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

Jeanne Lafortune[†]   José Tessada[‡]

January 29, 2010

Abstract

How does the presence of networks influence the location and occupation decisions of immigrants and how does it change as the immigrant spends time in her new country? This paper tries to answer this question by using the location and occupation choices of immigrants implicit in the gross and net immigration flows by ethnicity, occupation and state coming to the United States between 1900 and 1930. We compare the actual distribution of immigrants by intended state of residence at their entry into the United States and the one by actual state of residence later on to counterfactual distributions constructed by allocating new flows according to the distribution of previous migrants, our measure of networks. We find that the distribution of immigrants by intended state of residence, which is measured ex-ante, is most closely approximated by allocating these new migrants to locations where they have ethnic networks and "occupational" networks. However, the actual distribution of immigrants once settled is mostly driven by the presence of ethnic-specific occupational networks. Furthermore, the importance of ethnic-specific networks decreases and that of occupational networks (but not ethnic-specific) increases as an immigrant spends more time in the United States. These results are consistent with migrants selecting their state of residence based on the presence of individuals from their ethnic network and then selecting a new occupation at their arrival based on the occupations of that network. As they spend more time in the US, immigrants seem to need to rely less on their ethnic network, a result that may indicate a slow acquisition of knowledge and information about the local labor markets, thus increasing the relevance of their own skills in their occupational choice. Overall, these results emphasize the role of networks as a transitional mechanism and highlight the potential loss of occupation-specific human capital during this transition period.

It is the most natural thing in the world for an immigrant to want to settle where there are numbers of others of his immediate kind.
–Henry Pratt Fairchild, 1925

1   Introduction

The fact that immigrants of a particular country or ethnic group tend to locate in similar places once in their new adoptive country has been observed for a long period. The anti-immigration discourses of the early 20th century were using this pattern as an argument against immigration over that period. A similar pattern has now been observed in the most recent wave of migration to the United States, mostly from Mexico and Central America. Various studies have used this fact 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). 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 destinations, suggesting that this clustering may be linked to the capacity of immigrants to help each other find work through referrals. Patel and Vella (2007) emphasize that referrals may help explain the occupational concentration of different ethnicities in different cities that are observed in the current data. However, the dynamic process through which networks may be providing labor market assistance is still unexplored. How does the role of networks appear to change as the migrant spend more time in her new country of adoption and learns about labor markets in the United States? What is the role of networks in the dynamic process of adaptation to a new environment? This paper attempts to answer these questions by first setting up a simple theoretical framework and then empirically evaluating it using the location and occupation choices of immigrants implicit in the gross and net immigration flows by ethnicity, occupation and state coming to the United States between 1900 and 1930.

First, this study sets up a simple two-period model where migrants who have just arrived to their new country are unable to find employment unless they are being referred by another individual. Furthermore, the size of their network is limited in such a way that, in many cases, none of their network members can provide them referral for the employment in which their level of skills is best matched. In the second period, that is once they have acquired a better knowledge of their labor markets, they are capable of finding employment in the sector of their choice. Knowing this, immigrants who are deciding in which location to settle should pick a location where they have the combination of two factors: they have a network of individuals who can provide them referral/employment for the first period and the labor market for the set of skills they possess is promising. On the other hand, once we observe them in the first period and look at the occupation in which they are currently employed, we would find it to be

highly correlated with the occupations of their networks. This dependence should decrease in the second period.

The conclusions from this simple framework are then explored using multiple sources of data to study the large wave of migration to the United States between 1900 and 1930. A data source, yet unexplored to the best of our knowledge, first allows us to measure the gross flow of immigrants at the port of entry in the United States classified by the ethnic group one pertains to, the *intended* state of residence and the occupation of the migrant *in his country of origin*. These data were tabulated and published by the Commissioner of Immigration in his annual report from 1899 until 1930, and then again in 1932.[1] We match the cohorts of immigrants from this data to the ones measured in the decennial Census where the *actual* state of residence of newly arrived immigrants, their country of birth, and their *current* occupation is detailed.

To measure the influence of networks, we contrast the predictive power of counterfactual distributions of flows of immigrants over time by state constructed by allocating the national flow using shares based on proxies for different types of networks. All shares are based on the stock of previous immigrants, as is relatively standard in this literature.[2] We first assign the total flow of immigrants to each state according to two possible rules: all migrants follow either the historic location patterns of their ethnicity or that of their occupation. We then expand this strategy further by dividing the "ethnic" and "labor market" effects and creating three measures of networks. One of these proxies includes the share of the stock of immigrants from the same ethnicity but from a different occupation who have located previously in a particular state. The second is computed using the stock of previous migrants sharing one's occupation but not one's ethnic background. The last measures only the stock of individuals sharing both one's occupation *and* one's ethnicity. We then use these counterfactuals to estimate how the observed change in the ex-ante (intended) and ex-post distributions by state of immigrants of different ethnic groups and/or occupations are related to the concentration of individuals from the same ethnic group, or the same occupation or the same ethnic group and occupation in each state.

We find that individuals' choice of intended state of residence, which were reported ex-ante, appear to be highly driven by the existence of a network of individuals from the same ethnic group in their selected state of residence although there appears to be some role, albeit smaller, for decisions to be driven by the existence of a labor market suited for the skills of the new migrant. However, once one attempts to explore the ex-post location decisions of individuals, the relative size of the occupation in the local labor market becomes much more relevant to the decision. Once we attempt to separate more carefully these two effects, one finds that the ex-ante decision is particularly driven by the presence of individuals of a given ethnic group but

---

[1] 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.

[2] We argue that our measure of past stocks represents a measure of a network and try to construct it in a way as to diminish the potential reflection problem as emphasized by Manski (1993)

of a different occupation than the migrant himself. The ex-post decision is, on the other hand, strongly influenced by the presence of individuals of one's own ethnic group who also appear to have the same set of skills than the migrant.

These conclusions are qualitatively similar whether we attempt to predict the changes in the flow of immigrants over time in each state by ethnicity, by occupation or by the combination of both. Results are also robust to the measure of past stocks used to construct network proxies. Furthermore, they do not appear to be driven by a particular ethnic group. In addition, including another proxy using stocks from migrants of a similar ethnic background but a different ethnicity does not alter our conclusions. These results are particularly strong for large ethnic groups and occupations and for immigrants arriving before World War I.

The difference between the factors influencing the distribution of immigrants observed before and after their arrival to the US appears to be consistent with our theoretical framework where individuals rely on their network at their arrival in order to find employment in the occupations available within their networks. Further evidence supporting this includes the fact that the difference between the ex-ante and ex-post distribution is particularly more marked for immigrants recently arrived to the United States than for those who have been residing for a longer period. Moreover, the pattern highlighted above is particularly visible for immigrants whose occupation in their country of origin was not classified as high skill, which is logical since these individuals would have the least to loose by changing their occupation at arrival. Two other alternative hypotheses are also explored. First, the difference between the ex-ante and ex-post distributions could be driven by selective return migration through which migrants without the support of their networks return to their origin country. This appears to be unlikely as the difference between ex-ante and ex-post decisions is similar across ethnic groups with high and low rates of return migration. The pattern we highlight could also be driven by immigrants relocating within the United States at their arrival if their networks cannot provide them with the support needed. The evidence we present, although it does not rule out internal migration playing a role, it does not seem to support it as the main or sole mechanism through which this patterns arises. First, the fact that it is mostly recent immigrants who are displaying a difference in their ex-ante and ex-post distribution could only be explained by internal migration if individuals changed their location temporarily at their arrival to the United States. The low predictive power of our methodology when trying to explain the location of individuals by their actual occupation and state of residence when allocating the flow of immigrants by their occupation in their country of origin is also indicative that individuals are changing occupations at their arrival to the United States, not simply electing a new state of residence. We believe this set of evidence is consistent with ethnic networks playing a role upon arrival and the first settling years, leading migrants to adopt new occupations, but also with migrants learning about the effective labor market opportunities over time, which affects the immigrants' decisions in subsequent years.

3

Related Literature. These conclusions contribute to our understanding of the role networks have in the decisions of migrants and in immigration in general. First, the studies exploiting local variation at the city level to study the impact of immigration (see for example Card (2001); Card and Lewis (2005); Cortés (2008); Cortés and Tessada (2009); Peri and Sparber (2009) among several others) have used past distributions of immigrations to generate variation in the flow of immigrants to different areas; because of the existence of ethnic networks, this variation is argued to be unrelated to contemporaneous local labor market conditions. Our results 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 labor market conditions.

