INSTITUTO DE ECONOMÍA

DOCUMENTO de TRABAJO

427

2012

Smooth(er) Landing? The Dynamic Role of Networks in the Location and Occupational Choice of Immigrants

Jeanne Lafortune; José Tessada.

www.economia.puc.cl • ISSN (edición impresa) 0716-7334 • ISSN (edición electrónica) 0717-7593

Versión impresa ISSN: 0716-7334
Versión electrónica ISSN: 0717-7593

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA

Oficina de Publicaciones
Casilla 76, Correo 17, Santiago
www.economia.puc.cl

SMOOTH(ER) LANDING? THE DYNAMIC ROLE OF NETWORKS IN THE LOCATION AND OCCUPATIONAL CHOICE OF IMMIGRANTS

Jeanne Lafortune*
José Tessada

Documento de Trabajo Nº 427

Santiago, septiembre 2012

* **jlafortune@uc.cl**

INDEX

Smooth(er) Landing? The Dynamic Role of Networks in the Location and Occupational Choice of Immigrants*

Jeanne Lafortune †   José Tessada‡

September 2012

Abstract

This paper studies the dynamic effect of networks on location and occupation decisions of immigrants to the United States between 1900 and 1930. We compare the distributions of immigrants both by intended and actual state of residence to counterfactual distributions constructed by allocating the national-level flows according to the distribution of previous immigrants and to measures of demand for occupations at the state level. Our results are consistent with migrants using ethnic networks as a transitory mechanism while they learn about their new labor markets and not with other hypotheses that do not account for the dynamic patterns we document.

JEL Codes: F22, J61, N31

*We thank Carol Graham, Sabrina Pabilonia, Peter Temin, Madeline Zavodny, and seminar participants at PUC-Rio, Universidad Alberto Hurtado, LSE/UCL Development Growth Seminar, University of North Carolina – Chapel Hill, Yale, American University, Clemson, PUC-Chile, University of British Columbia, UCL, the VIII Workshop of the Regional Integration Network (RIN) of LACEA, the Population Association of America 2010 Annual Meetings, SOLE 2011, and the Maryland Population Research Center for their comments, and particularly Judy Hellerstein for insightful discussions and detailed comments on this project. We wish to acknowledge Carolina Gonzalez-Velosa for excellent research assistance, Gérard Lafortune for careful data entry, and Eileen Gallagher for editorial assistance. Tessada thanks financial support from Fondecyt (Grant Iniciación #11110101). The usual disclaimer applies.

†Pontificia Universidad Católica de Chile. Email: jlafortune@uc.cl.

‡Pontificia Universidad Católica de Chile. Email: jtessada@gmail.com.

1 Introduction

The fact that immigrants of a particular country or ethnic group tend to locate in similar regions once in their new adoptive country has been observed for a long time. For example, this pattern has been observed in the most recent wave of migration to the United States, this time coming mostly from Mexico and Central America.[1] This pattern of concentration has not only been used in policy discussion: various studies have used it as a way to generate (plausibly exogenous) variation in the number of migrants to a particular location (see for example Altonji and Card 1991, Card 2001, and more recently Cortés 2008 and Peri and Sparber 2009, among many others). Research on migration has also looked at the connection between this pattern and the existence of networks in the labor markets; in a well-known paper Munshi (2003) argues that the size of the ethnic network (in his case, at the level of the village of origin) increases the probability that a new migrant will find employment in the destination, suggesting that this clustering may be linked to the capacity of immigrants to help each other find work through referrals. Relatedly, Patel and Vella (2007) emphasize that referrals, through the ethnic networks, may help explain the occupational concentration of different ethnicities in different cities that are observed in the current data in the United States.

However, the dynamic effects of ethnic networks on the occupation and location decisions of immigrants have not received significant attention previously. Do migrants select their destination by matching their set of skills to that of the main occupations within their network before they depart? Or do they just look for the state with the best match for their skills once they have arrived? How does the role of networks appear to change as the migrant spends more time in her adopted country and learns about labor markets in the United States? This paper attempts to shed light on these questions studying the dynamics of occupational and geographical choices of immigrants that arrived to the United States between 1900 and 1930. Using a simple motivating framework of location-occupation choice, we exploit a large dataset with gross and net

[1] For example, Card (2009) shows some figures for the case of Filipino immigrants to the United States in recent decades. For the immigration waves during the early 20th century Wegge (1998) shows that "chain migration" was prevalent among German immigrants. Furthermore, the anti-immigration discourses of the early 20th century used this pattern, among other issues, as an argument against (the mostly European) immigration over that period, the so-called "new immigrants".

immigration flows to the United States between 1900 and 1930 by ethnicity, occupation and state of residence (both intended when arriving and actual residence). The data allow us to contrast concentrations across states, ethnicities and occupations, measured at arrival to the United States and after immigrants have spent time in the country.

Overview. This study first sets up a simple location-occupation decision framework to motivate our empirical results. We argue that when ethnic networks only provide ethnic-specific amenities or consumption goods, then location choices should be related to the strength of pure ethnic networks but should have little or no relation to the specific occupations of the members of these networks beyond some ethnic-specific skill set.[2] If frictions or other features of the destination's labor market force immigrants to rely permanently on their network for job seeking, then their location decisions would be driven solely by the presence of individuals of the same ethnicity and occupation. On the contrary, if immigrants eventually assimilate in the local labor market but initially need previous immigrants to provide job referrals that serve as a (transitory) safety network, then we should observe that the occupations of recently arrived immigrants are highly correlated with the most prevalent occupations among their ethnic group in that particular destination. In addition, and contrary to the other two alternatives, this effect should change over time as immigrants who have spent more time in the country of destination may be finding employment in occupations better related to their skills and not to what is done by their fellow countrymen. Finally, in this case an immigrant would select initially the state according to the presence of immigrants from the same ethnic group and/or because of the match between her skills and the (relative) demand in that state (reflecting the transitory role of the referrals within the ethnic group).

We then exploit two sources of data covering migration to the United States between 1900 and 1930, a period that spans part of the Great Migration era, the years right after World War I, and the years when major immigration reforms were enacted. At that time, when immigrants arrived

[2]For example, Italians are more likely to work in pizza restaurants because they have specific knowledge about the *tasks* (i.e. they have the skills), but that is not a reflection of the role of networks neither does it depend on the specific destination. An exception to that would be if the occupation captures some general social class and ethnic-specific amenities or if consumption goods are related to ethnicity and social class.

to the United States (and when they boarded the ship), they were asked a series of questions including their ethnicity, the state where they *intended* to live and their occupation *in their country of origin*. We demonstrate that this answer appears to be highly related to their actual state of residence soon after arrival. These answers were tabulated and presented in the annual reports of the Commissioner of Immigration from 1899 to 1930, a data source previously unexplored to the best of our knowledge.[3] As they settled in the United States, immigrants were also surveyed during the decennial Censuses, reporting their year of arrival, country of birth, "actual" state of residence and "current" occupation. We match the cohorts of immigrants from these two data sources using the year of immigration and ethnicity to form a panel at the cohort-level.

Using these sources, we study whether the patterns of location and occupation choice implicit in the *gross* (measured at arrival) and *net* flows (measured in the census) of immigration are compatible with the hypothesis that ethnic networks provide temporary support in the labor market. To do this, we construct three separate counterfactual distributions of immigrants by allocating the national flow of immigrants by ethnicity and occupation for every year of our analysis. These counterfactual distributions are intended to separately measure the strength of the pure ethnic benefits, the state level demand for a occupation, and the ethnic-specific occupation match for the state by using the distribution of previous immigrants who shared one's ethnicity, occupation or both within the stock of immigrants living in the United States in 1900.[4] We then simultaneously compare the fit of the three counterfactual distributions to the actual distributions of immigrants at two points in time: at their arrival to the United States (using their *reported* occupations in their country of origin and their *intended* state of residence), that we call *ex-ante* distribution, and once they have settled, or *ex-post* distribution, using their actual occupations and states of residence.

To better illustrate our empirical strategy, let us imagine that a state in 1900 had 10 percent of all Italians, 20 percent of all bakers and 30 percent of all Italian bakers in the United States but also only 5 percent of all carpenters and 2 percent of all Italian carpenters. Further assume

[3]See section 3 for a detailed description of the original data contained in the annual reports of the Commissioner of Immigration and the processed data used in this paper.

[4]Using shares of past immigrants in a decade before the period under study is relatively standard in the immigration literature (see for example Altonji and Card 1991). Our measure of past stocks could be considered as a proxy measure of a network, and we construct them this way to try to diminish the potential reflection problem, as emphasized by Manski (1993). We also control for a large set of fixed effects to capture other confounding factors.

that 100 more Italian bakers and 100 more Italian carpenters arrive to the United States in 1915 compared to 1910 and that in the 1920 Census, 150 more Italian bakers and 50 more Italian carpenters declare having arrived to the United States in 1915 compared to 1910.[5] The empirical exercise performed in this paper measures which of the three proposed counterfactual allocations is closest to the actual allocation of migrants by state, both ex-ante and ex-post.[6] Table 1 details what would be the predictions made by the allocating shares in each case. If individuals select their ex-ante location based only on the concentration of their ethnic group, we would expect that 10 more Italian bakers and carpenters would declare to be heading to this state at their arrival to the United States. Were we to see 20 bakers but only 5 carpenters, labor market factors would appear to be determining ex-ante choices.

Our results show that ex-ante decisions are particularly related to the presence of individuals of the immigrant's ethnic group irrespective of their occupations. In contrast, ex-post decisions, taken in the years shortly after arrival are strongly influenced by the presence of individuals of one's own ethnic group who also have the same set of skills as the migrant. We present an example graphically displaying this result in Figure 1. One can see in that figure that the change in the number of migrants declaring to be moving to that state is best matched by predictions constructed by ethnicity only. On the other hand, the change in the number of migrants actually living in the state is more precisely predicted by allocating national flows by ethnicity *and* occupation. These results are qualitatively similar whether we attempt to match the changes in the flow of immigrants over time in each state by ethnicity, by occupation or by the combination of both (following italians, carpenters, or italian carpenters, respectively). Results are also robust to alternative definitions and sample selections.

Overall, our estimates are consistent with a situation where networks serve as insurance upon landing with ethnic benefits and demand for skills playing a larger role later on. Moreover, these results are not consistent with simpler alternatives that ignore the dynamic aspects we uncover when we compare the importance of these three counterfactual distributions for immigrants who

[5] Since our analysis uses time fixed effects, we here explain the empirical exercise using changes from one year to another instead of levels.

[6] We perform the exercise for all states, ethnicities and occupations simultaneously within the limits of the data as explained in Section 3.

have just arrived to the United States (ex-ante distribution), those who have spent a few years in the United States, and those who have spent more time in the country. We observe that the difference between the ex-ante and ex-post distribution is more marked for immigrants recently arrived to the United States than for those who have been residing for longer periods. We also find additional support for our hypothesis that these allocation patterns are transitory and linked to the ability to participate in the labor market. First, the pattern highlighted above is especially visible for immigrant groups that are likely to face more complicated situations in the labor markets shortly upon arrival (immigrants who report occupations with low occupational scores[7] and immigrants from non-English speaking groups[8]). Second, the pattern we observe is most marked for occupations that produce "tradeable" goods, which is logical as providing employment in the same occupation might be costlier in terms of wages for individuals producing non-traded goods.[9]

Besides from changes in the occupations of the migrants, which is the mechanism we highlighted and emphasized before, two other mechanisms could lead to observing a difference between ex-ante and ex-post distributions: selective return migration, and changes in the location decisions of migrants. We explore the two alternatives in the paper but we do not find evidence they play an important role, and thus conclude it is unlikely they can explain the differences we see in the distributions. While we cannot conclude that any of the estimates necessarily measure a network effect, the pattern uncovered is consistent with the proposed role networks may play on the labor markets for immigrants.

Related Literature. The results presented in this paper contribute to our understanding of the factors that affect the location and occupational decisions of migrants, including ethnic networks

[7]Using occupations at origin to predict occupation distribution of immigrants at destination has been used in the literature. For example, Friedberg (2001) uses this approach in her study of Russian immigration to Israel. Our results imply that such an strategy will be more successful for more highly skilled immigrants, especially if used to predict occupations long after arrival to the host country.

[8]Related results have been observed in the context of the recent immigration wave to the United States. Peri and Sparber (2009) show that differences in communication skills are related to the sorting of native and immigrant workers into different occupations.

[9]Alternatively, it could be costlier or harder for previous immigrants working in non-tradeable occupations to provide employment if their total output is determined by local demand for the services and goods. This is true even if wages are equalized across states for equivalent workers.

and labor demand conditions. While Bauer, Epstein, and Gang (2005) study the role of networks on location choices of migrants in a static framework, our results shed some light on the dynamics of this pattern, something that has been explored theoretically (see for example Calvo-Armengol and Jackson 2004, 2007) but where empirical evidence has focused on the resulting segregation and workplace concentration. Munshi (2009) argues that networks may be able to break occupation-based traps and presents evidence from the Mumbai diamond industry supporting his hypothesis. Edin, Fredriksson, and Aringslund (2003) suggest, using an exogenous placement policy of migrants in Sweden, that living in a ethnic enclave benefits lower-skilled migrants, a finding that consistent with our results but Aslund (2005) look at the initial and subsequent location choices of migrants in Sweden and finds little variation over time in the factors influencing the decisions, in contrast to our findings. Most closely related to our paper, Beaman (2011) exploits the allocation process of refugees resettled in the United States to study these effects on employment and other labor market outcomes. She finds evidence that ethnic networks influence the access to local labor markets of newly arrived refugees, and that the effect depends on the "age" of the network: while the size of the recently arrived cohorts of refugees negatively affects the outcomes of the current arrivals, the size of older cohorts (two or more years before the current refugees) improves their outcomes. Our contribution is complimentary to Beaman (2011) as we argue that the role of networks itself changes as a migrant spends more time in the United States, thus looking at the dynamics as cohorts "age" rather than the differential effect of the cohorts already present at the moment of arrival.

