cemmap
centre for microdata methods and practice

Please call me John: name choice and the assimilation of immigrants in the United States, 1900-1930

Pedro Carneiro
Sokbae Lee
Hugo Reis

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP28/15

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Please Call Me John: Name Choice and the Assimilation of Immigrants in the United States, 1900-1930*

Pedro Carneiro
University College London,
Institute for Fiscal Studies and Centre for Microdata Methods and Practice

Sokbae Lee
Seoul National University,
Institute for Fiscal Studies and Centre for Microdata Methods and Practice

Hugo Reis
Bank of Portugal and Católica Lisbon School of Business and Economics

22 June 2015

Abstract

The vast majority of immigrants to the United States at the beginning of the 20th century adopted first names that were common among natives. The rate of adoption of an American name increases with time in the US, although most immigrants adopt an American name within the first year of arrival. Choice of an American first name was associated with a more successful assimilation, as measured by job occupation scores, marriage to a US native and take-up of US citizenship. We examine economic determinants of name choice, by studying the relationship between changes in the proportion of immigrants with an American first name and changes in the concentration of immigrants as well as changes in local labor market conditions, across different census years. We find that high concentrations of immigrants of a given nationality in a particular location discouraged members of that nationality from taking American names. Poor local labor market conditions for immigrants (and good local labor market conditions for natives) led to more frequent name changes among immigrants.

Key words: Americanization, culture, first name, identity, immigration
JEL Classification Codes: J15, N32

*We would like to thank Christian Dustmann and Imran Rasul for their helpful comments. We gratefully acknowledge financial support from the Economic and Social Research Council (grant reference RES-589-28-0002 through the Centre for Microdata Methods and Practice) and from the European Research Council (ERC-2009-StG-240910-ROMETA).

1 Introduction

Immigrant assimilation is often associated with cultural change. As emphasized by Lazear (1999), adoption of the native culture by immigrants facilitates trade with natives and is therefore more likely in settings where trading opportunities with natives are large and trading opportunities with immigrants are small.

In Lazear's (1999) empirical work, adopting the native culture is measured by the adoption of the native language. There are, however, several other dimensions of native culture that could potentially be taken up by immigrants. Watkins and London (1994) emphasize one in particular: the adoption of American first names. As in Lazear (1999), they relate name (cultural) change, with the degree of social interaction with natives. They state: "Although individual and personal, names take on their significance in social interaction. Since the context of social interaction changes with immigration, names can be expected to change as well." Adoption of an American name is a relatively easy task and, as documented by Watkins and London (1994), it was common among Italians and Jews in the 1910 US census. This was seen as a sign of rapid cultural change among immigrants arriving at the turn of the 20th century.

In this paper, we start by expanding the analysis in Watkins and London (1994) to most immigrants (not only Italians and Jews), in the 1900 to 1930 US censuses (not only in the 1910 census). We use a conservative definition of American names: they consist of the 100 most common names in the US Social Security records of the 1900s, plus names of American origin. With our definition of what is an American name, we show that at any given time between 1900 and 1930, around 77% of male immigrants in the US had an American first name. In contrast, American first names were very uncommon among immigrants at the time of their arrival (according to arrival records from Ellis Island). If we consider only the top three American names for each country, they are much more prevalent among immigrants than among natives. For example, around 40% of Czech male immigrants, about 30% of Portuguese male immigrants and around 20% of Italian male immigrants held a name from the respective top three American names. In contrast, only 15% of US natives held any of the three top names.

Having shown that the adoption of an American name was common among immigrants, and how it varied across country of origin, we proceed to document associations with other correlates of assimilation. Immigrants with American names are shown to work in occupations with better occupation scores, to be more likely to have a US born spouse and to be more likely to be a US citizen.

Finally, we study economic determinants of the choice of an American name. To

start with, as in Watkins and London (1994) and Lazear (1999), we examine the extent to which the degree of potential social interaction with other immigrants of the same country of origin affects the probability that an immigrant takes up an American name. We measure the degree of potential interaction using the concentration index developed by Lazear (1999), who argued that as this index increased, so did opportunities for trade in the immigrant community. The concentration index is 100 times the number of individuals aged 16 to 65 in a county who were born in the given immigrant's native country, divided by the total number of individuals residing in the county of residence of each immigrant.[1] However, we go much beyond Watkins and London's (1994) and Lazear's (1999) cross-sectional analysis, by relating changes in the concentration index within county over time and changes in the naming decisions of immigrants residing in that county. We find that, on average, a one standard deviation increase in the concentration index leads to about a 1.4 percentage point decline in the share of immigrants adopting American names.