The dynamic effects of networks, in particular in the context of immigration, is still not well understood. Theoretical models of such dynamics have been offered (see for example Calvo-Armengol and Jackson 2004, 2007) but little empirical evidence has been provided. In a recent paper, Beaman (2008) 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 this network: while the size of the recently arrived cohorts of refugees affects negatively the outcomes of the current arrivals, the size of older cohorts (two or more years before the current refugees) improves the outcomes. Our contribution is complimentary to hers as we argue that the role of networks itself changes as a migrant spends more time in the United States.

Finally, our results help understanding the occupational concentration among immigrants. The work of Patel and Vella (2007) builds on the previous evidence on the role of ethnic networks, and networks in general, in the labor market performance of immigrants to study the occupation choice of immigrants from the same country in different cities. They start from the observation that in many cities immigrants from particular countries have developed specific occupational niches.[3] Using data from the 1980, 1990 and 2000 Census in the United States they find that 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.[4] This evidence would be consistent with the interpretation of the results we present here.

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 networks dynamics of immigration.

---

[3]See also Federman et al. (2006).
[4]Munshi and Wilson (2008) study the connection between ethnic networks when the American Midwest was first settled and occupational choice today.

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.

2   Motivating Theory

Assume a migrant from ethnic group j and occupation o must decide which location s to settle in. Once he moves to his new country, he will live for two periods and he discounts the future using a discount factor of β.

The labor market functions as followed. The wage offered to individuals in occupation q and state s with skills of occupation o is denoted by w_{qso}. That wage is larger when $o = q$ as individuals have the skills appropriate for their occupation.

However, there are frictions in the labor market which makes finding employment difficult for new immigrants. Let the probability for a migrant of finding employment in period t given by P_{jqst}.

In the first period, this probability is dependent on his network. He will only find employment in occupation o if his network includes an individual who practices that occupation. Denote the network of a migrant from ethnic group j, occupation o by the set Θ_{jo} and the components of that set by the θ_{kpr} where k denotes the ethnicity of each member, p their occupation and r their state of residence. Then, we can define the employment probabilities in the first period as:

$$P_{jos1} = \begin{cases} 1 & \text{if } \exists \theta_{jos} \in \Theta_{jo} \\ 0 & \text{if } \nexists \theta_{jos} \in \Theta_{jo} \end{cases}$$

As the migrant stays longer in his new environment, his knowledge of the labor market increases and his capacity to find employment given his skills is no longer dependent on his network and thus we assume $P_{jqs2} = 1$ when $q = o$ or when the individual worked in this occupation in the first period.

The migrant's decision is thus given by:

$$\max_{s} \quad \left\{ \max_{q} \quad \left\{ P_{jqs1} * w_{qso} \right\} + \beta \max_{q} \quad \left\{ P_{jqs2} * w_{qso} \right\} \right\}$$

The migrant must thus make three decisions. First, he must select the location where he migrates (we will for the time being assume no internal migration once one elects a location). Second, in both periods, he must also elect which occupation he will work in. The second period

decision simplifies as the individual will simply work in the occupation that is best matched with his skills (since he is sure to find such employment in the second period). In the first period, the migrant will select the highest paid occupation among individuals in his network given his skills o. Denote that occupation as $m(s)$, the problem thus reads as:

$$\max_s \quad w_{m(s)so} + \beta w_{oso}$$

The migrant will thus select the state based on two elements: first, the wage that is paid to individuals of his own skills in that state. In addition, he will also value states where a member of his network can provide him with the best employment in the first period.

A few conclusions can be drawn from this simple model. First, migrants will tend to select states where their skills are valued and thus where we will encounter other individuals using the same skills. Second, they will tend to locate where they have network members to smooth their transition as they migrate. If we assume that networks tend to be formed within ethnic group, this will generate ethnic clustering. Finally, migrants will also select their location anticipating that they will have to rely on their networks in the first period. While they would like to locate in a place where they have a network member who shares their occupation, they may be unable to do so. If that is the case, we would observe the following pattern:

1. Based on their ex-ante skills, we would find that migrants select their location based on the overall attractiveness of this location for their occupation and on the abundance of member from their ethnic network.

2. In the first period, however, we would find that migrants would tend to be concentrated in locations where they have network members who share both their ethnicity and their occupation as they would have taken on the occupation of their network members to smooth their transition.

3. In the second period, however, we would find that most migrants are locating where they have network members and where their occupation is rewarded but we would no longer see occupation clustering within a network.

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" 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. (2009) and from the original published summary tables, which contain data for the first year of each decade.[5]

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.[6] 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 when arriving to the United States.[7] 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 groups. All the information taken from these reports is labeled as "administrative data" in the rest of the paper and was digitalized for the purpose of this paper.

3.2 Census Data

The Decennial Census collects the 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

---

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

[6]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 in a paper format, it would have been prohibitively expensive to enter the information it contains. Hopefully this information will be available in the future.

[7]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.

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.[8] 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 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.[9]

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 (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.[10]

3.3 Matching both Data Sources

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 occupations groups are relatively similar, specific occupations are grouped in different ways and some specific categories are used only by one of the two 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 list of subgroups (for each of the first three groups). We match groups of administrative data occupations to groups from the Census classifications.

---

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

[9]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.

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

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 since this was the only classification available. The exact groups are in Table A-2 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 [11] 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.[12] 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 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.[13]

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).

3.4 Stylized Facts and Summary Statistics

The late 19th and early 20th centuries were periods of massive migration. For example at the turn of the 20th century around 15% of the United States population was foreign born, a number that remains very high until 1930.[14] Annual gross immigration flows of more than a

---

[11]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.

[12]Those groups are described in Appendix Table A-3.

[13]Lafortune (2008) presents a similar classification of countries into groups using Census data from the same period.

[14]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

million people per year are observed between 1900 and 1914, a trend that is 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.[15] 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 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 1 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.[16]

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 2 shows the pattern

---

slightly above 12%.

[15]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.

[16]This change could be related to the modifications to the immigration laws introduced after the end of World War I.

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 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. However, given that the stock and the flow of immigrants at a given period of time could be determined by other factors such as specific demand shocks, we will include the stock of immigrants with given characteristics who are residents in a specific location in a determined period in the past as proxies of the strength of networks.[17] Specifically, we measure our networks mostly in 1900, and include in the regressions only the immigration flows from 1905 to 1930 to avoid

---

[17]We recognize that this could be a biased measure of networks if past immigrants simply responded to long-lasting demand shocks to which current migrants are also responding at their moment of arrival.

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.[18] Finally, our interest will lie more in the comparison of various proxies for networks than in the significance or magnitude of a given coefficient and all the proxies used could suffer from the same bias.

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:

$$n_{jost} = \alpha A_{jos} n_{jot} + \beta B_{jos} n_{jot} + \gamma C_{jos} n_{jot} + \mu_{jos} + \nu_{jot} + \omega_{jst} + \eta_{ost} + \varepsilon_{jost} \tag{1}$$

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

$$A_{jos} \equiv \left(\frac{N_{j\sim os}}{N_{j\sim o}} \right) \tag{2a}$$

$$B_{jos} \equiv \left(\frac{N_{\sim jos}}{N_{\sim jo}} \right) \tag{2b}$$

$$C_{jos} \equiv \left(\frac{N_{jos}}{N_{jo}} \right) \tag{2c}$$

where N represents the stock of immigrants with certain characteristics already in the United States and the sign $\sim$ refers to individuals who do not satisfy one's characteristic. Thus, A_{jos} refers to the share of individuals from one's ethnic group but not from one's occupation who elected to live in state s in the past. The second share, B_{jos} refers to the geographical distribution of immigrants who share one's occupation but not one's ethnicity. 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 a "pure" ethnic effect, to B as a "pure" labor/occupational effect and to C as an ethnic-specific labor market component.

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

---

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

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 (2) 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 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. However, 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 (1) 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, one obtains

$$n_{kst} = \alpha \sum_{k'} A_{jos} n_{jot} + \beta \sum_{k'} B_{jos} n_{jot} + \gamma \sum_{k'} C_{jos} n_{jot} + \mu_{ks} + \nu_{kt} + \omega_{st} + \varepsilon_{kst} \tag{3}$$

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 (3) 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.