This paper is also related to the literature that studies the performance of immigrants in the United States around the turn of the 20th century. Our results suggesting that the ethnic-occupation network becomes less relevant as the immigrants spend more time in the United States are consistent with the evidence presented in Minns (2000), who finds evidence of positive wage growth within cohorts of immigrants arriving during the same period we study. Furthermore, Minns (2000) also reports that immigrants showed high mobility into well-paid occupations which could be because migrants return to the occupation they were performing before their arrival, as we postulate.[10] The work of Abramitzky, Boustan, and Eriksson (2012) also explores the

[10]See also Hatton and Williamson (2005); O'Rourke and Williamson (1999) for a more comprehensive view of

same period but looking at the selection process from the country of origin, not by location or occupational choice at arrival.

Finally, our results improve our understanding of the occupational and geographical concentration of immigrants. Our results may justify why new immigrants are more likely to choose the same occupation previous immigrants from the same country have chosen, and that those who choose these occupations perceive a benefit in their earnings too as documented by Patel and Vella (2007) for the United States and Chen, Jin, and Yue (2010) for China.[11] However, all these studies are constrained to look at occupations once the individual has migrated and thus cannot distinguish whether this pattern is driven by individuals of similar skills selecting the same location or by immigrants selecting the same occupation once they have migrated. Ours results also offers some support for the use of local variation in networks strength as a source of exogenous variation to estimate the impact of immigration (see for example (Card 2001; Card and Lewis 2005; Cortés 2008; Cortés and Tessada 2011; Peri and Sparber 2009) among several others) as they suggest that while the ex-post location of immigrants may be more endogenous to local labor market conditions than the ex-ante one, this occurs also because immigrants change their occupation in response to the relative advantage of their networks and not only because they re-optimize their location decision based on current labor market conditions.

Layout. The remainder of this paper is organized as follows. Section 2 sketches a framework to help us better understand the factors that could influence the network dynamics of immigration. Section 3 describes the data and section 4 explains the empirical methodology and the results are presented in section 5. Section 6 then explores what could be the reason behind the pattern identified in the previous section. Finally, in the last section we summarize the results and offer some conclusions.

globalization, in general, and migration, in particular, during this period.

[11]See also Federman, Harrington, and Krynski (2006). Munshi and Wilson (2008) study the connection between ethnic networks when the American Midwest was first settled and occupational choice today.

2 Motivating Theory

One important element in our empirical analysis is the difference in the patterns of occupation and geographical location according to the timing of the decision and the relative strength of the different channels/networks at the moment the decision is taken. We introduce a simple framework highlighting the key elements that affect a migrant's decision about geographical location and occupational choice, and the moments when this decision is made. The main empirical hypothesis are then motivated using this framework.

2.1 (Basic) Framework

Assume a migrant from ethnic group j and *skills* o, i.e. she worked in occupation o in her country of origin and thus has only that exact set of skills, has just arrived to location s. The immigrant lives for two periods and discounts the future using a discount factor β. We assume migrants are risk-neutral, although risk-averse behavior does not change the main implications of our framework.

The labor market for immigrants at destination s functions as follows. The wage (w_{sqo}) offered to individuals in occupation q with skills o in location s is a function of the match between the occupation and the skills, reaching a maximum when $q = o$. Furthermore, we assume that the wage is higher for occupations that are in high demand in s, which will be reflected by a higher past concentration of individuals in that occupation N_{sq}. We also assume that frictions in the labor market are more important for immigrants than for natives (or immigrants that have stayed in the country for one period already), and that these frictions make it difficult for an immigrant to find employment unless they have some referral from their own network. In particular, let's assume that the probability for a migrant of ethnicity j of finding employment in occupation q in state s upon arrival is given by $P\left(N_{jsq}\right)$, where N_{jsq} denotes a measure of the relative strength of the migrant's ethnic network in occupation q in state s, such that $\sum_q P_{jqst} = 1$ for a given (j, s) pair.[12]

[12] In our empirical work we will approximate it with the stock of migrants of one's ethnicity and occupation in that state.

As the migrant lives longer in her new environment, her knowledge of the labor market increases and her capacity to find employment given her skills may become less and less dependent on her network. If such learning occurs, which we will denote by the parameter $\alpha = 1$, she will be able to pick the occupation in her state which will give her the highest income, most probably the one corresponding to her own skills. If there is no learning ($\alpha = 0$), then she will remain in the occupation she obtained in the first period.

The migrant's utility for a given s is

$$U(j,s,o) = \gamma(N_{js}) + \sum_{q} P(N_{jqs}) \left(w_{sqo} + \beta \left[\alpha(\max_{q'} w_{sq'o}) + (1-\alpha) w_{sqo} \right] \right) \tag{1}$$

where $\gamma(\cdot)$ represents benefits of having individuals of one's ethnicity which are unrelated to labor markets, such as having restaurants, temples, etc.

The migrant's location decision s^* can then be written as

$$s^* = \arg\max_{s} \quad U(j,s,o).$$

From this simple framework, the following general conclusions can be drawn. First, if the role of networks in the labor market is limited ($P'(N_{jqs}) \approx 0$), we should observe that migrants would locate where they have fellow countrymen, regardless of their occupation and because of the non-labor market benefits of ethnic networks, and where there is demand for their skills (captured by the stock of past migrants in the same occupation). In this case the distribution of immigrants, by state (s) and occupation (q), would then be better approximated using the distribution of previous immigrants of the same ethnic group, and this would not change as they spend more time in their destination.

Secondly, if $\alpha = 0$, migrants would elect a location where they have a large number of their countrymen exercising their own occupation as this will be their only way of working with their skills from their own country in the United States. Once more, this conclusion should be the same whether we consider the location choice of migrants based on their skills o or their actual occupation once they have migrated q. If $\gamma(\cdot)$ is large enough, pure ethnic networks will also play a role in the location decisions. Furthermore, while we may expect the importance of ethnic externalities to decrease as the migrant spends more time in her adoptive country, the *relative* importance of each element of the model is unlikely to be influenced by the time spent by the immigrant at her destination.

However, if $\alpha > 0$ and $P'(N_{jqs}) > 0$, a different location pattern could emerge. Conditional on the migrant's initial skill set o and ethnicity j, the probability that a given state will be selected will depend on the size of the ethnic network in that state, the wage that one could earn in the occupations that are already taken by individuals of the network with one's skills, and the general attractiveness of that state to migrants with similar skills sets (given the anticipation that in the second period, such employment will become available).[13] Conditional on the migrant's first occupation in the United States $\tilde{q}$, more migrants will be found (in the first period) where individuals of one's ethnic network who perform that occupation were already settled, i.e. the new migrants' occupation distribution will be influenced by the occupations that are more popular among previous migrants from the same ethnicity, this it what we call the *ethnic-occupation specific* effect. Without that network, a migrant would be unable to find employment in that sector. More migrants would also be found in locations where this occupation $\tilde{q}$ is well rewarded as the migrant would have been unlikely to select a state where the wage in the first period would be, in expectations, low. Finally, conditional on the migrant's occupation in the second period (which will be either o or $\tilde{q}$), migrants should be found in states where the wages for that occupation are highest.

Assuming that the overall popularity of a location for past migrants in a given occupation group is an acceptable proxy for the wage that a migrant can earn in that occupation[14], this simple model would suggest that we would observe the following pattern if the mechanism we've described is at play:

1. Based on their ex-ante skills, migrants would select locations where many migrants of their ethnic network (which provides with help, information and assistance during shortly after

[13] If the agents are too optimistic about the chances of obtaining a job in their own occupation this will imply a bias towards picking states based on the demand for their own occupation.

[14] Naturally, the wage would be set by factors influencing both the supply and the demand but we'll take labor supply as being a good indication of the attractiveness of a location for a particular occupation

arrival), and of their skills have previously located (which we take as a proxy of the demand for their set of skills).

2. In the first years in their adoptive country, however, we would find that migrants would be located in states where many migrants of their ethnic network (and less importantly, other previous immigrants) also share their *current* occupations.

3. After some time in their new country, we would find that migrants would be located in states where their current occupation is well rewarded but not necessarily where the members of their ethnic network practice the same one.

4. This pattern should be more marked for groups where α will be large or where the loss in human capital associated with switching occupations may not be large.

These elements will thus be the key insights we will be looking for in the empirical sections that follow.

3 Data Description

In order to study the dynamic role of networks in the location and occupational choices of immigrants to the United States in the framework presented in the previous section, we need to measure the immigrants' "skill set" at entry and their subsequent occupational choices as well as their location choices over time. Without detailed individual data, we collect data on gross flows constructed with information recorded at entry or arrival to the United States, and net flows recorded ex-post. All this information is provided by the combination of two main data sources: the United States Census for the ex-post data and the Report of the Commissioner of Immigration (henceforth RCI) for the information at entry to the United Sates. Data from the United States Census is taken both from the Public Use Micro Sample (PUMS), as compiled by Ruggles et al. (2010), and from the original published summary tables, which contain data for the first year of each decade.[15]

[15]The summary tables include tabulations based on the full sample of each Decennial Census.

3.1 Administrative Data from the Commissioner of Immigration

The RCI was published annually from 1899 to 1932 (except for 1931) and presented *summary* tables constructed using micro data from the questionnaires each immigrant coming to the United States had to answer.[16] For each year, immigrants are classified according to their ethnicity, self-reported occupation in the country of departure and their *intended* state of residence; this information was originally taken from the individual data each of them had to report when boarding at origin and/or when arriving to the United States.[17] Many may fear that the answers provided regarding the intended state of residence were nothing more than a random answer provided by the migrant and not related to the migrant's real plans. Two elements make us believe differently. First, in some cases, the immigration authorities would purchase a train ticket to the declared intended location and make sure the migrant was boarding the train leading to that destination (CITE NEEDED HERE). Second, we compared cohorts arrived the year before the Census (1909, 1919 and 1929) by their actual location and their intended location. We find that we can explain about 40 to 45 percent of the cross year variation in the actual flow to each destination by using the flow by intended state of residence, which, given the rate of return migration and the effort involved in matching ethnicities detailed before, we take as implying that the intended state of residence was a meaningful information provided by the migrants.

Using the tables presented in the RCI, we can create an annual series of gross entry flows for each ethnic group according to their intended state of residence, for each ethnic group according to their occupation in the country of departure, and for each intended state of residence according to their occupation in the country of departure. The three-way categorization, i.e., flows by ethnicity and occupation for each state, is not published in the RCI. Overall, we have data available for all states and territories in the United States for all years between 1899 and 1930 and for 1932, with a total of 75 occupation categories, grouped in 3 major categories, and 42 ethnic

[16] We have tried locating these tables for years 1932-1940 without success. Furthermore, while part of the micro-data may be available at the National Archives or a similar location in paper format, it would have been prohibitively expensive and time consuming to enter the information it contains. Hopefully, this information will be available in the future.

[17] It is our understanding that there was little incentive for a migrant to misreport their answers as this had no bearing on their acceptance to the United States, except in the case where their answers when boarding the vessel and at the port of entry were different. It is only in the early 1920s that entry restrictions affecting most of the potential immigrants were effectively imposed in the United States.

groups. All the information taken from these reports is labeled as "administrative data" in the rest of the paper and was digitized for the purpose of this paper.[18]

3.2 Census Data

The Decennial Census collects data every 10 years, and includes data on country of birth and year of first arrival in the United States (until 1930). From the 1900, 1910, 1920 and 1930 Census, we obtain information on country of birth, labor market status, occupation and state of residence for all those who are surveyed. Our main sources for the Decennial Census data are the 1% samples, plus the 5% sample for 1900, publicly available through the Public Use Micro Samples.[19] Variables measured from any of these two sources are referred to as "Census data" or "IPUMS data" in later sections.

We define an immigrant as anyone who reports being born outside the United States according to the value recorded in the *birthplace* variable in the Census. Using this definition of immigrant, for each Census we compute the stock of immigrants in each state and group them according to their occupation and country of birth. Also, with the information on year of first arrival to the US, we are able to create net flows by country of birth and actual state of residence and occupation. In order to improve accuracy in the calculation of shares and to avoid small-cell measurement problems, whenever possible we use the publicly available summary tables which were constructed using the original data set of the 1900 Decennial Census.[20]

We want to emphasize one important difference between the administrative data and the Census, and that is the timing of the information with respect to the moment of arrival to the United States. In the administrative data the information on occupations is collected before arrival, "ex-ante", while that in the Census reflects the occupations after agents have interacted upon and after arrival, hence we label it "ex-post". This is also true for the state distribution. Furthermore, while the Census data measures immigrants who have settled in the United States

[18] To the best of our knowledge the disaggregated data contained in the RCI has not been used before at this level of detail.

[19] Available online at http://usa.ipums.org/usa/.

[20] The 1890 Decennial Census micro samples are not available yet. While the summary tables are available, there are significant differences between the classification of occupations and countries of birth used in them and those used in the samples for subsequent years.

(or have at least stayed until the Census date), the administrative data measures entry flows and thus could differ from the Census measures if there is return or internal migration.[21]

The RCI and the Census use different classifications for occupations and for the origin of the immigrants. The RCI uses ethnic groups as classification, but the Census data records birthplace. Also, while the large occupation groups are relatively similar, specific occupations are grouped in different ways and some specific categories are used only by one of the two sources. In the appendix A we present a detailed description of the matching procedure used to combine both data sources.

