect estimation yields results that are more negative than direct estimation. For example, my sample includes males and females, while theirs only includes males. Further, theirs uses individual-level data that corrects for age, while I cannot do the same. Their estimate is based on assimilation estimates from repeated cross-sections (permanent and temporary migrants) and panel data (only permanent migrants), which difference should reflect only temporary migrant's earnings.

Figure A1: Self-Selection of Return Migrants using Alternative Occupational Scores



Notes: Data is from the Annual Report of the Commissioner General of Immigration (1908-1930), and IPUMS (1910-1930). The difference is between the logged occupational scores, a positive value indicated return migrants have higher occupational scores than foreign born. The foreign-born occupational score is weighted to match gender and length of stay of return migrant population.

Table A1: Comparison to BRV Estimates, 1910-1920

Country	BRV	Official Statistics	Percent of BRV
Italy	2,281,362	723,483	31.71
Austria-Hungary	1,338,729	378,795	28.29
Russia	1,085,628	191,757	17.66
Great Britain	2,772,513	90,279	3.25
Germany	787,711	44,898	5.69
Ireland	522,350	22,898	4.38
Sweden	398,894	18,690	4.68
Greece	393,909	126,128	32.01
Norway	453,094	20,152	4.44
Spain	562,680	22,820	4.05
Canada	157,123	299,367	190.53

Notes: Comparison of out-migrant estimates from Bandiera, Rasul and Viarengo (2013, Table 4) and official statistics from the Annual Report of the Commissioner General of Immigration (1911-1920). Austria and Hungary summed from official records for Austria-Hungary, and Canada in BRV is British North American in RCI.