One limit to the last regression equation is that the network measures *A*, *B* and *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:

$$n_{kst} = \alpha \sum_{k'} \left(\frac{N_{ks}}{N_k} \right) n_{jot} + \beta \sum_{k'} \left(\frac{N_{k's}}{N_{k'}} \right) n_{jot} + \mu_{ks} + \nu_{kt} + \omega_{st} + \varepsilon_{kst} \tag{4}$$

once more for $k = j, o$. In this regression equation, the networks are much more coarsely defined: the first regressor allocates the national flow of migrants based on the share of past immigrants with characteristic k who chose to settle in state s while the second allocates them based on the share of past immigrants with characteristic k' who settled in state s in the past. 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-state, ethnic-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 state-ethnicity but similar results were obtained when clustering by state only. All regressions are un-weighted.

5 Results

5.1 Ethnic vs. Labor networks

We first begin our empirical exploration by contrasting the role of two general types of networks, ethnic and labor markets, as presented in equation (4). 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.

The results of this strategy are presented in Table 3 and Table 4 when estimating the flow of immigrants selecting a particular state and of a given ethnic group and occupation. 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. 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.

Each of these tables is organized as follows. 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 results in the first panel correspond

14

to specifications of equation 4 where we use as right-hand side variable the predicted flows computed using n_{kt} from the administrative data; in the second panel we use IPUMS data for the variable n_{kt} of equation 4. In panel C we combine both sets of right-hand side variables, i.e. predicted flows using administrative data and predicted flows using Census data. 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 3 indicate that when trying to explain the ex-ante "intended" location decisions of migrants according to their ethnicity, the presence of individuals of their own ethnic group is particularly important. When using flows from the administrative data to construct our predicted location, the past attractiveness of a state for a given occupation also matters, although in a relatively smaller fraction. If we combine both sets of predictors as in Panel C, the shares constructed from administrative data appear to be the ones with the most predictive power.

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 3. In this case, the importance of having individuals of one's occupation who previously located in that state becomes much more important and that of ethnic networks is reduced in all panels. Once more, when we combine the two sets of regressors, the ones that are the most closely related to the left-hand side variable are usually the ones that have the best predictive power. However, whether combined or alone, the conclusions regarding which factor is the most important remain.

Turning to the regressions by occupation, the first three columns of Table 4 (where we attempt to predict the location decisions of migrants according to their reported occupation in the country of departure) highlight the particularly strong role of the presence of individuals from one's ethnicity in the past. This is slightly less the case when shares are computed from the administrative data where both factors appear to be roughly as important as the other. Combining both sets of regressors usually highlights the particularly strong role of the variables that were constructed from the same source as the dependent variable. 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 towards there, this is not linked to 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. 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.[19]

However, once we take a look at the actual distribution of immigrants by occupation, recorded ex-post in the Census, as shown in the last three columns, a different conclusion emerges. In that case, all of our predictive power lies in the past attractiveness of this state for a given occupation. The coefficient on the prediction based on the past attractiveness of the state for each ethnicity is generally insignificant and small. Combining the regressors as in Panel C does little to alter the results from Panel B. In this case, we would thus conclude that a state would enumerate a larger number of immigrants of a given occupation because 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 various share measures. We also confirmed that the effect we find is not simply driven by similar patterns among immigrants, regardless of their ethnicity. In Table 5, we add as additional regressors, the counterfactual distribution one would obtain by allocating each period's flow using similar ethnicities. Formally, we include all ethnicities within an ethnic group as defined in Table A-3 except one's own. For example, Irish immigrants are allocated using the geographical distribution from 1900 IPUMS of other British Isles, Australia and Canada. 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 network measures are constructed, we now turn to the estimation of the impact of our three distinct measures of networks: "pure" ethnic, "pure" 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 do such a distinction.

---

[19]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.

Table 6 first presents the results of estimating equation (3) for the location choice of individuals by ethnicity. This table presents once more the regressions where ex-ante decisions are explored in the first three columns and those about ex-post decisions are in the following three. The results from columns (1) and (2) suggest that the previous importance of ethnic networks in the ex-ante location decisions of immigrants was not linked to labor markets insofar as they are now fully captured by the presence of individuals of the same ethnic group but of a different occupation. This is particularly the case in column (1) while in column (2), the results are very noisy thus making it difficult to reject the hypothesis that one set of predictors is more important than another. Once we combine all regressors, we find a similar results except that the number of individuals of one's ethnicity *and* occupation appears to be potentially also an important determinant of location but only when used to allocate individuals by their ex-post occupations. The regressions look quite different once we attempt to predict the location choices of individuals once they have settled in the United States. Already in column (4), the relative role of the attractiveness of a given state based on one's occupation was larger than the role of one's ethnicity. But, as we start using the ex-post national distribution of occupations and ethnicities from the IPUMS as regressors, the best fit to the actual distribution we can find becomes the counterfactual based on the number of individuals who selected this state in the past who shared both an occupation and an ethnicity with the new migrants. Taking the example of Italian bakers, these results would suggest that a state which had received 10 percent more immigrants of that ethnic group but a different occupation in the past would attract 11 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 10 percent more of these immigrants ex-post.

We explore the robustness of these results in various ways. We first estimated these regressions for each of our nine 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. We only report regressions corresponding to columns (1) and (5) of table 6 but similar results were obtained in other cases. Columns (1) and (2) contrast the results for groups whose inflow was above or below the median during the entire period. These results first suggest the pattern highlighted above is particularly strong for larger ethnic groups. In this case, ex-ante decisions are particularly linked to "pure" ethnic networks while ex-post decisions appear strongly driven by the presence of individuals of one's ethnic group *and* one's occupation. Small groups, on the other hand, rely particularly on labor market attractiveness to select their location, both ex-ante and ex-post. Ethnic components do appear to be slightly more important in the case of the ex-post distribution for small ethnicities.

We also computed the above regressions for different time periods. Particularly column (3) present the results for immigrants arriving before the First World War (1905-1915) while column

(4) focuses on immigrants who arrived afterwards (1915-1930). That first set of years corresponds to the largest immigration waves but also to the least restricted period regarding occupations and origin. We find that the pattern we estimated previously is particularly accurate for cohorts arriving before WWI. For those, the ex-ante distribution is best matched by a counterfactual based on the location choices of individuals of one's ethnicity while the ex-post distribution seems to be best explained by the presence of ethnic-specific occupational networks. For immigrants arriving after 1915, the presence of individuals sharing both one's occupation and one's ethnicity appears to be influence the location decisions both ex-ante and ex-post. This may be due to the fact that fewer immigrants with less skills were admitted after WWI probably as a consequence of the dramatic changes in immigration legislation in the United States.

Table 8 turns to the results obtained for equation (3) but this time using occupation instead of ethnicity as the variable of interest. Once more the "pure" ethnic effect appears particularly crucial when one is using on the right-hand side of the estimating equation the ex-ante distribution of occupations. In those regressions, the ethnic-specific occupation variable does not appear to have any explanatory power. On the other hand, the "pure" labor market effect becomes more important as one tries to explain the ex-post distribution of occupations than when trying to estimate the changes in ex-ante occupational distribution. Once more, it's when one attempts to allocate immigrants by their ex-post distribution of occupation that the role of ethnic-specific labor markets takes most of its importance, although in this case, the size of the coefficient remains more similar to that of the "pure" labor market effect. Taken together, this would suggest that a state increases the number of individuals of a given ethnic occupation who claim to want to settle there because that state was attractive in the past to ethnicities which now have more individuals with those skills at arrival. On the other hand, we would observe an increase in the number of migrants actually practicing a given occupation in a state mostly because that state in the past was popular for individuals of that given occupation (irrespective of their ethnicity) than because that state attracted individuals of a particular ethnicity and there are now more immigrants of that ethnicity practicing that occupation nationally.

Table 9 explore how these results change depending on the sub-sample one looks at. It first contrasts large and small occupations, as defined by occupations which had more or less than the median among all immigrants over this period. The results presented in Columns (1) and (2) first suggest that the ex-ante location decisions of individuals appear to be driven by the importance of their ethnic network, irrespective of the size of the occupation. On the other hand, the relative role of general labor market attractiveness and that of ethnic-specific labor market attractiveness in shaping the ex-post decision depends greatly on the size of the occupation. In the case of the smaller occupations, an immigrant is not more likely to report working in a given occupation in a state where individuals of their own ethnic group practiced their occupation in the past but simply if that state had many immigrants of the same ethnicity or the same occupation located

there in 1900. In the case of larger occupations, the past attractiveness of that state for one's given ethnicity *and* occupation is slightly more important than the fact that it attracted individuals of that occupation but from a different ethnicity. This is consistent with a situation where networks need a certain size to play a significant role, with smaller occupations is less likely that a specific ethnic group will have the chance to enough of a concentration in a particular state.