3.3 Stylized Facts and Summary Statistics

The late 19th and early 20th centuries were periods of massive migration, with the U.S. and other countries in the *New World* receiving significant inflows. For example at the turn of the 20th century around 15% of the United States population was foreign born, a number that remained very high until 1930.[22] Annual gross immigration flows of more than a million people per year are observed between 1900 and 1914, a trend that was broken with the onset of the war in Europe. After the war, a significant recovery in the flows is observed but they decline again after the introduction of changes in the immigration: the 1917 Immigration Act, the 1921 Emergency Quota Act and the 1924 Johnson-Reid Act. By the end of the 1920 decade and the beginning of the Great Depression era the gross immigration flows are small. While we can consider the limits on migration before 1917 to be almost non-existent, they increased progressively over this period.[23] These restrictions particularly limited the entry of "unskilled" workers and favored family reunification. Furthermore, while a majority of the immigration over this period was European (or from countries with European ties such as Canada and Australia), an increase in the number of Hispanics migrants was observed after 1920. One important dimension in which

[21] Mortality rates is another reason why these two flows could differ.

[22] As a reference, using data from the 2000 Census micro sample we calculate that the fraction of foreign born in the United States lies around 11%; according to the 2007 and 2008 American Community Survey the same fraction lies slightly above 12%.

[23] See O'Rourke and Williamson (1999), Hatton and Williamson (2005) and the references therein for a more detailed description of the patterns of immigration for this period, including data for other countries besides the United States, and of the possible causes behind the protectionist backslash. Also see Bandiera, Rasul, and Viarengo (2010) for an analysis of the validity of the immigration and emigration data for this period based on flows recorded at Ellis Island.

immigrants from different countries differed was in the return and seasonality patterns; while for some groups return migration and seasonal migration was not particularly prevalent, other groups had important return rates, which were close to 30% for Spaniards and Italians between 1890 and 1914 (see O'Rourke and Williamson 1999).

Geographic clustering is definitely visible in this period: we find in both data sets, annual immigrations reports and decennial Census, an extremely high level of geographical clustering by ethnic group. States popular among immigrants today were also popular in the early 20th century: New York, Texas, California, etc. In Table 2 we present the top 3 states and the percentage going to the 10 most popular states for each ethnic group included in the previous figures. We can observe that for all groups in all periods at least 3 out of 4 immigrants went to one of the top 10 destinations of the group; furthermore, it is also evident that there is a significant degree of persistence in the top 3 destinations in the Census and in the administrative data although there is no perfect coincidence between the top 3 intended states and the top 3 actual states of residence for each of the ethnic groups. Furthermore, this appears to be fairly constant over time: the geographical concentration of the stock of immigrants in 1900 generally reflects that of the ex-ante decisions made by subsequent migrants as well as their ex-post decisions. Geographical clustering appears to be larger in ex-ante than ex-post decisions, suggesting that more individuals turn to less traditional destinations once they set foot in their country of adoption or that those who eventually decide to stay in the United States are not necessarily those who picked the most popular destinations. However, this difference is also decreasing with time as the last migration cohorts of this period become more and more concentrated ex-post in the United States.[24]

We also document something that has often gone unnoticed in some previous studies which is that there is also a very important occupational clustering by ethnicity. Table 3 shows the pattern of occupational choice according to the declared occupation in the United States and according to the occupation in the country of origin. We observe there is also high concentration, with the 10 most popular occupations for each ethnic accounting for 65% or more of the total members of each group. While some occupations are particularly important for all immigrants (such as

[24]This change could be related to the modifications to the immigration laws introduced after the end of World War I which not only reduced immigration overall but also change the composition towards skilled workers and family reunification.

laborers and farm laborers), there is also evidence of ethnic-specific clustering on different skills. This specialization in a given skill set appears also to be relatively constant over time. Once more, ex-ante occupational clustering is larger than ex-post but the difference remains more or less constant throughout the period. When we compare the Census and the administrative data we can observe that in this case we see that unskilled occupations such as laborers, farm laborers and servants are among the most popular ones in both data sets. We also observe that for most of the groups there is a certain level of agreement between the top three categories in each of the data sets, we also see that certain occupations are more popular ex-post than ex-ante, e.g., miners, hotel keepers, manufacturers, etc.

Finally, the United States tend to demand different occupations in different states, leading to geographical clustering among individuals of a given occupation. This concentration is slightly less marked than the two previous ones as the top ten states by occupation in 1900 captured around 70 per cent of all workers in that occupation, but it is still noticeable.

4 Empirical strategy

We explore whether we can find any evidence of the empirical regularities mentioned in section 2 by attempting to contrast the "fit" of counterfactual distributions built using different measures of networks to the actual ex-ante and ex-post distribution of immigrants. We are constrained in our empirical strategy by the fact that we do not have access to micro-level data with the individual records of migrants entering the US during this period. However, we are, as we explained in section 3, capable of constructing flows by ethnicity (j), occupation (o), state (s) and time (t), which we will denote as n_{jost}, measured at the moment of arrival to and after settling in the United States.

We will approximate the (relative) strength of networks in the same way traditionally used in the immigration literature, i.e. using the distribution by states of a given group of immigrants in a base period. Specifically, we measure our networks mostly in 1900, and include in the regressions only the immigration flows from 1905 to 1930 to avoid too much proximity between our flow measure and the shares employed in the prediction. This argument is further strengthened by the fact that immigration had significantly dampened between 1890 and 1900 and thus that our measure of stocks probably reflects the location choices of immigrants who arrived mostly before 1890.[25] Clearly, these measures could also capture something else than networks: Scandinavians migrants may all select the same type of state not because they like living together but because the climate of some parts of the United States is more similar to their own. However, most of these alternative explanations would not generate the change in patterns will we demonstrate across ex-ante and ex-post data sources. Thus, our empirical exercise will consist in displaying a pattern similar to the one suggested by the framework and not by alternative hypotheses, not in necessarily interpreting all coefficients found as valid estimates of "network effects". Finally, our interest will lie more in the comparison of various proxies for networks than in the significance or magnitude of a given coefficient. If all proxies suffer from the same bias, the comparison would still be valid.

In order to control for other factors such as a particular match between an ethnicity and a state (because of a particular endowment of natural resources for example), we include a number of fixed effects to attempt to capture this and other elements.

The framework presented in section 2 suggests an empirical regression of the following type:

$$n_{jost} = \alpha A_{js} n_{jot} + \beta B_{os} n_{jot} + \gamma C_{jos} n_{jot} + \mu_{jos} + \nu_{jot} + \omega_{jst} + \eta_{ost} + \varepsilon_{jost} \tag{2}$$

where n represents the immigrant flow and the shares A, B and C represent different ways of allocating the national flow n_{jot} to different states, related to distinct types of networks.

Specifically, we define

$$A_{js} \equiv \left(\frac{N_{js}}{N_j} \right) \tag{3a}$$

$$B_{os} \equiv \left(\frac{N_{os}}{N_o} \right) \tag{3b}$$

$$C_{jos} \equiv \left(\frac{N_{jos}}{N_{jo}} \right) \tag{3c}$$

where N represents the stock of immigrants with certain characteristics already in the United

[25]To the best of our knowledge there is no available digitalized sample from the 1890 US Census.

States. Thus, A_{js} refers to the share of individuals from one's ethnic group who elected to live in state s in the past. The second share, B_{os} refers to the geographical distribution of immigrants who share one's occupation. Finally, C_{jos} represents the share of individuals from the same ethnicity *and* the same occupation who lived in a particular state in the past. We will refer to A as measuring an ethnic effect, to B as a labor/occupational effect and to C as an ethnic-specific occupational component. [26]

The regression equation includes fixed effects for all triple interactions between ethnicity, occupation, state and time. This allows to control for all possible confounding effects that are affecting the immigration rates of a particular sub-group or the overall effects in a particular state. We have used fewer fixed effects and the results are similar. Also, the standard errors are clustered by ethnicity-occupation-state cells although very similar results were obtained with much more aggressive clustering.

Intuitively, this regression attempts to test whether a state has a differential growth in a particular occupation between two different ethnic groups compared to a different state for one of the three following reasons. Is it because that state was popular for a particular ethnic group and that ethnic group has a growth in the number of people of that occupation coming from that ethnic group? Or because that state was popular for a particular occupation and there's now an increase among that ethnic group of that occupation? Or because that state was popular for individuals of that ethnic group and that occupation and there's a growth in the number of individuals that satisfy both categories? Each one of the three components defined in equation (3) attempt to capture each one of these reasons separately, thus enabling us to compare them with the actual observed distributions.

The theoretical framework presented in section 2 would suggest that when s represents the intended state of residence and o the skills with which an individual arrives in the United States, we should obtain estimates of α and β to be positive and significant while γ should not be. On the other hand, when s represents the actual state of residence and o the occupation in which an

[26] One could be worried about the potential collinearity between these three proxies. In order to verify that the results are not corrupted by this problem, specifications where the ethnic share was built excluding individuals of the same occupation and the occupation share was built excluding individuals of the same ethnicity were also run and the results are qualitatively and quantitatively very similar.

immigrant is working in the United States, one would expect to find the parameter γ to be large and significant with little importance of either α or β, and that the magnitude of the estimated coefficient should be smaller when we consider the case of immigrants that have spent more time in the United States.

Unfortunately, n_{jost} is only measured ex-post in our data, that is where s represents the actual state of residence and o one's occupation in the United States. While we will run this regression, it will not provide us with the ex-ante, ex-post comparison we desire. Because the flows n_{ost} and n_{jst} are observed both ex-ante and ex-post, it will be useful to estimate a simpler version of equation (2) by summing over all ethnicities/occupations in a given cell. Denoting the variable over which one is actually summing by k' and the other by k, where $k, k' = j, o$, one obtains

$$n_{kst} = \alpha \sum_{k'} A_{js} n_{jot} + \beta \sum_{k'} B_{os} n_{jot} + \gamma \sum_{k'} C_{jos} n_{jot} + \mu_{ks} + \nu_{kt} + \omega_{st} + \varepsilon_{kst} \tag{4}$$

where the left-hand side variable n_{kst} represents the flow of immigrants arriving in period t, either intending on residing or living in state s, depending on whether we use the administrative or the census data, respectively, and from ethnicity (when $k = j$) or occupation (when $k = o$).

Regression equation (4) again contrasts the strength of the three different networks and it does so by comparing the correlation between the predicted distributions across states of individuals of characteristic k to the actual distribution observed in the data. Exactly as before the variables A, B and C are our measures of the strength of the pure ethnic, the occupation/labor market, and the ethnic-occupation specific networks, respectively.

In this case, all the double-interactions between characteristic k, time t and state s are included and standard errors are clustered at the ks level. We will estimate these equations using only cells where $k = j$ and where $k = o$ separately as well as jointly. In the latter case, the estimation constrains the parameters α, β and γ to be the same but allows for a more efficient estimation as the parameters ω_{st} are common to both sets of regressions and thus estimated using more data points.

One limit to the last regression equation is that the network measure C can only be obtained from one data source (IPUMS) because we require information on the distribution according to the *jos* cells over time. To verify that this is not driving our results, another simplification can be made and the following equation is estimated:

$$n_{kst} = \alpha \sum_{k'} \left(A_{js}\right) n_{jot} + \beta \sum_{k'} \left(B_{os}\right) n_{jot} + \mu_{ks} + \nu_{kt} + \omega_{st} + \varepsilon_{kst} \tag{5}$$

once more for $k = j, o$. In this regression equation, we simply exclude the role of ethnic-occupation network. The impact of these will be bundled up with that of ethnic effects (in A_{js}) or occupational factors(in B_{os}). While the results obtained here are less informative, they give us an opportunity of verifying the robustness of our results to the use of different measures of networks.

The above regression includes a full set of fixed effects to control for any ethnic/occupation-state, ethnic/occupation-time and state-time confounding factors. However, we also used a smaller set of fixed effects and found similar results. The standard errors are clustered by *ks* but similar results were obtained when clustering by state only. All regressions are un-weighted.

Note that because the mean and standard deviations of our predicted flows measures are all roughly similar, we can simply compare the coefficients to understand the relative strength of the match from each of the counterfactual distributions.

5 Results

5.1 Ethnic vs. Labor networks

We first begin our empirical exploration by contrasting the role of A_{js} and B_{os} as presented in equation (5). This strategy, although less informative in terms of the factors driving the location and occupational choices of migrants, has the advantage of allowing us to compare the results across a variety of alternative proxies for the relative strength of networks, which we will be unable to do for the rest of the analysis.

The results of this strategy are presented in Table 4 when estimating the flow of immigrants selecting a particular state and of a given ethnic group (in Panel A) and occupation (in Panel

B).[27]

In the first three columns, we attempt to explain the ex-ante location choices (as measured in the administrative data) while the last three use as the left-hand side variable ex-post decisions from IPUMS. The measures of predicted flows are constructed by allocating, using the two types of shares as detailed in equation (5), the national flow (n_{kt}) from the administrative data in the first three columns and the national flow as measured from the IPUMS in the last three. Finally, columns also differ by how the network measures are computed. In columns (1) and (4) we report results using the average share of individuals with a characteristic k or k' who declare state s as their intended state of residence for the first 6 years of the administrative data, that is from 1899 to 1904. In columns (2) and (5) we use the published Census summary tables to compute the number of individuals from a given ethnicity or from a given occupation residing in each state in 1900. This has the advantage of offering large sample sizes and reducing the probability of small sample bias. On the other hand, it also assumes that immigrants who are active in a given set of occupations are locating in the same states as their natives counterpart, because we do not have summary tables with employment levels by state and birthplace. Finally, Columns (3) and (6) compute the share of immigrants living in a particular state from a given ethnic group or a given occupation in the 1900 Census Public Use Micro-Sample (IPUMS).

The first three columns of Table 4 indicate that when trying to explain the ex-ante "intended" location decisions of migrants, the presence of individuals of their own ethnic group is particularly important. When we attempt to match the distribution of immigrants by ethnicity, we find that the past location of immigrants by occupation also contribute to increase the match with the actual distribution, although the magnitudes of the coefficient on the ethnic component are always larger than those on the occupational element.

On the other hand, when we explore the factors behind the ex-post location decisions of individuals in the IPUMS, we find quite a distinct pattern, as observed in the last three columns of Table 4. In this case, the importance of having individuals of one's occupation who previously

[27] These tables include all individuals who report an occupation but very similar results were obtained when all individuals, including those reporting no occupation, were incorporated in the sample. Results that also include individuals without an occupation in the sample are available upon request from the authors. Similarly, in these regressions a period t corresponds to a 5-year period, with no overlap between two consecutive periods; we also explored the same specification using annual variation and obtained similar results.

located in that state becomes much more important and that of ethnic networks is reduced in both panels.

These results would thus suggest that when one attempts to explain why a particular state now has a larger set of individuals with a given set of skills claiming to be heading there, the fact that this state was particularly attractive to individuals with those skills in the past and that more of them are now landing in the United States is of little relevance. Instead, our results suggest that it is because a larger number of individuals with those particular sets of skills are coming from ethnic groups that had a large fraction of their previous migrants already living in that given state by 1900.[28] On the other hand, the increase in the number of immigrants performing a particular occupation in a particular state ex-post appears to be driven by the fact that more immigrants from that occupation are now within the United States and this particular state was attractive to that occupation in the past. We find little evidence that this is linked to the fact that this state was attractive to a given ethnic group in the past and that this ethnic group now has more individuals reporting working as that occupation after their arrival.

The results appear to be fairly robust across the shares as measured from the IPUMS or the Census tables, even if the occupational distributions from the Census table include all individuals in the United States, natives or immigrants. The counterfactual distributions built with shares from the administrative data appear to be more weakly related to the actual flows than the ones built from Census (table or micro-sample) data. This is comforting because it indicates that immigrants were less likely to follow the location decisions of recent migrants (as measured by flows) than the ones of established migrants (as measured by stock). This would suggest that our results are unlikely to be driven by contemporary shocks since in that case, one would expect the flows to have a higher predictive power than the stocks. Furthermore, this is consistent with the empirical results of Beaman (2011).

We also confirmed that the effect we find is not simply driven by similar patterns among immigrants, regardless of their ethnicity. First, all regressions include fixed effects for the in-

[28]Notice that our fixed effects allows us to remove the variation that comes from permanent characteristics that create an intrinsic match between states and occupations or ethnicities. For example, we rule out that our results are driven by a state having during all this period a dominant industry that offers a perfect match for a country of origin that happens to have the same dominant industry.

teraction of state and time, thus capturing all characteristics specific to a particular location at a given moment. A more stringent robustness test is provide in Table 5, where we add as additional regressors, the counterfactual distribution one would obtain by allocating each period's flow using similar ethnicities. Specifically, we include all ethnicities within a broader ethnic group as defined in Table A-2 except one's own. For example, Irish immigrants are allocated using the geographical distribution from 1900 IPUMS of other British Isles, Australia and Canada. In almost no case are these variables entering significantly in the regression equation and in all cases, the main results remain extremely similar. The results are fairly similar across samples and suggest that it is the location choices of one's particular ethnicity which influences locations decisions and not that of fairly close ethnic groups.

5.2 More Disaggregated Measures of Networks

Having shown in the previous section that the results are more or less robust to the way the shares are constructed, we now turn to the estimation of the impact of our three distinct proxies of networks: ethnic, occupational and ethnic-specific occupational networks. In this case, all shares will be computed from the IPUMS as this is the only source that allows us to include the third factor.

Table 6 presents our baseline results for four distinct samples. The first three panels correspond to the results of estimating equation (4) for the location choice of individuals by ethnicity (when $k = j$), by occupation (when $k = o$) and jointly (when $k = j, o$). Finally, panel D shows the results of estimating equation (2). This table presents once more the regressions where ex-ante decisions are explored in the first two columns and those about ex-post decisions are in the following two. Since the flow by ethnicity, occupation and state is only available in the IPUMS data, the last panel only includes results about ex-post location decisions.

The results from column (1) are remarkably similar across samples and all emphasize that the key factor driving the ex-ante location decisions of individuals lies in the presence of individuals in 1900 that were of the same ethnicity than the newly arrived migrants. Allocating newly arrived migrants based on the past distribution of their occupation appears to only contribute to match the actual distribution by ethnicity, but not by occupation or in the joint estimation. Finally, in all regressions, the ethnic-specific occupational factor is small and insignificant.

Our theoretical framework suggests that while an individual would pick ex-ante a location based on the presence of an ethnic network, that location would be particularly attractive to her when the occupations within this network are more similar to hers. This is explored in column (2) where an interaction term is added which captures the absolute difference between the average occupational score of the network and that of the occupation of the migrant. As is expected, this is negative in all specifications and significantly so in the first panel, suggesting that while individuals select ex-ante locations based on the presence of a network of individuals of their ethnicity, they are particularly likely to pick one where the occupations of that network are similar to theirs (although not identical).

The pattern is strikingly different in columns (3) and (4) where the ex-post distributions are explored. The role of the ethnic factors disappears in this case and turns negative in all regressions. The counterfactual distributions based on the past location choices of immigrants of one's ethnicity are if anything negatively related to the actual distribution, conditional on the other two counterfactual distributions we constructed. The results presented in this table emphasize that the ex-post distributions appear to be much more closely correlated with our predicted flows based on occupational factors and more specifically, on the past location choices of immigrants sharing one's ethnicity and one's occupation.

Taken together, these results suggest that an increase in the number of individuals that *intend to reside* in a particular state, compared to the average for all states, is due mostly to the fact that this state was attractive in the past to ethnicities which are now receiving larger national flows. At the same, the results also imply that if we observe an increase in the number of migrants *actually* practicing a given occupation in a particular state, compared to the average for all states, this is due mostly to that state being popular in the past among individuals who practiced that occupation and who were of ethnicities that are currently experiencing larger flows. The magnitudes of the coefficients should be interpreted as follows. Taking the example of Italian bakers, these results would suggest that a state which had received 10 percent more immigrants of that ethnic group (irrespective of their occupation) in the past would attract 12 percent more immigrants of those newly arrived immigrants ex-ante. However, it is in a state where one found, in 1900, 10 percent more Italian bakers that one would eventually find 6-11 percent more of these immigrants ex-post, i.e. after they have settled in the United States.

Robustness Checks. We explore the robustness of these results in various ways. We first estimate these regressions for each of our nine broader ethnic categories separately. While the results are fairly noisy, they do not indicate that one ethnic group is driving the results more than another. Table 7 further explores whether the results depend on the sample over which these estimates are obtained. One may first be worried that our results may be sensitive to problems related to small cells. Columns (1) and (4) thus restrict the sample to groups whose national flows over the period were above the median. In Panel A, this corresponds to ethnicities with large inflows, in Panel B, to occupations with large inflows, and in the last panel, to ethnicities and occupations that were observed in large numbers (ex-post). The results from these regressions are extremely similar to the ones presented in the previous table, suggesting that our main empirical conclusions are not driven by small cells. Although not presented, we find that for small ethnicities, occupational groups are more relevant in driving the location choices and for rare occupations, ethnic networks appear to matter more. This is logical if one needs a network of a significant size to be able to derive any benefits from it or if the returns to a network are increasing with its size.

One could also be worried that after the introduction of limits on immigration, migrants may have had more of an incentive to provide false information to the immigration officer at arrival, thus contaminating our ex-ante measures. Furthermore, migrants arriving after the 1921 Act were more skilled and more likely to come for the purpose of family reunification, thus potentially explaining the importance of ethnic networks in that group. Columns (2) and (4) verify whether these concerns are founded by limiting the sample to immigrants arriving before 1920, that is before any immigration restrictions were imposed. Once more, these results are extremely similar to those presented in Table 6.

Finally, the state of New York was the largest recipient of immigrants and, given that our estimating equation is in levels and not in a logarithmic form, one could be worried that this particular state is driving our results. The results presented in columns (3) and (6) find little indication of this: excluding this particular state has very limited impact on the overall results.

6 Exploring the difference between ex-ante and ex-post decisions

The previous section explored the factors influencing ex-ante and ex-post location decisions of immigrants to the United States in the first three decades of the twentieth century. We find striking differences between the factors that are determining location decisions at their arrival in the United States and once they have settled. In the first case, ethnic networks and labor market effects appear to drive the ex-ante location decisions of immigrants. Ex-post, however, we find a very important role to be played by the presence of individuals of the same ethnicity *and* occupation, which we interpret as evidence of the role played by ethnic networks in the incorporation of immigrants to the local labor markets.

In the theoretical framework presented in section 2 we suggest that this result is compatible with newly arrived immigrants taking on the same occupations of individuals in their ethnic networks at their arrival because they are more likely to receive job offers, or to obtain referrals for positions, through them. In this sense, networks among immigrants become a tool to smooth the transition into the new labor market.

6.1 Differences Across Occupations and Ethnicities

If our hypothesis that networks are the reason behind this pattern is correct, then, as we explained in our motivating framework, we should be able to find some additional evidence consistent with it. For example, networks are more likely to play an important role among individuals who arrive in the United States with low levels of skills.[29] A doctor is much less likely to change his occupation than a general laborer, for example. Furthermore, higher-skilled individuals may only offer a low chance of referrals in the same area of specialization. This hypothesis is explored in the first two columns of Table 8 where we divide the sample between occupations

[29] Hellerstein, McInerney, and Neumark (2008) use matched employer-employee data and find that networks generated by residential proximity are relatively more important for minorities and less-skilled workers, particularly Hispanics.

that had an occupational score (based on the average wage in 1950) above or below the median, which we take to be a proxy for the skill level of the occupation. On the one hand, for highly paid occupations, ex-ante decisions are very closely related to the presence of an ethnic network but the ex-post decisions are driven by pure labor market considerations.[30] On the other hand, individuals of lower paid occupations display the reliance on one's ethnic-specific occupational network as presented above. Although not shown here, very similar results were obtained if one divides the occupations based on the RCI classification of occupations into "professionals", "skilled" and "unskilled". In that case, it is for individuals with occupations identified as "unskilled" that the pattern shown above is particularly marked, with both skilled tradesmen and, particularly, professionals relying ex-post more heavily on labor market considerations.

Columns (3) and (4) of Table 8 explore another distinction between occupations. We attempted to classify occupations based on whether their main product is a tradable or a non-tradable (see Table A-1 for which occupations were placed into each group). This is naturally a very gross matching as some occupations may produce a variety of goods, some which are traded and some that are not, depending on the industry in which they perform. Nevertheless, we would expect that the use of one's ethnic network as a transition tool may be much more likely to occur in the case of occupations where the output is traded, as the wage of these individuals would be less likely to be affected by the number of individuals performing that occupation in a given location. Referrals, for these individuals, would thus be less "costly". Column (3) clearly shows that the pattern we identified above is particularly visible for occupations producing traded products. Column (4) show that ex-post decisions of individuals working in "non-traded" sectors are more influenced by labor market concerns than by the presence of individuals sharing both one's ethnicity and occupation. Somewhat surprisingly, however, ex-ante decisions among that group appear to be somewhat correlated with the presence of an ethnic-specific occupational network. The difference in the coefficients between the two columns is significant at 10% level in Panel B and at 1% in Panel A.

Furthermore, learning about local labor markets might be particularly important for individ-

[30]The coefficients of interest are statistically different for both groups at 5% for Panel A (although not for any coefficient separately) and at 10% for Panel B.

uals who do not speak the same language as US natives. The first columns of 9 explore this issue by separating ethnic groups between those who spoke English (British Isles, Canada and Australia) and those who did not. The first group has limited variation that can be exploited and the results are much noisier in that case. Nevertheless, it seems that the pattern we identified in the overall sample is most closely related to that of non-English speaking individuals. English-speaking ethnicities do rely on the ethnic-specific occupational networks but they appear to do so consistently both at arrival and once settled, which is what would be expected if there is no learning involved in their adaptation process. The hypothesis of equal coefficients for both groups is not rejected in the ex-ante data but clearly rejected in the ex-post regressions.

6.2 Time since arrival

Our framework does suggest that if the lack of credentials or verifiable experience was one of the reasons why ethnic networks were important, then ethnic-specific occupational networks should become less relevant (i.e. reduce their correlation with the observed location and occupation choices) as immigrants stay for a longer time in their new country. Thus for a given cohort of immigrants the estimated effect of the ethnic-specific occupation network should be decreasing in the time since arrival to the United States.

The last two columns of Tables 8 and 9 explore this issue. The difference between ex-ante and ex-post decisions is particularly marked for individuals who have just arrived in the United States. It is for that group that we observe a very strong preference for electing a state of *intended* residence based on ethnicity and labor market considerations but a very strong preference for eventually residing in a state where individuals who share one's occupation and ethnicity can be found. For more established immigrants, we find a different pattern where all factors appear to determine almost equally ex-ante decisions but ex-post decisions depend much more on factors related to labor markets, although not particularly the ones related to one's ethnicity. An extremely similar pattern (if only maybe more marked) can be found in the corresponding columns of Table 8. The fact that the ex-ante decisions appear to be driven by similar factors is reassuring, suggesting that this is not because these two groups were somehow different at entry.[31]

[31]The hypothesis of equal coefficients cannot be rejected for either sets of parameters because of the noisiness of

Versión impresa ISSN: 0716-7334
Versión electrónica ISSN: 0717-7593

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA

Oficina de Publicaciones
Casilla 76, Correo 17, Santiago
www.economia.puc.cl

SMOOTH(ER) LANDING? THE DYNAMIC ROLE OF NETWORKS IN THE LOCATION AND OCCUPATIONAL CHOICE OF IMMIGRANTS

Jeanne Lafortune*
José Tessada

Documento de Trabajo Nº 427

Santiago, septiembre 2012

* **jlafortune@uc.cl**

INDEX

Furthermore, we have repeated this exercise adding data for the groups with even longer stays and found similar results to the ones we present; the effect of the ethnic-occupation network is diminished as the migrants spend more time in the United States, while the other two, the pure ethnic and the labor market effects, are strengthened.[32]

6.3 Alternative Channels

Two other factors could explain the difference between ex-ante and ex-post decisions. First, it may be that individuals without a group of past migrants sharing their ethnicity and occupation are more likely to leave the United States. As we discussed before, this is a period where return migration was a fairly important phenomena and thus could explain the patterns we observe. Since our measure of ex-ante flows include all immigrant entries into the United States but our ex-post measure involves only those who eventually stay, the difference in our results could be driven by selective return migration. In order to evaluate this alternative hypothesis, the first two columns of Table 9 estimate the same regressions but dividing the sample between ethnicities with high and low levels of return migration. We define a group as high return migration if the national (gross) flow of immigrants over the entire period of analysis obtained from the administrative data is, in percentage term, above the median compared to the net flow obtained from the Census data.[33] The results presented in Table 9 do not, however, suggest that the location choices of these two groups are extremely different. In both cases, the ex-ante distribution appears to be best approximated when the national flow is allocated according to the presence of individuals of one's ethnic group while the ex-post distribution appears to be best approximated when using the ethnic-specific occupational networks.[34] This does not seem to suggest that the selective return migration can explain the patterns we have highlighted above.