We also estimated the above regression model for different time periods and we present the results in columns (3)and (4) of Table 9. The importance of ethnic-specific occupation networks appear to be particularly important before the First World War with the "pure" ethnic effect remaining much stronger in the post-War period. However, the results before WWI are not very precisely estimation, although the point estimates of the coefficient for the pure occupation and the ethnicity occupation proxies are relatively robust, which makes comparing the two sets of regressions more difficult. The "pure" ethnic effect is larger after WWI which may be because family reunification was more encouraged in the later immigration periods. Nevertheless, because the estimates are somewhat noisier, it is difficult to draw very strong conclusions from these results.

5.3   Location Decision by Ethnicity and Occupation

To complete our analysis, we turn to regressions that attempt to explain the geographic distribution of immigrants of a particular ethnic group in a particular occupation and relate those to the importance of the "pure" ethnic effect, the "pure" occupation effect and to that of ethnic-specific occupation networks as presented in equation (1). These results are presented in Table 10. The results are extremely similar when one include or excludes individuals without an occupation, we present here only the results with individuals that report an occupation.[20]

As we explained before, we can only use Census, and in particular the IPUMS microsamples, to construct the left-hand side variable and the measures of the strenght of the networks. We present the basic results for equation (1) in columns (1) through (3). One finds that, in all three sets of regressions, the importance of labor markets is particularly marked. The difference is somewhat substantial between the regressions where the predictors are constructed from the administrative data inflow compared to those made from the IPUMS flows. In particular, the role of "pure" labor markets is much stronger compared to that of the ethnic-specific labor market effect when using the administrative data. The ethnicity-occupation combination is, on the other hand, dominating when using IPUMS to construct our measures of predicted flows. When we add both sets of predictive values we observe that it is actually the ethnic-specific occupation network, as measured using IPUMS data, that actually has the largest coefficient.

---

[20]Results that also include individuals without an occupation in the sample are available upon request from the authors.

The next six columns of Table 10 explore the robustness of the results presented above. Columns (4) and (5) compare the results for immigrants arriving before WWI to those arriving afterwards. Before WWI, the attractiveness of a location appears to have been deeply rooted in the presence of individuals of one's ethnic group practicing the same occupation. None of the other predictors is either large in magnitude or significant for that period. For the period after 1915, a different conclusion altogether arises. The coefficients on the prediction based on "pure" ethnic networks and those particular to one's ethnic-occupation pair are the only significant and large ones and they are roughly of the same magnitude. This indicates that location choices over that period mostly depended on the past location choice of one's countrymen, whether or not they share one's occupation.

Columns (6) and (7) then divide the sample by the relative size of the ethnic group as explained before. In this case, we find that only in the case of large ethnic groups does the presence of individuals of one's ethnicity and occupation determine one's location choice. In that case, it is the only strong predictor of location. This is logical since one would need a large enough network to use connection for labor-market purposes. On the other hand, for small ethnic groups, location decisions are mostly determined by labor markets and we find little evidence that ethnic-specific occupation networks are determining ex-post location decisions.

Finally, the last two columns of Table 10 compare the results for large and small occupations. For large occupations, the results are almost identical to those for the full sample where the past location of individuals from the same ethnicity and occupation is the most important determinant of ex-post decisions. For small occupations, all three regressors constructed from IPUMS are roughly similar suggesting an equal weight for occupation and ethnic networks and no additional role for individuals who share both characteristics.

We've also estimated these equations separately for groups of ethnic groups and the results were fairly similar although very noisy. Overall, these results suggest that explaining the ex-post location choices of individuals, even within this extremely detailed regression framework, lead to similar results where there is a particularly important role for the presence of individuals who shared one's ethnicity *and* occupation in the past.

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, "pure" ethnic networks and, to a lesser extent, "pure" 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

20

individuals of the same ethnicity *and* occupation.

The theoretical framework presented in Section 2 suggests that this could be due to the fact that newly arrived immigrants may take the occupations of individuals in their ethnic networks at their arrival as a tool to smooth the transition to the new labor market. If this were the case, we would expect this phenomenon to be particularly important for individuals who arrive in the United States with low levels of skills. A doctor is much less likely to change his occupation than a general laborer, for example. This hypothesis is explored in the first three columns of Table 11. We first see that for professionals, both ex-ante and ex-post decisions appear to be linked to the past location choices of individuals of their own ethnicity or their own occupation but no role appears in either decision for the interaction ethnic-specific occupation network. The observed pattern for skilled tradesmen is quite different. Those individuals appear to first select their desired location based on the presence of individuals of their own ethnic group, and maybe slightly more so if these individuals also share the same occupation. Ex-post, labor market forces, both outside and within one's ethnic group, are stronger determinants of location choices. Finally, it is for the unskilled that the pattern we highlighted above is the most visible. Ex-ante decisions appear to be driven by the location choices of individuals of the same ethnicity, irrespective of their occupation. However, ex-post, we find that the relative importance of ethnic-specific labor market considerations increases substantially. This pattern is consistent with our hypothesis that individuals with the lowest levels of human capital would be more likely to change occupation upon arrival based on the opportunities of their ethnic networks.

Furthermore, our framework does suggest that the importance of ethnic-specific occupational networks should decrease as an immigrant passes more time in his new country. The last two columns of Tables 11 and 12 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 "pure" ethnicity and "pure" 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 11. 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.[21]

Two other factors could explain the difference between ex-ante and ex-post decisions. First, it

---

[21]Results are available upon request.

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 12 estimates 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. The results presented in Table 12 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. 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 consistent 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 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 their ex-post location decisions may respond to different factors. 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.

We can further explore the role of internal migration by returning to the results of the tables presented in the previous section. 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 would have a distinct pattern than the one presented above. 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 equation (4) or equation (3) whether or not the left-hand side variable is built using administrative or Census data. Contrasting the results of columns (1) to (2) and (4) to (5) in Table 6 and Table 8 clearly does not appear to give credential to this hypothesis, particularly not when the left-hand side variable measures the ex-post distribution of immigrants. While we are willing to admit that this is not a direct test of the hypothesis, we feel comfortable saying that internal migration cannot be the only reason through which this adjustment is made.

7   Conclusion

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 early years of the twentieth century. We have shown that the presence of individuals of one's ethnicity is strongly influential in the intended location reported by an immigrant at her arrival in the United States. However, ex-post 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.

This is an important finding in helping us understand how networks are playing a role in the location choice of immigrants. While most studies have looked individually 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, a fact that has not been carefully documented before. It even seems to indicate that the ex-ante distribution of immigrants across states by ethnicity is particularly influenced by the role of local ethnic networks. While we are not presenting direct individual level evidence on the role of networks specifically, these conclusions are consistent with ethnic networks, both related and unrelated to labor markets. More work must be devoted to better argue that the pattern we document and estimate in this paper is due to networks.

23

While our specific data allows us to look at a historical period, our results might be relevant for today's immigration debate. This past immigration wave resembled that of today in many aspects: migrants were less skilled than natives (or at least were perceived to be less skilled), they represented a substantial fraction of the population and increased fairly rapidly to then be limited by changes in immigration policy, 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.

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. It can also have different implications for the type of immigrants a country might target as well as how it may restrict their location decisions.

References

Altonji, J. G. and D. Card (1991). The effects of immigration on the labor market outcomes of less-Skilled natives. In J. M. Abowd and R. B. Freeman (Eds.), *Immigration, Trade and Labor*, Chicago, IL. University of Chicago Press.

Beaman, L. (2008). Social networks and the dynamics of labor market outcomes: Evidence from refugees resettled in the U.S. Mimeo, Northwestern University.

Calvo-Armengol, A. and M. O. Jackson (2004, June). The effects of social networks on employment and inequality. *American Economic Review 94*(3), 426–454.

Calvo-Armengol, A. and M. O. Jackson (2007, January). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory 132*(1), 27–46.

Card, D. (2001, January). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics 19*, 22–64.

Card, D. and E. Lewis (2005, August). The diffusion of mexican immigrants during the 1990s: Explanations and impacts. NBER Working Paper 11552, National Bureau of Economic Research.

Cortés, P. (2008, June). The effect of low-skilled immigration on U.S. prices: Evidence from cpi data. *Journal of Political Economy 116*(3), 381–422.

Cortés, P. and J. A. Tessada (2009, June). Low-skilled immigration and the labor supply of highly educated women. Mimeo The Brookings Institution and Chicago GSB.

Fairchild, H. P. (1925). *Immigration: a world movement and its American significance* (rev. edition ed.). Macmillan Co.