Also, the importance of the pure labor markets do appear to be more important for groups that have low levels of return migration, particularly for the ex-ante distribution, which is con-

some results. However, the test statistic is much larger for ex-post regressions than ex-ante ones.

[32]Results are available upon request.

[33]This definition is in agreement with Bandiera, Rasul, and Viarengo's (2010) assessment of the emigration patterns for this period. However, we also conducted the same analysis using the emigration flows from the administrative data to classify the ethnicities into the high and low return groups, and we obtained similar results.

[34]The equality of coefficients can be rejected in Panel A because of the difference in the second and third coefficient; however, we cannot reject equality of the coefficients in Panel B.

sistent with our framework. When individuals are forecasting staying in their new location for a long period of time, the quality of the labor market for their skills becomes more crucial.

Secondly, it may also be that individuals first locate (or say they will locate) based on general considerations but then relocate within the United States to a location where they have someone from their ethnic group who also share their occupation. The difference in the pattern observed would then not be driven by the fact that immigrants change their occupation in response to the composition of their networks but rather that they re-optimize their location decision to bring themselves closer to a network that can provide them with referrals in their given occupation, for example. To explore this, the ideal data would have included information regarding internal migration, as the one currently compiled by the Census. Unfortunately, for this period, no question on whether an individual has moved recently is available. However, the results presented above suggest that the pattern we observe is particularly strong for individuals recently arrived to the United States. If internal migration is an important explanation between the phenomena we just described, we would expect that individuals who have been in the United States for a longer time period would have had more time to re-optimize and thus that they would be the ones whose decisions would look more different than the ones at arrival. The opposite, however, seems to be found in the results we present. Thus, internal migration would only explain this pattern if individuals, at their arrival to the United States, moved temporarily to a location where their ethnic-specific occupational network is more important to then return to one where their skills are more valued later on.

Finally, we can further explore the role of internal migration by exploring the differences between IPUMS and administrative data. The major difference between a change in occupation and a change in the state of residence is that the former potentially affect the national flow being allocated (n_{jot}) in our estimating equations while the latter does not. Thus, if immigrants were mostly changing their state of residence but not their occupation in response to the presence of networks, the results of regressions where we attempt to predict the ex-post distribution by allocating the ex-ante national flows and vice versa should be identical. This is because the national flow n_{jot} should be the same when measured in the administrative data than in the Census data if individuals do not change their occupation upon arrival. Thus, the same pattern should be observed when estimating equations (4) or (5) whether or not the right-hand side variable is built using administrative or Census data. The result of this exercise is presented in Table 10. Each panel again represents one of our sample. In the first column, the left-hand side variable is measured ex-ante from the administrative data but the predicted flows are built using the national flows from the IPUMS dataset. The reverse is true in column (2). The results of that table are very different from those presented in Table 6. In the first column, that is when trying to match the distribution by intended state of residence using the ex-post national flow, we see that the importance of the ethnic component remains. This is not surprising since the change in occupation would not influence this result. However, in the second column, the results are very different from what was observed previously. In this case, the importance of ethnic-specific occupational factors are always absent. This is again logical if people of a given ethnicity change occupation which generates a different national flow n_{jot} to be allocated.[35]

7 Conclusions

We have thus presented, through various methods, evidence regarding the role played by networks and labor market characteristics in the determination of the location and occupation choices of immigrants in the United States in the first three decades of the twentieth century. We have shown empirical evidence consistent with the hypothesis that the presence of individuals of one's ethnicity, but not necessarily in the same occupation, is strongly influential in the *intended* location reported by an immigrant at her arrival in the United States. However, the *actual* location and occupation choices appear to be driven much more by considerations linked to the labor markets, in particular to the presence of individuals of one's ethnicity who also share one's occupation, a measure that we call ethnic-specific occupation network. We have also shown suggestive evidence that the reason behind this change in behavior lies in the fact that unskilled immigrants are likely to change their occupation once they arrive into the United States to benefit from the labor market connections of their ethnic networks.

[35] If the change in occupation was not a real change in occupation but rather a simple error in the matching of occupations, we would expect that this error of matching be independent of our proxies of networks, thus not generating the patterns highlighted above.

The findings in this paper help us understand the role networks play in the location choice of immigrants. While most studies have looked separately at location or occupation choices, our study combines both decisions and seems to suggest important interactions between the two choices. It also emphasizes the dynamic aspect of the process as immigrants spend more time in their destination country, a fact that has not been carefully documented before. It even seems to indicate that ethnic networks are particularly useful as a smoothing mechanism for newly arrived immigrants and are not something that migrants need to permanently rely upon.

These conclusions might also be relevant in shaping optimal immigration policy. If immigrants of different skill levels select their location and occupation in their new country using networks differently, this has implications for the impact that these immigrants can have on native and previous migrant workers. Understanding the mechanics of the effects and its connection to immigrant geographical and occupational concentration can thus shed light on potential measures aimed at improving immigrants incorporation into labor markets and to benefit native population.

While our specific data allows us to look at a historical period, our results might be relevant for today's immigration debate. The immigration wave of the early 20th century resembles that of today in some important aspects: migrants were less skilled than natives (or at least were perceived to be less skilled), they were perceived to be culturally different than the remainder of the population, and their arrival generated controversy over their potential adverse effects. More importantly for our results and analysis, immigrants today also share ethnic and occupational networks as defined in this paper. Whether the patterns we have highlighted here are also present in the current migration wave is a subject of future research.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102 (5):1832–56. URL http://www.aeaweb.org/articles.php?doi=10.1257/aer.102.5.1832.

Altonji, Joseph G. and David Card. 1991. "The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives." In *Immigration, Trade and Labor*, edited by John M. Abowd and Richard B. Freeman. Chicago, IL: University of Chicago Press.

Aslund, O. 2005. "Now and Forever? Initial and Subsequent Location Choices of Immigrants." *Regional Science and Urban Economics* 35:141–165.

Bandiera, Oriana, Imran Rasul, and Martina Viarengo. 2010. "The Making of Moden America: Migratory Flows in the Age of Mass Migration." URL http://www.homepages.ucl.ac.uk/~uctpimr/research/ellis_accounting.pdf. Mimeo, Harvard University, LSE and UCL.

Bauer, T., G. S. Epstein, and I. N. Gang. 2005. "Enclaves, Language, and the Location Choice of Migrants." *Journal of Population Economics* 18:649–662.

Beaman, Lori. 2011. "Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the U.S." *Review of Economic Studies* Forthcoming.

Calvo-Armengol, Antoni and Matthew O. Jackson. 2004. "The Effects of Social Networks on Employment and Inequality." *American Economic Review* 94 (3):426–454.

———. 2007. "Networks in labor markets: Wage and employment dynamics and inequality." *Journal of Economic Theory* 132 (1):27–46. URL http://ideas.repec.org/a/eee/jetheo/v132y2007i1p27-46.html.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19:22–64.

———. 2009. "Immigration and Inequality." *American Economic Review* 99 (2):1 – 21.

Card, David and Ethan Lewis. 2005. "The Diffusion of Mexican Immigrants During the 1990s: Explanations and Impacts." NBER Working Paper 11552, National Bureau of Economic Research.

Chen, Yuyu, Ginger Jin, and Yang Yue. 2010. "Peer Migration in China." Mimeo, University of Maryland–College Park.

Cortés, Patricia. 2008. “The Effect of Low-Skilled Immigration on U.S. Prices: Evidence from CPI Data.” *Journal of Political Economy* 116 (3):381–422. URL http://www.journals.uchicago.edu/doi/abs/10.1086/589756.

Cortés, Patricia and José A. Tessada. 2011. “Low-Skilled Immigration and the Labor Supply of Highly Skilled Women.” *American Economic Journal: Applied Economics* 3 (3):88–123.

Edin, Per-Anders, Peter Fredriksson, and Olof Aringslund. 2003. “Ethnic Enclaves And The Economic Success Of Immigrants - Evidence From A Natural Experiment.” *Quarterly Journal of Economics* 118 (1):329–357.

Federman, May, David Harrington, and Kathy Krynski. 2006. “Vietnamese Manicurists: Displacing Natives or Finding New Nails to Polish?” *Industrial and Labor Relations Review* 59 (2):302–318.

Friedberg, Rachel M. 2001. “The Impact Of Mass Migration On The Israeli Labor Market.” *Quarterly Journal of Economics* 116 (4):1373–1408.

Hatton, Timothy J. and Jeffrey G. Williamson. 2005. *Global Migration and the World Economy: Two Centuries of Policy and Performance*. Cambridge: MA: MIT Press.

Hellerstein, Judith, Melissa McInerney, and David Neumark. 2008. “Measuring the Importance of Labor Market Networks.” Mimeo, University of Maryland.

Lafortune, Jeanne. 2010. “Making Yourself Attractive: Pre-Marital Investments and the Returns to Education in the Marriage Market.” URL http://sites.google.com/site/jeannelafortune/research. Mimeo, University of Maryland.

Manski, Charles F. 1993. “Identification of Endogenous Social Effects: The Reflection Problem.” *Review of Economic Studies* 60 (3):531–542. URL http://www.jstor.org/stable/2298123.

Minns, Chris. 2000. “Income, Cohort Effects, and Occupational Mobility: A New Look at Immigration to the United States at the Turn of the 20th Century.” *Explorations in Economic History* 37:326–350.

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Immigrants in the US Labor Market." *Quarterly Journal of Economics* 118 (2):549–599.

———. 2009. "Strength in Numbers: Networks as a Solution to Occupational Traps." Mimeo, Brown University.

Munshi, Kaivan and Nicholas Wilson. 2008. "Identity, Parochial Institutions, and Career Decisions: Linking the Past to the Present in the American Midwest." URL http://www.econ.brown.edu/fac/Kaivan_Munshi/midwest7.pdf. Mimeo, Brown University.

O'Rourke, Kevin H. and Jeffrey G. Williamson. 1999. *Globalization and History: The Evolution of a Nineteenth-Century Atlantic Economy*. Cambridge: MA: MIT Press.

Patel, Krishna and Francis Vella. 2007. "Immigrant Networks and Their Implications for Occupational Choice and Wages." Discussion Paper 3217, IZA.

Peri, Giovanni and Chad Sparber. 2009. "Task Specialization, Immigration, and Wages." *American Economic Journal: Applied Economics* 1 (3):135–69. URL http://ideas.repec.org/a/aea/aejapp/v1y2009i3p135-69.html.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Matthew B. Schroeder, and Matthew Sobek. 2010. *Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]*. Minneapolis, MN: Minnesota Population Center.

Wegge, Simone A. 1998. "Chain Migration and Information Networks: Evidence From Nineteenth-Century Hesse-Cassel." *The Journal of Economic History* 58 (04):957–986.

Figure 1. An example: Flows ex-ante and ex-post for Bohemian immigrants in Ohio

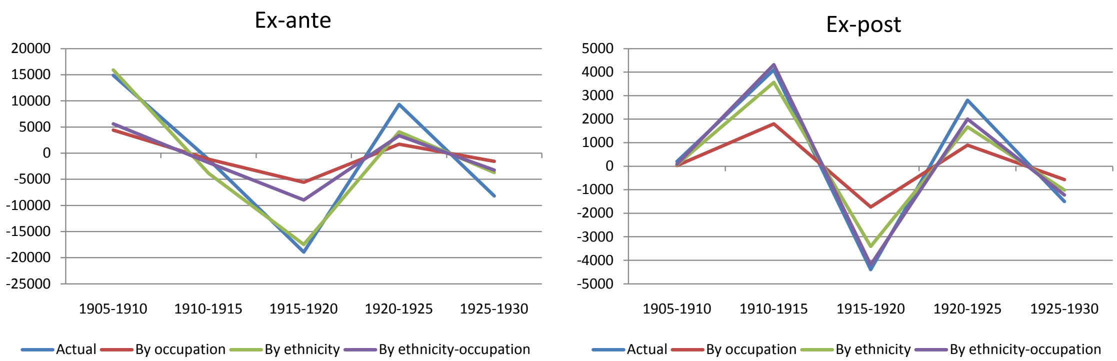

Table 1. An example: Changes from 1910 to 1915 for a given state

	Italian bakers		Italian carpenters	
	Ex-ante	Ex-post	Ex-ante	Ex-post
Change in national totals	100	150	100	50
Predicted state change according to...				
Ethnicity	10	15	10	5
Occupation	20	30	5	2.5
Ethnicity-occupation	30	45	2	1

Table 2. Spatial distribution of immigrants by ethnic group, over time, Census and administrative data contrasted

	Stocks 1900				Flows 1901-1910				Flows 1911-1920				Flows 1921-1930			
	1st	2nd	3rd	TOP 10	1st	2nd	3rd	TOP 10	1st	2nd	3rd	TOP 10	1st	2nd	3rd	TOP 10
Panel A: From Census data																
British ancestry	NY	MA	PA	75.4	PA	NY	MA	78.1	MI	MA	NY	78.4	MA	MI	NY	90.6
French	MA	MI	NY	75.6	CA	MI	NY	76.5	CA	MI	NY	83.9	MI	MA	NY	86.2
South Europeans	NY	PA	MA	86.3	MA	PA	NY	84.5	MA	PA	NY	87.5	PA	NJ	NY	94.3
Hispanics	TX	AZ	FL	93.9	AZ	CA	TX	89.7	NY	CA	TX	89.8	NY	TX	CA	88.0
Germanic	NY	IL	PA	77.1	IL	NY	PA	80.0	OH	PA	NY	82.6	NJ	IL	NY	89.7
Scandinavians	MN	IL	WI	76.3	WI	IL	MN	74.2	IL	MN	NY	77.9	MN	IL	NY	83.9
Russians and others	NY	PA	IL	83.6	IL	PA	NY	86.8	IL	PA	NY	87.9	PA	IL	NY	92.1
Other Europeans	NY	PA	IL	85.9	NY	OH	PA	77.5	NY	OH	PA	78.9	IL	PA	NY	91.7
Other countries	HI	CA	MA	86.4	NY	WA	CA	72.2	NY	HI	CA	85.5	WA	NY	CA	90.3
Panel B: From administrative data																
British ancestry					NY	MA	PA	84.2	NY	MA	MI	80.3	NY	MI	MA	87.3
French					NY	PA	MA	89.9	NY	PA	MA	88.7	NY	PA	MA	91.8
South Europeans					NY	IL	MI	77.0	NY	MA	MI	78.0	MA	NY	MI	84.9
Hispanics					FL	NY	TX	96.1	TX	NY	AZ	95.1	TX	CA	NY	94.8
Germanic					NY	PA	IL	81.3	NY	IL	PA	77.7	NY	IL	NJ	84.9
Scandinavians					NY	MN	IL	80.4	NY	MN	IL	77.9	NY	IL	MN	80.1
Russians and others					NY	PA	IL	93.1	NY	PA	IL	90.6	NY	PA	IL	90.9
Other Europeans					PA	NY	IL	88.1	NY	PA	OH	83.6	NY	PA	OH	87.5
Other countries					HI	CA	NY	88.9	CA	HI	NY	85.8	CA	NY	HI	83.9

Table 3. Occupational distribution of immigrants by ethnic group, over time, Census and administrative data contrasted

	Stocks 1900				Flows 1901-1910				Flows 1921-1930			
	1st	2nd	3rd	TOP 10	1st	2nd	3rd	TOP 10	1st	2nd	3rd	TOP 10
Panel A: From Census data												
British ancestry	Laborers	Farmers	Servants	68.1	Laborers	Servants	Clerks	69.4	Laborers	Servants	Clerks	72.3
French	Laborers	Textile wrkers	Farmers	75.9	Laborers	Farm laborers	Servants	73.0	Laborers	Servants	Farm laborers	68.9
South Europeans	Laborers	Miners	Manufacturers	82.5	Laborers	Miners	Farm laborers	84.8	Laborers	Manufacturers	Miners	72.7
Hispanics	Laborers	Farm laborers	Servants	84.0	Laborers	Farm laborers	Miners	90.6	Laborers	Farm laborers	Servants	87.2
Germanic	Laborers	Farmers	Manufacturers	71.0	Laborers	Miners	Servants	79.7	Laborers	Servants	Machinists	64.1
Scandinavians	Farmers	Laborers	Servants	77.4	Laborers	Servants	Farm laborers	79.6	Laborers	Servants	Clerks	70.7
Russians and others	Laborers	Tailors	Manufacturers	82.8	Laborers	Miners	Tailors	81.2	Laborers	Merchants	Manufacturers	72.3
Other Europeans	Laborers	Farmers	Miners	83.0	Laborers	Miners	Servants	85.5	Laborers	Servants	Hotel keepers	73.0
Other countries	Laborers	Farm laborers	Other unskilled	94.3	Laborers	Farm laborers	Servants	89.3	Farm laborers	Hotel keepers	Other unskilled	89.1
Panel B: From administrative data												
British ancestry					Servants	Laborers	Clerks	75.3	Servants	Laborers	Clerks	72.3
French					Laborers	Servants	Farm laborers	71.9	Laborers	Servants	Farmers	75.6
South Europeans					Laborers	Farm laborers	Servants	91.7	Laborers	Servants	Farm laborers	84.8
Hispanics					Laborers	Tobbaco wkers	Servants	82.6	Laborers	Servants	Clerks	89.1
Germanic					Servants	Laborers	Farm laborers	75.3	Servants	Farmers	Clerks	72.8
Scandinavians					Laborers	Servants	Farm laborers	91.0	Servants	Farmers	Farm laborers	80.8
Russians and others					Laborers	Farm laborers	Servants	90.3	Servants	Laborers	Tailors	78.1
Other Europeans					Laborers	Farm laborers	Servants	95.7	Laborers	Servants	Farm laborers	87.5
Other countries					Farm laborers	Laborers	Merchants	88.2	Merchants	Laborers	Servants	77.1

Table 4. Explaining changes in location choices of immigrants (excluding individuals with no occupation)

	From administrative data			From IPUMS data		
	Admin	Census	IPUMS	Admin	Census	IPUMS
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: By ethnicity (N=7140)			
Pred. flow by ethnicity	0.623***	0.916***	0.838***	0.273*	0.590***	0.539***
	(0.085)	(0.048)	(0.089)	(0.108)	(0.076)	(0.098)
Pred. flow by occupation	0.390**	0.641***	0.537***	0.540***	0.794***	0.626**
	(0.130)	(0.132)	(0.127)	(0.146)	(0.229)	(0.192)
R-square	0.963	0.947	0.936	0.907	0.929	0.930
			Panel B: By occupation (N=15810)			
Pred. flow by ethnicity	0.549***	1.124***	1.163***	0.121	0.437***	0.102
	(0.150)	(0.170)	(0.238)	(0.108)	(0.092)	(0.104)
Pred. flow by occupation	0.344*	0.109	-0.166	0.564**	0.835***	0.918***
	(0.150)	(0.141)	(0.226)	(0.185)	(0.174)	(0.141)
R-square	0.969	0.937	0.930	0.931	0.944	0.960

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnicity/occupation electing a particular state in a particular period of migration and the right-hand side variables are predicted flows as detailed in the left-hand column. Actual and predicted flows are measured from the administrative data (ex-ante) in the first three columns but from the IPUMS (ex-post) in the last three. In columns (1) and (4), these predicted flows are built using location shares from the administrative data, in columns (2) and (5), from the 1900 Census tables and in columns (3) and (6), from the 1900 IPUMS. All regressions include fixed effects for the double interactions of state, ethnicity/occupation and period of immigration.

Standard errors are clustered at the group-state level.

*: 5% significance, **: 1% significance, ***: 0.1% significance

Table 5. Robustness checks: using close ethnic groups as an alternative

	From administrative data			From IPUMS data		
	Admin	Census	IPUMS	Admin	Census	IPUMS
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: By ethnicity (N=6885)					
Pred. flow by ethnicity	0.625***	0.898***	0.839***	0.274*	0.577***	0.536***
	(0.086)	(0.057)	(0.088)	(0.110)	(0.075)	(0.097)
Pred. flow by occupation	0.378**	0.604***	0.506***	0.522**	0.775**	0.620**
	(0.133)	(0.131)	(0.123)	(0.166)	(0.253)	(0.215)
Pred. flow by close eth. grp	0.015	0.078	0.025	0.035	0.059	0.018
	(0.008)	(0.050)	(0.052)	(0.051)	(0.040)	(0.052)
R-square	0.964	0.949	0.937	0.909	0.930	0.930
	Panel B: By occupation (N=15810)					
Pred. flow by ethnicity	0.541***	1.063***	1.309***	0.110	0.370***	0.120
	(0.150)	(0.163)	(0.234)	(0.100)	(0.086)	(0.092)
Pred. flow by occupation	0.324*	0.116	-0.124	0.501**	0.813***	0.983***
	(0.139)	(0.139)	(0.227)	(0.184)	(0.180)	(0.119)
Pred. flow by close eth. grp	0.034	0.083	-0.223*	0.122*	0.159	-0.136
	(0.029)	(0.090)	(0.099)	(0.051)	(0.111)	(0.095)
R-square	0.969	0.937	0.933	0.933	0.946	0.962

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnicity (Panels A) or a particular occupation (Panels C) electing a particular state in a particular period of migration and the right-hand side variables are predicted flows as detailed by the list. Actual and predicted flows are measured from the administrative data (ex-ante) in the first three columns but from the IPUMS (ex-post) in the last three. In columns (1) and (4), predicted flows are built using location shares from the administrative data, in columns (2) and (5), from the 1900 Census tables and in columns (3) and (6), from the 1900 IPUMS. All regressions include fixed effects for the double interactions of state, ethnicity/occupation and period of immigration.

Standard errors are clustered at the group-state level.

*: 5% significance, **: 1% significance, ***: 0.1% significance

Table 6. Explaining changes in location choices of immigrants, networks more finely defined

	From administrative data		From IPUMS data	
	(1)	(2)	(3)	(4)
	Panel A: By ethnicity (N=7140)			
Pred. flow by ethnicity	1.226***	2.541***	-0.609**	-0.782**
	(0.174)	(0.381)	(0.231)	(0.275)
Pred. flow by ethnicity*occ. score diff.		-0.097***		0.023
		(0.028)		(0.014)
Pred. flow by occupation	0.613***	0.615***	0.275***	0.278***
	(0.165)	(0.152)	(0.072)	(0.072)
Pred. flow by ethn.-occupation	-0.225	-0.264	1.346***	1.338***
	(0.148)	(0.139)	(0.240)	(0.239)
R-square	0.929	0.938	0.953	0.953
	Panel B: By occupation (N=15810)			
Pred. flow by ethnicity	1.193***	1.409***	-0.182	-0.196
	(0.245)	(0.217)	(0.103)	(0.125)
Pred. flow by ethnicity*occ. score diff.		-0.012		0.002
		(0.014)		(0.006)
Pred. flow by occupation	-0.072	-0.076	0.561***	0.561***
	(0.155)	(0.155)	(0.099)	(0.100)
Pred. flow by ethn.-occupation	-0.104	-0.143	0.536***	0.636***
	(0.120)	(0.103)	(0.115)	(0.116)
R-square	0.937	0.939	0.966	0.966
	Panel C: Joint estimation (N=22950)			
Pred. flow by ethnicity	1.249***	1.682***	-0.236	-0.315
	(0.210)	(0.270)	(0.145)	(0.193)
Pred. flow by ethnicity*occ. score diff.		-0.026		0.011
		(0.018)		(0.008)
Pred. flow by occupation	0.136	0.129	0.323***	0.323***
	(0.118)	(0.116)	(0.061)	(0.062)
Pred. flow by ethn.-occupation	-0.179	-0.249*	0.839***	0.836***
	(0.139)	(0.103)	(0.126)	(0.127)
R-square	0.924	0.927	0.954	0.954
	Panel D: By ethnicity and occupation (N=442680)			
Pred. flow by ethnicity			-0.212	-0.200
			(0.116)	(0.125)
Pred. flow by ethnicity*occ. score diff.				-0.002
				(0.007)
Pred. flow by occupation			0.159*	0.159*
			(0.065)	(0.065)
Pred. flow by ethn.-occupation			0.891***	0.891***
			(0.101)	(0.101)
R-square			0.747	0.745

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnicity/occupation electing a particular state in a particular period of migration and the right hand side variables are predicted flows constructed as detailed by the left-hand column using shares from 1900 IPUMS. Actual and predicted flows are measured from the administrative data (first two columns) and from IPUMS (last two). All regressions include fixed effects for the double interactions of state, group and period of immigration for the first three panels and for all triple interactions of state, ethnicity, occupation and period of immigration for panel D.

Standard errors are clustered at the group-state level.

*: 5% significance, **: 1% significance, ***: 0.1% significance

Table 7. Robustness checks: networks more finely defined

	From administrative data			From IPUMS data		
	Large	Bef. 1920	Excl. NY	Large	Bef. 1920	Excl. NY
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: By ethnicity			
Pred. flow by ethnicity	1.195***	1.188***	0.948***	-0.645**	-0.484*	-0.529*
	(0.178)	(0.206)	(0.220)	(0.228)	(0.250)	(0.246)
Pred. flow by occupation	0.651***	0.740***	0.461***	0.254**	0.295***	0.290***
	(0.173)	(0.221)	(0.139)	(0.077)	(0.079)	(0.074)
Pred. flow by ethn.-occupation	-0.231	-0.236	-0.237	1.409***	1.218***	1.255***
	(0.138)	(0.181)	(0.128)	(0.242)	(0.250)	(0.255)
R-square	0.938	0.939	0.846	0.958	0.965	0.927
N	3570	4284	7000	3570	4284	7000
			Panel B: By occupation			
Pred. flow by ethnicity	1.185***	1.200***	1.525***	-0.178	-0.181	-0.032
	(0.244)	(0.279)	(0.252)	(0.102)	(0.110)	(0.081)
Pred. flow by occupation	-0.070	-0.053	-0.225	0.557***	0.451***	0.407***
	(0.154)	(0.195)	(0.138)	(0.099)	(0.123)	(0.093)
Pred. flow by ethn.-occupation	-0.103	-0.093	-0.209	0.535***	0.664***	0.627***
	(0.117)	(0.140)	(0.112)	(0.115)	(0.139)	(0.102)
R-square	0.938	0.951	0.920	0.967	0.974	0.954
N	7905	9486	15500	7905	9486	15500
			Panel C: By ethnicity and occupation			
Pred. flow by ethnicity				-0.215	-0.278*	-0.265*
				(0.118)	(0.121)	(0.131)
Pred. flow by occupation				0.157*	0.110	0.214***
				(0.065)	(0.067)	(0.070)
Pred. flow by ethn.-occupation				0.894***	0.971***	0.902***
				(0.101)	(0.104)	(0.123)
R-square				0.746	0.744	0.658
N				244290	265608	434000

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnicity/occupation electing a particular state in a particular period of migration and the right hand side variables are predicted flows constructed as detailed by the left-hand column using shares from 1900 IPUMS. Actual and predicted flows are measured from the administrative data (first three columns) and from IPUMS (last three). All regressions include fixed effects for the double interactions of state, group and period of immigration for the first two panels and for all triple interactions of state, ethnicity, occupation and period of immigration for panel C.