Federman, M., D. Harrington, and K. Krynski (2006). Vietnamese manicurists: Displacing natives or finding new nails to polish? *Industrial and Labor Relations Review 59*(2), 302–318.

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

Lafortune, J. (2008). Making yourself attractive: Pre-marital investments and the returns to education in the marriage market. Mimeo, University of Maryland.

Manski, C. F. (1993, July). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies 60*(3), 531–542.

Munshi, K. (2003). Networks in the modern economy: Mexican immigrants in the US labor market. *Quarterly Journal of Economics 118*(2), 549–599.

Munshi, K. and N. Wilson (2008, April). Identity, parochial institutions, and career decisions: Linking the past to the present in the american midwest. Mimeo, Brown University.

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

Patel, K. and F. Vella (2007, December). Immigrant networks and their implications for occupational choice and wages. Discussion Paper 3217, IZA.

Peri, G. and C. Sparber (2009, July). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics 1*(3), 135–69.

Ruggles, S., M. Sobek, T. Alexander, C. A. Fitch, R. Goeken, P. K. Hall, M. King, and C. Ronnander (2009). Integrated public use microdata series: Version 4.0 [machine-readable database].

Table 1. 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 2. 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
Hispanic	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
Hispanic					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 3. Explaining changes in location choices of immigrants by ethnicity (excluding individuals with no occupation)

| | Flows from administrative data | | | Flows from IPUMS | | |
	Admin	Census	IPUMS	Admin	Census	IPUMS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Admin flows pred. by ethnicity	0.623***	0.916***	0.838***	0.146**	0.198***	0.177***
	(0.085)	(0.048)	(0.089)	(0.051)	(0.043)	(0.050)
Admin flows pred. by occupation	0.390**	0.641***	0.537***	0.105	0.295**	0.214**
	(0.130)	(0.132)	(0.127)	(0.055)	(0.091)	(0.080)
R-square	0.963	0.947	0.936	0.777	0.775	0.768
N	7140	7140	7140	7140	7140	7140
Panel B						
IPUMS flows pred. by ethnicity	1.768**	2.239**	2.147**	0.273*	0.590***	0.539***
	(0.645)	(0.681)	(0.739)	(0.108)	(0.076)	(0.098)
IPUMS flows pred. by occupation	-0.176	-0.111	0.099	0.540***	0.794***	0.626**
	(0.561)	(0.814)	(0.591)	(0.146)	(0.229)	(0.192)
R-square	0.837	0.836	0.832	0.907	0.929	0.930
N	7140	7140	7140	7140	7140	7140
Panel C						
Admin flows pred. by ethnicity	0.562***	0.857***	0.768***	0.065*	0.013	0.008
	(0.104)	(0.058)	(0.107)	(0.028)	(0.024)	(0.028)
Admin flows pred. by occupation	0.416**	0.608***	0.449***	-0.070	-0.112	-0.048
	(0.138)	(0.167)	(0.133)	(0.044)	(0.091)	(0.070)
IPUMS flows pred. by ethnicity	0.248	0.228	0.237	0.161	0.573***	0.532***
	(0.171)	(0.173)	(0.232)	(0.143)	(0.105)	(0.130)
IPUMS flows pred. by occupation	-0.133	-0.012	0.203	0.620***	0.886**	0.665**
	(0.148)	(0.269)	(0.208)	(0.170)	(0.279)	(0.238)
R-square	0.964	0.948	0.939	0.910	0.930	0.930
N	7140	7140	7140	7140	7140	7140

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnic group electing a particular state in a particular period of migration. This flow is measured from the administrative data (ex-ante) in the first three columns but from the IPUMS (ex-post) in the last three. The right-hand side variables are predicted flows by ethnicity, by state, by period of immigration constructed as detailed by the variables listed in each panel. 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 and period of immigration.

Standard errors are clustered at the ethnic-state level.

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

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

| | Flows from administrative data | | | Flows from IPUMS data | | |
	Admin	Census	IPUMS	Admin	Census	IPUMS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Admin flows pred. by ethnicity	0.549***	1.124***	1.163***	-0.026	0.158	0.055
	(0.150)	(0.170)	(0.238)	(0.283)	(0.116)	(0.107)
Admin flows pred. by occupation	0.344*	0.109	-0.166	0.228	0.366**	0.328*
	(0.150)	(0.141)	(0.226)	(0.273)	(0.134)	(0.133)
R-square	0.969	0.937	0.930	0.727	0.764	0.766
N	15810	15810	15810	15810	15810	15810
Panel B						
IPUMS flows pred. by ethnicity	0.419**	1.201***	1.045***	0.121	0.437***	0.102
	(0.139)	(0.204)	(0.318)	(0.108)	(0.092)	(0.104)
IPUMS flows pred. by occupation	0.641**	0.156	0.296	0.564**	0.835***	0.918***
	(0.225)	(0.373)	(0.374)	(0.185)	(0.174)	(0.141)
R-square	0.688	0.676	0.677	0.931	0.944	0.960
N	15810	15810	15810	15810	15810	15810
Panel C						
Admin flows pred. by ethnicity	0.494**	1.130***	1.151***	0.065	0.009	-0.014
	(0.167)	(0.215)	(0.244)	(0.058)	(0.012)	(0.011)
Admin flows pred. by occupation	0.403*	0.029	-0.258	-0.097	-0.050	-0.005
	(0.157)	(0.074)	(0.200)	(0.069)	(0.033)	(0.015)
IPUMS flows pred. by ethnicity	0.126	-0.122	-0.043	0.080	0.428***	0.117
	(0.109)	(0.220)	(0.201)	(0.102)	(0.092)	(0.103)
IPUMS flows pred. by occupation	-0.122	0.381	0.321	0.638***	0.889***	0.925***
	(0.112)	(0.266)	(0.226)	(0.177)	(0.192)	(0.143)
R-square	0.970	0.938	0.933	0.933	0.945	0.960
N	15810	15810	15810	15810	15810	15810

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. This flow is measured from the administrative data (ex-ante) in the first three columns but from the IPUMS (ex-post) in the last three. The right-hand side variables are predicted flows by occupation, by state, by period of immigration constructed as detailed by the variables listed in each panel. 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, occupation and period of immigration.

Standard errors are clustered at the occupation-state level.

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

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

| | Flows from administrative data | | | Flows from IPUMS | | |
	Admin	Census	IPUMS	Admin	Census	IPUMS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Explaining location choices by ethnicity (N=6885)						
Admin flows pred. by ethnicity	0.625***	0.898***	0.839***	0.146**	0.193***	0.177***
	(0.086)	(0.057)	(0.088)	(0.051)	(0.042)	(0.050)
Admin flows pred. by occupation	0.378**	0.604***	0.506***	0.104	0.287**	0.211*
	(0.133)	(0.131)	(0.123)	(0.057)	(0.087)	(0.086)
Admin flows pred. by close eth. grp	0.015	0.078	0.025	0.001	0.020	0.003
	(0.008)	(0.050)	(0.052)	(0.011)	(0.015)	(0.016)
R-square	0.964	0.949	0.937	0.777	0.776	0.764
Panel B: Explaining location choices by ethnicity (N=6885)						
IPUMS flows pred. by ethnicity	1.774**	2.233**	2.148**	0.274*	0.577***	0.536***
	(0.653)	(0.721)	(0.755)	(0.110)	(0.075)	(0.097)
IPUMS flows pred. by occupation	-0.229	-0.145	0.074	0.522**	0.775**	0.620**
	(0.585)	(0.795)	(0.554)	(0.166)	(0.253)	(0.215)
IPUMS flows pred. by close eth. grp	0.075	0.039	0.023	0.035	0.059	0.018
	(0.224)	(0.245)	(0.236)	(0.051)	(0.040)	(0.052)
R-square	0.837	0.836	0.832	0.909	0.930	0.930
Panel C: Explaining location choices by occupation (N=15810)						
Admin flows pred. by ethnicity	0.541***	1.063***	1.309***	-0.025	0.152	0.161
	(0.150)	(0.163)	(0.234)	(0.288)	(0.165)	(0.154)
Admin flows pred. by occupation	0.324*	0.116	-0.124	0.231	0.366**	0.358**
	(0.139)	(0.139)	(0.227)	(0.270)	(0.133)	(0.137)
Admin flows pred. by close eth. grp	0.034	0.083	-0.223*	-0.005	0.009	-0.163
	(0.029)	(0.090)	(0.099)	(0.096)	(0.101)	(0.135)
R-square	0.969	0.937	0.933	0.727	0.764	0.775
Panel D: Explaining location choices by occupation (N=15810)						
IPUMS flows pred. by ethnicity	0.406**	1.123***	1.041***	0.110	0.370***	0.120
	(0.129)	(0.185)	(0.277)	(0.100)	(0.086)	(0.092)
IPUMS flows pred. by occupation	0.567*	0.131	0.284	0.501**	0.813***	0.983***
	(0.223)	(0.379)	(0.437)	(0.184)	(0.180)	(0.119)
IPUMS flows pred. by close eth. grp	0.141**	0.187	0.024	0.122*	0.159	-0.136
	(0.054)	(0.149)	(0.189)	(0.051)	(0.111)	(0.095)
R-square	0.689	0.677	0.677	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 and B) or a particular occupation (Panels C and D) electing a particular state in a particular period of migration. This flow is measured from the administrative data (ex-ante) in the first three columns but from the IPUMS (ex-post) in the last three. The right-hand side variables are predicted flows by ethnicity, by state, by period of immigration constructed as detailed by the variables listed in each panel. 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 ethnic-state level.

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

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

	Flows from administrative data			Flows from IPUMS data		
	(1)	(2)	(3)	(4)	(5)	(6)
Admin flows pred. by pure ethnicity	1.129***		0.988***	0.185***		0.002
	(0.171)		(0.205)	(0.050)		(0.031)
Admin flows pred. by pure occupation	0.591**		0.540**	0.236**		0.018
	(0.155)		(0.172)	(0.077)		(0.033)
Admin flows pred. by ethnicity-occupation	-0.128		-0.158	0.025		-0.007
	(0.153)		(0.128)	(0.057)		(0.023)
IPUMS flows pred. by pure ethnicity		1.770	-0.510		-0.427*	-0.434*
		(1.510)	(0.466)		(0.216)	(0.217)
IPUMS flows pred. by pure occupation		0.030	-0.233		0.232***	0.213*
		(0.429)	(0.285)		(0.067)	(0.086)
IPUMS flows pred. by ethnicity-occupation		0.443	1.203**		1.203***	1.215***
		(1.079)	(0.463)		(0.219)	(0.219)
R-square	0.925	0.833	0.935	0.753	0.951	0.951
N	7140	7140	7140	7140	7140	7140

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnic group electing a particular state in a particular period of migration. This flow is measured from the administrative data (ex-ante) in the first three columns but from the IPUMS (ex-post) in the last three. The right-hand side variables are predicted flows by ethnicity, by state, by period of immigration constructed as detailed by the variables. All location shares are obtained from 1900 IPUMS. 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

31

Table 7. Explaining changes in location choices of immigrants by ethnicity, by different groups and period

	Small	Large	Pre-War	Post-War
	(1)	(2)	(3)	(4)
Panel A: Flows from administrative data				
Admin flows pred. by pure ethnicity	0.389	1.101***	0.845	0.146
	(0.311)	(0.173)	(0.763)	(0.407)
Admin flows pred. by pure occupation	0.829***	0.624***	0.612	0.615**
	(0.249)	(0.165)	(0.345)	(0.193)
Admin flows pred. by ethnicity-occupation	0.136	-0.137	-0.076	1.029*
	(0.186)	(0.143)	(0.406)	(0.447)
R-square	0.793	0.935	0.984	0.958
N	3570	3570	2856	4284
Panel B: Flows from IPUMS				
IPUMS flows pred. by pure ethnicity	0.419	-0.443*	-0.347	-0.480
	(0.215)	(0.224)	(0.305)	(0.546)
IPUMS flows pred. by pure occupation	0.345*	0.215**	0.225	0.249
	(0.166)	(0.072)	(0.132)	(0.138)
IPUMS flows pred. by ethnicity-occupation	0.233	1.244***	1.145***	1.498***
	(0.156)	(0.229)	(0.320)	(0.546)
R-square	0.748	0.956	0.968	0.935
N	3570	3570	2856	4284

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnic group electing a particular state in a particular period of migration. This flow is measured from the administrative data (ex-ante) in Panel A but from the IPUMS (ex-post) in Panel B. The right-hand side variables are predicted flows by ethnicity, by state, by period of immigration constructed as detailed by the variables. All location shares are obtained from 1900 IPUMS. Column (1) is computed only for the ethnic groups that had an overall flow over the period smaller than the median while Column (2) uses those which total flow exceeded the median. Column (3) only includes 1905-1914 while column (4) includes 1915-1930 as periods of migration. 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 8. Explaining changes in location choices of immigrants by occupation, networks more finely defined

	Flows from administrative data			Flows from IPUMS data		
	(1)	(2)	(3)	(4)	(5)	(6)
Admin flows pred. by pure ethnicity	1.128***		1.147***	0.041		0.003
	(0.236)		(0.246)	(0.098)		(0.007)
Admin flows pred. by pure occupation	-0.088		-0.132	0.202		0.029
	(0.160)		(0.149)	(0.113)		(0.022)
Admin flows pred. by ethnicity-occupation	-0.030		-0.101	0.126		-0.042**
	(0.101)		(0.076)	(0.131)		(0.014)
IPUMS flows pred. by pure ethnicity		1.187*	-0.507		-0.194*	-0.200*
		(0.572)	(0.395)		(0.098)	(0.098)
IPUMS flows pred. by pure occupation		0.692**	-0.111		0.487***	0.470***
		(0.245)	(0.196)		(0.100)	(0.091)
IPUMS flows pred. by ethnicity-occupation		-0.427	0.759		0.608***	0.647***
		(0.657)	(0.442)		(0.116)	(0.114)
R-square	0.928	0.677	0.934	0.771	0.966	0.967
N	15810	15810	15810	15810	15810	15810

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. This flow is measured from the administrative data (ex-ante) in the first three columns but from the IPUMS (ex-post) in the last three. The right-hand side variables are predicted flows by occupation, by state, by period of immigration constructed as detailed by the variables. All location shares are obtained from 1900 IPUMS. 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. Explaining changes in location choices of immigrants by occupations, by different groups and period

	Small	Large	Pre-War	Post-War
	(1)	(2)	(3)	(4)
Panel A: Flows from administrative data				
Admin flows pred. by pure ethnicity	1.435***	1.121***	1.638	0.661***
	(0.531)	(0.234)	(0.870)	(0.192)
Admin flows pred. by pure occupation	-0.091	-0.085	-0.127	0.156
	(0.256)	(0.159)	(0.606)	(0.125)
Admin flows pred. by ethnicity-occupation	0.067	-0.030	-0.450	0.331
	(0.351)	(0.098)	(0.878)	(0.199)
R-square	0.902	0.930	0.984	0.973
N	7905	7905	6324	9486
Panel B: Flows from IPUMS				
IPUMS flows pred. by pure ethnicity	0.386*	-0.191*	-0.254	0.613
	(0.182)	(0.097)	(0.236)	(0.326)
IPUMS flows pred. by pure occupation	0.690**	0.483***	0.445	0.318
	(0.235)	(0.098)	(0.335)	(0.289)
IPUMS flows pred. by ethnicity-occupation	0.105	0.607***	0.662	0.458
	(0.242)	(0.116)	(0.442)	(0.418)
R-square	0.923	0.967	0.979	0.937
N	7905	7905	6324	9486

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. This flow is measured from the administrative data (ex-ante) in Panel A but from the IPUMS (ex-post) in Panel B. The right-hand side variables are predicted flows by ethnicity, by state, by period of immigration constructed as detailed by the variables. All location shares are obtained from 1900 IPUMS. Column (1) is computed only for the occupation that had an overall flow over the period smaller than the median while Column (2) uses those which total flow exceeded the median. Column (3) only includes 1905-1914 while column (4) includes 1915-1930 as periods of migration. All regressions include fixed effects for the double interactions of state, occupations and period of immigration.

Standard errors are clustered at the occupation-state level.

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

Table 10. Explaining changes in location choices of immigrants by occupation and ethnicity.