Standard errors are clustered at the group-state level.

*: 5% significance, **: 1% significance, ***: 0.1% significance

Table 8. Exploring the differences between ex-ante and ex-post regressions, by occupation.

	High occ. score	Low occ. score	Traded	Non-traded	Within 5 years	Between 5-10 years
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Flows from administrative data					
Pred. flow by ethnicity	1.066**	1.186***	1.205***	0.428	1.239***	1.132***
	(0.342)	(0.243)	(0.242)	(0.397)	(0.303)	(0.242)
Pred. flow by occupation	0.130	-0.070	-0.135	0.224	-0.097	-0.102
	(0.194)	(0.154)	(0.154)	(0.179)	(0.206)	(0.147)
Pred. flow by ethn.-occupation	0.119	-0.102	-0.101	0.405*	-0.103	-0.108
	(0.200)	(0.118)	(0.110)	(0.205)	(0.154)	(0.101)
R-square	0.922	0.938	0.933	0.983	0.905	0.954
N	7905	7905	8925	6885	9486	6324
	Panel B: Flows from IPUMS data					
Pred. flow by ethnicity	0.165	-0.182	-0.186	0.072	-0.175	-0.186
	(0.146)	(0.102)	(0.103)	(0.153)	(0.121)	(0.147)
Pred. flow by occupation	0.764***	0.558***	0.553***	0.764***	0.516***	1.007**
	(0.213)	(0.099)	(0.099)	(0.112)	(0.122)	(0.348)
Pred. flow by ethn.-occupation	0.215	0.536***	0.542***	0.250*	0.573***	0.156
	(0.158)	(0.116)	(0.117)	(0.126)	(0.154)	(0.252)
R-square	0.885	0.967	0.966	0.962	0.966	0.967
N	7905	7905	8925	6885	9486	6324

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular occupation electing a particular state in a particular period of migration and the right hand side variables are predicted flows constructed as detailed by the left-hand column using shares from 1900 IPUMS. Actual and predicted flows are measured from the administrative data (Panel A) and from IPUMS (Panel B). The first 2 columns divide the sample by the occupational score of each occupation. Columns (3) and (4) compare occupations classified as traded and non-traded as detailed in Table A-1. Column (5) restricts the sample to immigrants observed in the Census within 5 years of their arrival to the United States while the last includes individuals observed in the Census 5-10 years after their arrival. All regressions include fixed effects for the double interactions of state, occupation and period of immigration.

Standard errors are clustered at the occupation-state level.

*: 5% significance, **: 1% significance, ***: 0.1% significance

Table 9. Exploring the differences between ex-ante and ex-post regressions, by ethnicity.

	English	Non English	Low return	High return	Within 5 years	Between 5-10 years
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Flows from administrative data					
Pred. flow by ethnicity	-0.424	1.222***	1.264***	0.943***	1.243***	1.203***
	(1.156)	(0.177)	(0.209)	(0.182)	(0.224)	(0.194)
Pred. flow by occupation	0.844	0.618***	0.710**	0.154	0.686**	0.404*
	(0.509)	(0.170)	(0.228)	(0.130)	(0.213)	(0.187)
Pred. flow by ethn.-occupation	1.263	-0.223	-0.311	0.216	-0.240	-0.214
	(1.152)	(0.148)	(0.172)	(0.175)	(0.211)	(0.125)
R-square	0.964	0.928	0.935	0.925	0.886	0.949
N	1020	6120	3570	3570	4284	2856
	Panel B: Flows from IPUMS					
Pred. flow by ethnicity	-0.724*	-0.593*	-0.688**	-0.230	-0.683**	-0.460
	(0.362)	(0.244)	(0.235)	(0.397)	(0.256)	(0.436)
Pred. flow by occupation	-0.317**	0.291***	0.256***	0.234	0.165*	0.634***
	(0.096)	(0.073)	(0.070)	(0.176)	(0.069)	(0.142)
Pred. flow by ethn.-occupation	2.043***	1.315***	1.445***	0.967*	1.503***	0.903*
	(0.358)	(0.253)	(0.246)	(0.425)	(0.264)	(0.450)
R-square	0.949	0.956	0.970	0.893	0.954	0.949
N	1020	6120	3570	3570	4284	2856

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnicity electing a particular state in a particular period of migration and the right hand side variables are predicted flows constructed as detailed by the left-hand column using shares from 1900 IPUMS. Actual and predicted flows are measured from the administrative data (Panel A) and from IPUMS (Panel B). The first 2 columns divide the sample by whether the ethnicity is English speaking or not. Columns (3) and (4) compare ethnicities by the share of return migration observed over this entire period. Column (5) restricts the sample to immigrants observed in the Census within 5 years of their arrival to the United States while the last includes individuals observed in the Census 5-10 years after their arrival. All regressions include fixed effects for the double interactions of state, ethnicity and period of immigration.

Standard errors are clustered at the ethnic-state level.

*: 5% significance, **: 1% significance, ***: 0.1% significance

Table 10. Explaining location choice using the "wrong" predicted flows

	Flows from administrative data	Flows from IPUMS data
	(1)	(2)
	Panel A: By ethnicity (N=7140)	
Pred. flow by ethnicity	1.786	0.205***
	(1.380)	(0.052)
Pred. flow by occupation	-0.015	0.255**
	(0.538)	(0.083)
Pred. flow by ethn.-occupation	0.435	-0.004
	(1.074)	(0.049)
R-square	0.830	0.758
	Panel B: By occupation (N=15810)	
Pred. flow by ethnicity	1.579*	0.041
	(0.722)	(0.096)
Pred. flow by occupation	0.780**	0.219
	(0.266)	(0.115)
Pred. flow by ethn.-occupation	-0.850	0.113
	(0.778)	(0.132)
R-square	0.681	0.772
	Panel C: Joint estimation (N=22950)	
Pred. flow by ethnicity	1.885*	0.090
	(0.770)	(0.088)
Pred. flow by occupation	-0.024	0.232**
	(0.525)	(0.078)
Pred. flow by ethn.-occupation	-0.154	0.086
	(0.656)	(0.083)
R-square	0.756	0.763
	Panel D: By ethnicity and occupation (N=442680)	
Pred. flow by ethnicity		0.050
		(0.046)
Pred. flow by occupation		0.142***
		(0.038)
Pred. flow by ethn.-occupation		0.049
		(0.029)
R-square		0.111

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular occupation electing a particular state in a particular period of migration and the right hand side variables are predicted flows constructed as detailed by the left-hand column using shares from 1900 IPUMS. Actual flows are from the administrative data and the predicted flows from IPUMS in Panel A and vice versa in Panel B. Column (1) represents the regression computed on the sample by ethnicity, column (2) on the one by occupation. Column (3) corresponds to the regressions on both ethnicity and occupation cells are estimated simultaneously. The final column includes all ethnicity-occupation combinations as observations. All regressions include fixed effects for the double interactions of state, group and period of immigration, Column (4) also include all triple interactions between state, ethnicity, occupation and period of immigration.

Standard errors are clustered at the group-state level.

*: 5% significance, **: 1% significance, ***: 0.1% significance

A Matching both Data Sources

The Census data in IPUMS provides two different occupations codes, the 1900 (for 1900 and 1910) and the 1950 Census occupation codes (for all years). Administrative data comes in a different classification; with four main groups –professional, skilled, unskilled, and no-occupation, being divided into very detailed lists of subgroups (for each of the first three groups). We match groups of administrative data occupations to groups from the Census classifications. We match the administrative data to the IPUMS samples using the 1950 Census occupation codes to preserve comparability across years. When matching to the 1900 Census published tables, however, occupations were paired using the 1900 Census codes because this was the only classification available. The exact groups are in Table A-1 together with the corresponding Census codes. The appendix table also shows our classification of occupations as large or small, defined as being above or below the median size for all immigrant flow over the period.

The Commissioner of Immigration classified immigrants by their "ethnicity" rather than their country of origin.[36] In order to reach a matching set of ethnicities, we grouped both countries of birth and ethnicities from the administrative data. Our final classification includes 28 "ethnicities" used in the regression.[37] Most of the pairings are fairly intuitive but we needed to make a few adjustments in order to represent the definitions used by both data sets. For example, the RCI classified all Blacks as Africans. However, over the period studied, most Black immigrants are from the Caribbean and not from Africa. This explains why they are paired with West Indians rather than with Africans. Similarly, Jewish immigrants were classified as "Hebrews". We allocate Jewish immigrants by their country of birth using the available information from the RCI, which presents a table with the distribution of individuals by their ethnicity and the country of last residence, although this pairing is definitely a gross approximation. We also collated our various "ethnicities" by groups to present more easily understandable summary statistics and to run some of the regressions separately for different ethnic groups.[38]

[36]For some years, a table highlights the distribution of individuals by their ethnicity and the country of last residence but does not indicate the intended state of residence nor the occupations by countries. We use the information in these tables to allocate certain ethnic groups to different countries as explained later in this section.

[37]Those groups are described in Appendix Table A-2.

[38]Lafortune (2010) presents a similar classification of countries into groups using Census data from the same period.

Finally, we also classify each ethnic group by whether its flow was above or below the median over this period and classified as “large” those that were above the median. Similarly, we computed the difference between the national flow as measured in the administrative data and in the Census data and classify as those that have a difference in percentage term larger than the median as high return migration ethnicities. Most of those correspond closely to the ones identified in the existing literature as having large circular migration flows (Hatton and Williamson 2005).

NOT FOR PUBLICATION

Table A-1. Matching of occupations between administrative and Census data

Administrative Data Categories	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?	Traded?
			TOTAL PROFESSIONAL		
Electricians	515	Electricians	Electricians	No	No
	603	Apprentice electricians			
Engineers (professional)	41	Engineers, aeronautical	Engineers (civil, etc) and surveyors	Yes	No
	42	Engineers, chemical	Designers, draughtsmen and inventors		
	43	Engineers, civil			
	44	Engineers, electrical			
	45	Engineers, industrial			
	46	Engineers, mechanical			
	47	Engineers, metallurgical, metallurgists			
	48	Engineers, mining			
	49	Engineers (n.e.c.)			
	35	Draftsmen			
	92	Surveyors			
Sculptors and artists	4	Artists and art teachers	Artists and teachers of art	No	No
	31	Dancers and dancing teachers			
Literary and scientific persons	6	Authors	Literary and scientific persons	No	No
	56	Librarians			
	301	Attendants and assistants, library			
	7	Chemists			
	61	Agricultural scientists			
	62	Biological scientists			
	63	Geologists and geophysicists			
	67	Mathematicians			
	68	Physicists			
	69	Miscellaneous natural scientists			
	83	Statisticians and actuaries			
Actors	1	Actors and actresses	Actors	No	No
			Theatretical managers, etc		
Musicians	51	Entertainers (n.e.c.)	Musicians and teachers of music	Yes	No
	57	Musicians and music teachers	Professional showmen		
Teachers	12	Agricultural sciences	Teachers and professors in colleges, etc	Yes	No
	13	Biological sciences			
	14	Chemistry			
	15	Economics			
	16	Engineering			
	17	Geology and geophysics			
	18	Mathematics			
	19	Medical sciences			
	23	Physics			
	24	Psychology			
	25	Statistics			
	26	Natural science (n.e.c.)			
	27	Social sciences (n.e.c.)			
	28	Nonscientific subjects			
	29	Subject not specified			
	93	Teachers (n.e.c.)			
Clergy	9	Clergymen	Clergymen	No	No
Officials (Government)	210	Inspectors, public administration	Officials (government)	No	No
	250	Officials and administrators (n.e.c.), public administration	Soldiers, sailors and marines (U.S.)		
	270	Postmasters			

Administrative Data Categories	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?	Traded?
	595	Members of the armed services			
	771	Marshals and constables			
	782	Sheriffs and bailiffs			
Physicians	75	Physicians and surgeons	Physicians and surgeons	No	No
Architects	3	Architects	Architects	No	No
Editors	36	Editors and reporters	Journalists	No	No
Lawyers	55	Lawyers and judges	Lawyers	No	No
Other professionals	2	Airplane pilots and navigators	Dentists	Yes	No
	5	Athletes	Other professional service		
	8	Chiropractors	Nuurses (trained)		
	32	Dentists	Nurses (not specified)		
	70	Optometrists	Undertakers		
	71	Osteopaths			
	10	College presidents and deans			
	33	Designers			
	34	Dieticians and nutritionists			
	52	Farm and home management advisors			
	53	Foresters and conservationists			
	54	Funeral directors and embalmers			
	58	Nurses, professional			
	59	Nurses, student professional			
	72	Personnel and labor relations workers			
	73	Pharmacists			
	76	Radio operators			
	77	Recreation and group workers			
	78	Religious workers			
	79	Social and welfare workers, except group			
	81	Economists			
	82	Psychologists			
	84	Miscellaneous social scientists			
	91	Sports instructors and officials			
	94	Technicians, medical and dental			
	95	Technicians, testing			
	96	Technicians (n.e.c.)			
	97	Therapists and healers (n.e.c.)			
	98	Veterinarians			
	99	Professional, technical and kindred workers (n.e.c.)			
	260	Officials, lodge, society, union, etc.			
	532	Inspectors, scalers, and graders, log and lumber			
	533	Inspectors (n.e.c.)			
	781	Practical nurses			
TOTAL SKILLED					
Barbers and hairdressers	740	Barbers, beauticians, and manicurists	Barbers and hairdressers	Yes	No
Clerks and accountants	0	Accountants and auditors	Bookeepers and accountants	Yes	No
	302	Attendants, physician's and dentist's office	Clerks and coypists		
	310	Bookkeepers	Messengers and errand and office boys		
	320	Cashiers	Stenographers and typewriters		
	321	Collectors, bill and account			
	325	Express messengers and railway mail clerks			
	335	Mail carriers			
	340	Messengers and office boys			
	341	Office machine operators			
	342	Shipping and receiving clerks			
	350	Stenographers, typists, and secretaries			
	360	Telegraph messengers			
	365	Telegraph operators			
	370	Telephone operators			