	Full sample			By time period		By ethnic group		By occupation	
				Pre-war	Post-war	Large	Small	Large	Small
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Admin flows pred. by pure ethnicity	0.058		0.010						
	(0.048)		(0.011)						
Admin flows pred. by pure occupation	0.110***		-0.036*						
	(0.030)		(0.014)						
Admin flows pred. by ethnicity-occupation	0.058		-0.015						
	(0.048)		(0.010)						
IPUMS flows pred. by pure ethnicity		-0.138	-0.139	-0.218	0.458*	-0.230	0.170	-0.138	0.145
		(0.111)	(0.109)	(0.127)	(0.204)	(0.127)	(0.107)	(0.109)	(0.080)
IPUMS flows pred. by pure occupation		0.149**	0.170**	0.091	-0.092	0.002	0.225*	0.148***	0.334***
		(0.052)	(0.057)	(0.069)	(0.103)	(0.060)	(0.108)	(0.051)	(0.076)
IPUMS flows pred. by ethnicity-occupation		0.831***	0.841***	0.898***	0.524**	0.950***	0.227*	0.831***	0.484***
		(0.091)	(0.092)	(0.100)	(0.193)	(0.098)	(0.103)	(0.088)	(0.076)
R-square	0.110	0.740	0.743	0.700	0.528	0.700	0.172	0.743	0.364
N	531216	531216	531216	265608	265608	265608	265608	265608	265608

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnicity, in a particular occupation electing a particular state in a particular period of migration, measured from the IPUMS (thus, ex-post). The right-hand side variables are predicted flows by ethnicity, by state, by period of immigration constructed as detailed by the variables listed in each panel using shares from the 1900 IPUMS. The first 3 columns include all possible ethnic-occupation-state-time cells. Column (4) restrict the sample to immigrants arriving before 1915, column (5) to those arriving after that date. Columns (6) and (8) include all ethnic groups/occupations where the flow over the 1900-1930 period was larger than the median while columns (7) and (9) only include those with flows less than the median. All regressions include fixed effects for the triple interactions of state, ethnicity, occupation and period of immigration.

Standard errors are clustered at the ethnic-state level.

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

Table 11. Exploring the differences between ex-ante and ex-post location choices, by occupation.

	Professionals	Skilled	Unskilled	Within 5 years	Between 5-10 years
	(1)	(2)	(3)	(4)	(5)
Panel A: Flows from administrative data					
Admin flows pred. by pure ethn.	0.558*	0.691***	1.116***	1.165***	1.085***
	(0.230)	(0.138)	(0.224)	(0.287)	(0.242)
Admin flows pred. by pure occ.	0.509*	-0.052	-0.086	-0.116	-0.110
	(0.243)	(0.107)	(0.156)	(0.212)	(0.153)
Admin flows pred. by ethn.-occ.	0.287	0.679***	-0.029	-0.018	-0.057
	(0.189)	(0.121)	(0.091)	(0.127)	(0.094)
R-square	0.961	0.973	0.931	0.890	0.947
N	3570	9180	3060	9486	6324
Panel B: Flows from IPUMS					
IPUMS flows pred. by pure ethn.	0.127	0.004	-0.294	-0.183	-0.217
	(0.355)	(0.039)	(0.155)	(0.112)	(0.153)
IPUMS flows pred. by pure occ.	1.096	0.770***	0.468***	0.443***	0.901**
	(0.799)	(0.179)	(0.105)	(0.136)	(0.330)
IPUMS flows pred. by ethn.-occ.	0.027	0.324*	0.709***	0.642***	0.271
	(0.301)	(0.145)	(0.166)	(0.148)	(0.242)
R-square	0.857	0.949	0.970	0.967	0.967
N	3570	9180	3060	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. This flow is measured from the administrative data (ex-ante) in Panel A but from the IPUMS (ex-post) in Panel B. The right-hand side variables are predicted flows by occupation, by state, by period of immigration constructed as detailed by the variables. All location shares are obtained from 1900 IPUMS. The first 3 columns divide the sample by the type of occupations: professionals, skilled and unskilled. Column (4) 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 12. Exploring the differences between ex-ante and ex-post location choices, by ethnicity.

	Low return ethnicity	High return ethnicity	Within 5 years	Between 5-10 years
	(1)	(2)	(3)	(4)
Panel A: Flows from administrative data				
Admin flows pred. by pure ethn.	1.165***	0.885***	1.134***	1.127***
	(0.196)	(0.191)	(0.225)	(0.172)
Admin flows pred. by pure occ.	0.689**	0.113	0.657**	0.401*
	(0.212)	(0.116)	(0.199)	(0.175)
Admin flows pred. by ethnicity-occ.	-0.218	0.300	-0.126	-0.144
	(0.164)	(0.160)	(0.222)	(0.110)
R-square	0.931	0.922	0.879	0.947
N	3570	3570	4284	2856
Panel B: Flows from IPUMS				
IPUMS flows pred. by pure ethn.	-0.473*	-0.252	-0.461	-0.315
	(0.237)	(0.168)	(0.245)	(0.295)
IPUMS flows pred. by pure occ.	0.206**	0.232	0.143*	0.584***
	(0.068)	(0.168)	(0.062)	(0.136)
IPUMS flows pred. by ethnicity-occ.	1.279***	0.987***	1.307***	0.799*
	(0.245)	(0.332)	(0.246)	(0.312)
R-square	0.968	0.893	0.951	0.948
N	3570	3570	4284	2856

The left-hand side variable of the regressions presented in this table is the flow of immigrants from a particular ethnic group electing a particular state in a particular period of migration. This flow is measured from the administrative data (ex-ante) in Panel A but from the IPUMS (ex-post) in Panel B. The right-hand side variables are predicted flows by ethnicity, by state, by period of immigration constructed as detailed by the variables. All location shares are obtained from 1900 IPUMS. The first column restrict the sample to ethnic groups with rates of return migration below that of the median, the next to groups with rates of return migration above it. Column (3) 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

A Additional Tables

<div align="center">**Table A-1.** Summary Statistics</div>

Variable	N	Mean	Standard deviation
Data set on flows by ethnicity-state-period			
Flow from administrative data	8568	2172.24	12656.07
Flow from IPUMS	8568	774.26	4103.93
Pred. flow based on ethn. (all admin shares, admin flow)	8568	2172.24	12618.09
Pred. flow based on ethn. (early admin shares, admin flow)	8568	2172.24	14097.17
Pred. flow based on ethn. (Census shares, admin flow)	8568	2170.83	11137.40
Pred. flow based on ethn. (IPUMS shares, admin flow)	8568	2171.24	11124.62
Pred. flow based on ethn. (all admin shares, IPUMS flow)	8568	774.26	4795.43
Pred. flow based on ethn. (early admin shares, IPUMS flow)	8568	774.26	5293.29
Pred. flow based on ethn. (Census shares, IPUMS flow)	8568	773.82	4087.17
Pred. flow based on ethn. (IPUMS shares, IPUMS flow)	8568	774.26	4007.12
Pred. flow based on occ. (all admin shares, admin flow)	8568	2175.35	10043.09
Pred. flow based on occ. (early admin shares, admin flow)	8568	2172.79	11284.60
Pred. flow based on occ. (Census shares, admin flow)	8568	1525.16	3927.16
Pred. flow based on occ. (IPUMS shares, admin flow)	8568	2152.28	7179.17
Pred. flow based on occ. (all admin shares, IPUMS flow)	8568	774.26	3863.86
Pred. flow based on occ. (early admin shares, IPUMS flow)	8568	772.11	4358.71
Pred. flow based on occ. (Census shares, IPUMS flow)	8568	763.62	2350.11
Pred. flow based on occ. (IPUMS shares, IPUMS flow)	8568	774.26	3040.25
Pred. flow based on pure occ. (admin flow)	8568	2151.29	7108.47
Pred. flow based on pure ethn. (admin flow)	8568	2151.29	11063.24
Pred. flow based on ethn.-occ. (admin flow)	8568	2110.58	10481.94
Pred. flow based on pure occ. (IPUMS flow)	8568	774.26	3007.65
Pred. flow based on pure ethn. (IPUMS flow)	8568	774.26	4008.40
Pred. flow based on ethn.-occ. (IPUMS flow)	8568	753.75	4079.75
Data set on flows by occupation-state-period			
Flow from administrative data	21216	615.03	6802.87
Flow from IPUMS	21216	308.39	2815.39
Pred. flow based on ethn. (all admin shares, admin flow)	21216	615.03	6481.79
Pred. flow based on ethn. (early admin shares, admin flow)	21216	614.00	7232.09
Pred. flow based on ethn. (Census shares, admin flow)	21216	615.03	3824.06
Pred. flow based on ethn. (IPUMS shares, admin flow)	21216	606.44	5010.54
Pred. flow based on ethn. (all admin shares, IPUMS flow)	21216	308.39	3074.37
Pred. flow based on ethn. (early admin shares, IPUMS flow)	21216	307.51	3600.83
Pred. flow based on ethn. (Census shares, IPUMS flow)	21216	308.38	1972.92
Pred. flow based on ethn. (IPUMS shares, IPUMS flow)	21216	308.39	2572.16
Pred. flow based on occ. (all admin shares, admin flow)	21216	615.94	6826.38

Variable	N	Mean	Standard deviation
Pred. flow based on occ. (early admin shares, admin flow)	21216	615.94	7474.97
Pred. flow based on occ. (Census shares, admin flow)	21216	615.54	5683.09
Pred. flow based on occ. (IPUMS shares, admin flow)	21216	615.59	5602.40
Pred. flow based on occ. (all admin shares, IPUMS flow)	21216	308.39	3352.56
Pred. flow based on occ. (early admin shares, IPUMS flow)	21216	308.39	3697.69
Pred. flow based on occ. (Census shares, IPUMS flow)	21216	308.21	2785.00
Pred. flow based on occ. (IPUMS shares, IPUMS flow)	21216	308.39	2735.42
Pred. flow based on pure occ. (admin flow)	21216	606.27	4975.31
Pred. flow based on pure ethn. (admin flow)	21216	606.27	5664.32
Pred. flow based on ethn.-occ. (admin flow)	21216	589.84	5444.40
Pred. flow based on pure occ. (IPUMS flow)	21216	308.39	2546.97
Pred. flow based on pure ethn. (IPUMS flow)	21216	308.39	2756.99
Pred. flow based on ethn.-occ. (IPUMS flow)	21216	300.10	2877.39
Data set on flows by ethnicity-occupation-state-period			
Flow from IPUMS	593844	11.02	252.65
Pred. flow based on pure occ. (admin flow)	593844	21.66	354.14
Pred. flow based on pure ethn. (admin flow)	593844	21.66	546.16
Pred. flow based on ethn.-occ. (admin flow)	593844	21.07	574.39
Pred. flow based on pure occ. (IPUMS flow)	593844	11.02	181.79
Pred. flow based on pure ethn. (IPUMS flow)	593844	11.02	244.05
Pred. flow based on ethn.-occ. (IPUMS flow)	593844	10.72	263.75

Table A-2. Matching of occupations between administrative and Census data

Administrative Data Final Groupings	1950 Occupation Classification		1900 Occupation Classification	Large?
	Code	Description		
		Total professional		
Electricians	515	Electricians	Electricians	No
	603	Apprentice electricians		
Engineers (professional)	41	Engineers, aeronautical	Engineers (civil, etc) and surveyors	Yes
	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
	31	Dancers and dancing teachers		
Literary and scientific persons	6	Authors	Literary and scientific persons	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
			Theatretical managers, etc	
Musicians	51	Entertainers (n.e.c.)	Musicians and teachers of music	Yes
	57	Musicians and music teachers	Professional showmen	
Teachers	12	Agricultural sciences	Teachers and professors in colleges, etc	Yes
	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
Officials (Government)	210	Inspectors, public administration	Officials (government)	No

Administrative Data Final Groupings	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?
	250	Officials and administrators (n.e.c.), public administration	Soldiers, sailors and marines (U.S.)	
	270	Postmasters		
	595	Members of the armed services		
	771	Marshals and constables		
	782	Sheriffs and bailiffs		
Physicians	75	Physicians and surgeons	Physicians and surgeons	No
Architects	3	Architects	Architects	No
Editors	36	Editors and reporters	Journalists	No
Lawyers	55	Lawyers and judges	Lawyers	No
Other professionals	2	Airplane pilots and navigators	Dentists	Yes
	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
Clerks and accountants	0	Accountants and auditors	Bookkeepers and accountants	Yes
	302	Attendants, physician's and dentist's office	Clerks and copyists	
	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		

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

Administrative Data Final Groupings	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?
Locksmiths	551	Mechanics and repairmen, office machine		
	552	Mechanics and repairmen, radio and television		
	553	Mechanics and repairmen, railroad and car shop		
	554	Mechanics and repairmen (n.e.c.)		
	600	Apprentice auto mechanics		
	605	Apprentice mechanics, except auto		
Printers	512	Compositors and typesetters	Printers, lithographers and pressmen	No
	575	Pressmen and plate printers, printing		
	613	Apprentices, printing trades		
Tinners	591	Tinsmiths, coppersmiths, and sheet metal workers	Tinplate and tinware makers	
Hat and Cap makers	645	Milliners	Hat and cap makers	No
Milliners			Milliners	
Seamstresses	633	Dressmakers and seamstresses, except factory	Seamstresses	Yes
Dressmakers			Dressmakers	
Shoemakers	582	Shoemakers and repairers, except factory	Boot and shoe makers and repairers	Yes
Tailors	590	Tailors and tailoresses	Tailors and tailoresses	Yes
Textile workers (not specified)	543	Loom fixers	Bleachery and dye works operatives	Yes
	634	Dyers	Carpet factory operatives	
			Cotton mill operatives	
			Hosiery and knittign mill operatives	
			Woolen mill operatives	
			Shirt, collar, and cuff makers	
Weavers and spinners	675	Spinners, textile	Other textile mill operatives	Yes
	684	Weavers, textile	Other textile workers	
Furriers and fur workers	525	Furriers		
Upholsterers	593	Upholsterers	Upholsterers	No
Mariners	240	Officers, pilots, pursers and engineers, ship	Boatmen and sailors	Yes
	623	Boatmen, canalmen, and lock keepers		
	673	Sailors and deck hands		
Bakers	500	Bakers	Bakers	Yes
Butchers	644	Meat cutters, except slaughter and packing house	Butchers	Yes
Millers	555	Millers, grain, flour, feed, etc.	Millers	Yes
Miners	650	Mine operatives and laborers	Miners and quarrymen	Yes
Photographers	74	Photographers	Photographers	No
	671	Photographic process workers		
Plasterers	573	Plasterers	Plasterers	No
Bookbinders	502	Bookbinders	Bookbinders	No
Engravers	521	Engravers, except photoengravers	Engravers	No
	571	Photoengravers and lithographers		
Pattern makers	570	Pattern and model makers, except paper	Model and pattern makers	No
Saddlers and harness makers	511	Cement and concrete finishers	Harness and saddle makers and repairs	Yes
Tanners and curriers	513	Cranemen, derrickmen, and hoistmen	Leather curriers and tanners	
Brewers	514	Decorators and window dressers	Brewers and malsters	
Cigar packers	522	Excavating, grading, and road machinery operators	Distillers and rectifiers	
Cigarette makers	523	Foremen (n.e.c.)	Tobacco and cigar factory operatives	
Tobacco workers	540	Linemen and servicemen, telegraph, telephone, and power	Switchmen, yardmen and flagmen	
Cigar makers	542	Locomotive firemen	Telegraph and telephone linemen	
Other skilled	562	Motion picture projectionists	Decorators and window dressers	
	563	Opticians and lens grinders and polishers	Weighters, gaugers, and measurers	
	565	Paperhangers	Paper hangers	
	572	Piano and organ tuners and repairmen	Roofers and slaters	
	581	Roofers and slaters	Brick and tile makers	

Administrative Data Final Groupings	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?
	592	Tool makers, and die makers and setters	Glass woekrs	
	594	Craftsmen and kindred workers (n.e.c.)	Potters	
	611	Apprentices, building trades (n.e.c.)	Butter and cheese makers	
	614	Apprentices, other specified trades	Confectionners	
	615	Apprentices, trade not specified	Trunk and leather-case makers	
	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

Administrative Data Final Groupings	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?
Farm laborers	640	Fruit, nut, and vegetable graders, and packers	Farm and plantation laborers	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
	123	Farm managers		
Fishermen	910	Fishermen and oystermen	Fishermen and oystermen	No
Laborers	690	Operative and kindred workers (n.e.c.)	Laborers (not specified)	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
	710	Laundresses, 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
	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
	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 chauffeurs		
	683	Truck and tractor drivers		
	960	Teamsters		
Manufacturers	290	Managers, officials, and proprietors (n.e.c.)	Foremen and overseers	No
	560	Millwrights	Manufacturers and officials	
Merchants and dealers	200	Buyers and department heads, store	Commercial travelers	Yes

Administrative Data Final Groupings	Code	1950 Occupation Classification Description	1900 Occupation Classification	Large?
	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	Auctioneers	
	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
	750	Bartenders	Boarding and lodging house keeprs	
	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
	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-3. Matching ethnic groups and countries.

Ethnic group	Ethnicities (from administrative data)	Country of birth (from Census)	High return?	Large?
British ancestry	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
	Finnish	Finland	No	No
Scandinavians	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

46