Administrative Data Categories	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?	Traded?
	390 450 520	Clerical and kindred workers (n.e.c.) Insurance agents and brokers Electrotypers and stereotypers			
Bankers	204 305 480	Credit men Bank tellers Stock and bond salesmen	Bankers and brokers Officials of banks and companies	No	No
Gardeners	930	Gardeners, except farm, and groundskeepers	Gardeners, florists, nurserymen, etc Garden and nursery laborers	Yes	No
Watch and clock makers Jewelers	534	Jewelers, watchmakers, goldsmiths, and silversmiths	Clock and watchmakers and repairers Gold and silver workers	Yes	Yes
Carpenters and joiners Shipwrights	510 602	Carpenters Apprentice carpenters	Carpenters and joiners	Yes	Yes
Cabinetmakers	505	Cabinetmakers	Cabinet makers	No	Yes
Woodworkers (not specified)	674	Sawyers	Coopers Saw and planing mill employees Other wood workers	No	Yes
Plumbers	574 610	Plumbers and pipe fitters Apprentice plumbers and pipe fitters	Plumbers and gas and steam fitters	No	No
Painters and glaziers	530 564 670	Glaziers Painters, construction and maintenance Painters, except construction or maintenance	Painters, glaziers and varnishers	Yes	Yes
Masons	504 601	Brickmasons, stonemasons, and tile setters Apprentice bricklayers and masons	Masons	Yes	Yes
Stonecutters	584 635	Stone cutters and stone carvers Filers, grinders, and polishers, metal	Marble and stone cutters	Yes	Yes
Blacksmiths	501 524	Blacksmiths Forgemen and hammermen	Blacksmiths	Yes	Yes
Engineers	203 541 583	Conductors, railroad Locomotive engineers Stationary engineers	Conductors (steam railroad) Engineers and firemen Engineers and firemen (not railroad)	Yes	No
Iron and steel workers Metal workers	503 531 535 561 580 585 612 642 685	Boilermakers Heat treaters, annealers, temperers Job setters, metal Molders, metal Rollers and roll hands, metal Structural metal workers Apprentices, metalworking trades (n.e.c.) Heaters, metal Welders and flame cutters	Iron and steel workers Steam boiler makers Stove, furnace and grate makers Wire workers Brass workers Other metal workers	Yes	Yes
Stokers	641	Furnacemen, smeltermen and pourers	Charcoal, coke and lime burners	No	Yes
Machinists	544 604	Machinists Apprentice machinists and toolmakers	Machinists	Yes	Yes
Wheelwrights Mechanics Locksmiths	545 550 551 552 553 554 600 605	Mechanics and repairmen, airplane Mechanics and repairmen, automobile Mechanics and repairmen, office machine Mechanics and repairmen, radio and television Mechanics and repairmen, railroad and car shop Mechanics and repairmen (n.e.c.) Apprentice auto mechanics Apprentice mechanics, except auto	Wheelwrights Mechanics (nec)	Yes	Yes
Printers	512	Compositors and typesetters	Printers, lithographers and pressmen	No	Yes

Administrative Data Categories	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?	Traded?
Tinners	575 613 591	Pressmen and plate printers, printing Apprentices, printing trades Tinsmiths, coppersmiths, and sheet metal workers	Tinplate and tinware makers		
Hat and Cap makers Milliners	645	Milliners	Hat and cap makers Milliners	No	Yes
Seamstresses Dressmakers	633	Dressmakers and seamstresses, except factory	Seamstresses Dressmakers	Yes	Yes
Shoemakers	582	Shoemakers and repairers, except factory	Boot and shoe makers and repairers	Yes	Yes
Tailors	590	Tailors and tailoresses	Tailors and tailoresses	Yes	Yes
Textile workers (not specified)	543 634	Loom fixers Dyers	Bleachery and dye works operatives Carpet factory operatives Cotton mill operatives Hosiery and knittign mill operatives Woolen mill operatives Shirt, collar, and cuff makers	Yes	Yes
Weavers and spinners Furriers and fur workers	675 684 525	Spinners, textile Weavers, textile Furriers	Other textile mill operatives Other textile workers	Yes	Yes
Upholsterers	593	Upholsterers	Upholsterers	No	Yes
Mariners	240 623 673	Officers, pilots, pursers and engineers, ship Boatmen, canalmen, and lock keepers Sailors and deck hands	Boatmen and sailors	Yes	No
Bakers	500	Bakers	Bakers	Yes	Yes
Butchers	644	Meat cutters, except slaughter and packing house	Butchers	Yes	Yes
Millers	555	Millers, grain, flour, feed, etc.	Millers	Yes	Yes
Miners	650	Mine operatives and laborers	Miners and quarrymen	Yes	Yes
Photographers	74 671	Photographers Photographic process workers	Photographers	No	No
Plasterers	573	Plasterers	Plasterers	No	Yes
Bookbinders	502	Bookbinders	Bookbinders	No	Yes
Engravers	521 571	Engravers, except photoengravers Photoengravers and lithographers	Engravers	No	Yes
Pattern makers	570	Pattern and model makers, except paper	Model and pattern makers	No	Yes
Saddlers and harness makers Tanners and curriers Brewers Cigar packers Cigarette makers Tobacco workers Cigar makers Other skilled	511 513 514 522 523 540 542 562 563 565 572 581 592 594 611 614 615	Cement and concrete finishers Cranemen, derrickmen, and hoistmen Decorators and window dressers Excavating, grading, and road machinery operators Foremen (n.e.c.) Linemen and servicemen, telegraph, telephone, and power Locomotive firemen Motion picture projectionists Opticians and lens grinders and polishers Paperhangers Piano and organ tuners and repairmen Roofers and slaters Tool makers, and die makers and setters Craftsmen and kindred workers (n.e.c.) Apprentices, building trades (n.e.c.) Apprentices, other specified trades Apprentices, trade not specified	Harness and saddle makers and repairs Leather curriers and tanners Brewers and malsters Distillers and rectifiers Tobacco and cigar factory operatives Switchmen, yardmen and flagmen Telegraph and telephone linemen Decorators and window dressers Weighters, gaugers, and measurers Paper hangers Roofers and slaters Brick and tile makers Glass woekrs Potters Butter and cheese makers Confectionners Trunk and leather-case makers	Yes	Yes

Administrative Data Categories	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?	Traded?
	620	Asbestos and insulation workers	Bottlers and soda makers		
	622	Blasters and powdermen	Box makers		
	624	Brakemen, railroad	Broom and brush makers		
	630	Chainmen, rodmen, and axmen, surveying	Glove makers		
	662	Oilers and greaser, except auto	rubber factory operatives		
	672	Power station operators	Tool and cutlery makers		
	681	Switchmen, railroad			
		TOTAL UNSKILLED			
Farm laborers	640	Fruit, nut, and vegetable graders, and packers	Farm and plantation laborers	Yes	Yes
	810	Farm foremen	Farm laborers (members of family)		
	820	Farm laborers, wage workers	Dairymen and dairywomen		
	830	Farm laborers, unpaid family workers	Stock raisers, herders and drovers		
	840	Farm service laborers, self-employed	Turpentine farmers and laborers		
Farmers	100	Farmers (owners and tenants)	Farmers, planters, and overseers	Yes	Yes
	123	Farm managers			
Fishermen	910	Fishermen and oystermen	Fishermen and oystermen	No	Yes
Laborers	690	Operative and kindred workers (n.e.c.)	Laborers (not specified)	Yes	Yes
	920	Garage laborers and car washers and greasers	Laborers (steam railroad)		
	950	Lumbermen, raftsmen, and woodchoppers	Laborers (street railway)		
	970	Laborers (n.e.c.)	Oil well and oil works employees		
	940	Longshoremen and stevedores	Other chemical workers		
			Other food preparers		
			Other miscellaneous industries		
Servants	700	Housekeepers, private household	Housekeepers and stewards	Yes	No
	710	Laundressses, private household	Servants		
	720	Private household workers (n.e.c.)			
	753	Charwomen and cleaners			
	764	Housekeepers and stewards, except private household			
	780	Porters			
Agents	280	Purchasing agents and buyers (n.e.c.)	Agents	No	No
	300	Agents (n.e.c.)	Station agents and employees (steam railraod)		
	380	Ticket, station, and express agents	Station agents and employees (street railroad)		
	470	Real estate agents and brokers			
Draymen, hackmen and teamsters	304	Baggagemen, transportation	Draymen, hackmen, teamsters	No	No
	322	Dispatchers and starters, vehicle	Baggagemen		
	625	Bus drivers	Brakemen		
	631	Conductors, bus and street railway	Conductors (street railway)		
	632	Deliverymen and routemen	Drivers (street railway)		
	660	Motormen, mine, factory, logging camp, etc.	Motormen		
	661	Motormen, street, subway, and elevated railway			
	682	Taxicab drivers and chauffers			
	683	Truck and tractor drivers			
	960	Teamsters			
Manufacturers	290	Managers, officials, and proprietors (n.e.c.)	Foremen and overseers	No	Yes
	560	Millwrights	Manufacturers and officials		
Merchants and dealers	200	Buyers and department heads, store	Commercial travelers	Yes	Yes
	201	Buyers and shippers, farm products	Hucksters and peddlers		
	205	Floormen and floor managers, store	Merchants and dealers (except wholesale)		
	400	Advertising agents and salesmen	Merchants and dealers (wholesale)		
	410	Auctioneers	Salesmen and saleswomen		
	420	Demonstrators	Auctionneers		
	430	Hucksters and peddlers	Newspaper carriers and newsboys		
	460	Newsboys			
	490	Salesmen and sales clerks (n.e.c.)			
Hotel keepers	621	Attendants, auto service and parking	Bartenders	No	No
	750	Bartenders	Boarding and lodging house keeprs		

Administrative Data Categories	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?	Traded?
	752	Boarding and lodging house keepers	Hotel keepers		
	754	Cooks, except private household	Restaurant keepers		
	760	Counter and fountain workers	Saloon keepers		
	784	Waiters and waitresses	Waiters		
Other miscellaneous	230	Managers and superintendents, building	Lumbermen and raftsmen	Yes	No
	643	Laundry and dry cleaning operatives	Wood choppers		
	680	Stationary firemen	Other agricultural pursuits		
	730	Attendants, hospital and other institution	Janitors and sextons		
	731	Attendants, professional and personal service (n.e.c.)	Launderers and laundresses		
	732	Attendants, recreation and amusement	Midwives		
	751	Bootblacks	Watchmen, policemen, firemen, etc.		
	761	Elevator operators	Other domestic and personal services		
	762	Firemen, fire protection	Hostlers		
	763	Guards, watchmen, and doorkeepers	Livery stable keepers		
	770	Janitors and sextons	Packers and shippers		
	772	Midwives	Porters and helpers (in stores)		
	773	Policemen and detectives	Other persons in trade and transportation (nec)		
	783	Ushers, recreation and amusement			
	785	Watchmen (crossing) and bridge tenders			
	790	Service workers, except private household (n.e.c.)			
		TOTAL NO OCCUPATION			
No occupation	980	Keeps house/housekeeping at home/housewife			
	981	Imputed keeping house (1850-1900)			
	982	Helping at home/helps parents/housework			
	983	At school/student			
	984	Retired			
	985	Unemployed/without occupation			
	986	Invalid/disabled w/ no occupation reported			
	987	Inmate			
	991	Gentleman/lady/at leisure			
	995	Other non-occupational response			

Table A-2. Matching ethnic groups and countries.

Ethnic group	Ethnicities (from administrative data)	Country of birth (from Census)	High return?	Large?
British anscestry	English	England, Canada (English), Australia	No	Yes
	Irish	Ireland	No	Yes
	Scotch	Scotland	No	Yes
	Welsh	Wales	No	No
French	Dutch and Flemmish	Belgium, Netherlands	No	No
	French	France, Canada (French)	No	Yes
South Europeans	Italian (North), Italian (South)	Italy	Yes	Yes
	Portuguese	Portugal	Yes	No
	Spanish	Spain	Yes	No
Hispanics	Mexican	Mexico	No	Yes
	Spanish-American	Central and South America	No	No
	Cuban, West Indian, African (Black)	All Carribeans Islands	Yes	No
Germans	Germans, German Hebrews	Austria, Germany, Switzerland	No	Yes
Scandinavians	Finnish	Finland	No	No
	Scandinavian	Denmark, Norway, Sweden	No	Yes
Russians and others	Lithuanian, Russian, Russian Hebrews	Russia	No	Yes
	Polish	Poland	Yes	Yes
	Romanian, Romanian Hebrews	Romania	Yes	No
Other Europe	Bohemian and Moravian, Ruthenian, Slovak	Bohemia (Czechoslovakia)	Yes	Yes
	Serbian and Montenegrin, Croatian and Slovenian, Bulgarian, Dalmatian, Bosnian, Herzegovinian	Bulgaria, Yugoslavia	Yes	Yes
	Greek	Greece	Yes	Yes
	Magyar	Hungary	No	Yes
Other	Chinese	China	Yes	No
	East Indian	India	Yes	No
	Japanese	Japan	No	No
	Pacific Islander	Pacific Islands	No	No
	Syrian, Turkish, Armenian	Turkey	Yes	No
	Others, other Hebrews, Korean	Korea, other countries	Yes	No