In addition, a novelty of our paper is the study of the impact of changes in local labor market conditions for immigrants and natives, and their interaction with changes in the concentration index, on immigrants' decisions to adopt American names. Local labor market conditions for immigrants could affect the opportunities for trading both with other immigrants and with natives. We find that, on average, a one standard deviation increase in the occupation score of other immigrants living in the same location leads to about a 1.3 percentage point decline in the adoption of American names. This effect is weaker for immigrants in locations with a higher concentration index, indicating that labor market conditions and ethnic concentration are substitutes with regard to name change. We do not find strong impacts of the local unemployment rate of immigrants on adoption of American names.

There is a substantial literature in economics on international migration to the US at the turn of the 20th century (see, for example, Hatton and Williamson (1998), Abramitzky, Boustan and Eriksson (2012, 2013, 2014) or Bandiera, Rasul and Viarengo (2013), among others). The paper by Abramitzky, Boustan and Eriksson (2014) is particularly related to our work given the authors' careful treatment of assimilation patterns for immigrants arriving in this era. Contrary to previous studies, they find that patterns of occupational convergence over time between immigrants and natives were limited to just a few nationalities, and that the typical immigrant experienced

[1]Watkins and London (1994) use instead very indirect measures of potential social interactions, such as: duration in the US (since longer duration is likely to mean stronger interaction with natives), arrival in the US before age 14 (since much socialization with natives is likely to begin at school) or ability to speak English.

neither convergence nor divergence relative to natives. Our paper is also closely related to Biavaschi, Giulietti and Siddique (2013), who focus on the impact of the Americanization of names on the labor market outcomes of migrants and found that migrants who Americanized their names experienced larger occupational upgrading. The main difference between our paper and Biavaschi, Giulietti and Siddique (2013) is that we focus primarily on the determinants of name choice, while their main interest is the impact of the Americanized name (about which we have little to say beyond some interesting and strong correlations).

Regarding name changes in broader contexts, this paper is related to Arai and Thoursie (2009) and Rubinstein and Brenner (2014). Arai and Thoursie (2009) studied the effects of surname change to Swedish-sounding or neutral names for immigrants from Asian, African and Slavic countries. They found an increase in annual earnings after a name change and argued that those changes are a response to discrimination. Rubinstein and Brenner (2014) used sorting into inter-ethnic marriage and differences between Israeli ethnic surnames to study ethnic discrimination in labor markets. Both papers relate closely to the literature looking to racial discrimination and black names (see, for example, Fryer and Levitt (2004), Bertrand and Mullainathan (2004) and Cook, Logan and Parman (2014), among others).[2] See also Olivetti and Paserman (2013) and Güell, Mora and Telmer (2015) for the use of informational contents of names to study intergenerational mobility.

Our paper is also related to the economics literature on identity. The simple model we consider follows Lazear (1999) in focusing primarily on the role of market interactions, but we could have written instead an economic model of immigrant identity. Identity may influence preferences (and behaviors and outcomes), and the fact that it operates intrinsically through groups leads to complex group dynamics and equilibria. For example, see Akerlof and Kranton (2000) and Benabou and Tirole (2006) for theoretical work, as well as Casey and Dustmann (2010) and Manning and Roy (2010) for empirical work.

The remainder of this paper is organized as follows. Section 2 describes the data and documents the extent to which immigrants at the turn of the 20th century adopted American names (for them and their children). Section 3 proceeds to show to what extent the adoption of an American name is associated with better assimilation, as measured by labor market and social outcomes. Section 4 presents a simple economic

[2]In experimental economics, Charness and Gneezy (2008) studied behavior in dictator and ultimatum games by comparing outcomes between the standard case of anonymity and the case when information on the last name of the participant is revealed. They presumed that knowing the last name of the counterpart in experiments would reduce the social distance between participants.

model of name choice, as a function of labor market and network variables. Section 5 provides estimates of this simple model and Section 6 concludes. Online appendices provide a detailed description of the classification of the name types and additional empirical results that are omitted from the main text.

2 Data

We use data from the Integrated Public Use Microdata Samples (IPUMS) of the US Decennial Census between 1900 and 1930 (Ruggles et al., 2010).[3] Up to the 1930 census, data from the IPUMS records the first name for most individuals (as well as the country of birth of each individual), which allows us to determine the type of name used by each immigrant.

The 1900 census consists of a 5% national random sample of the population. From 1910 onwards, the census data consists of a 1% national random sample of the population. In this paper, all reported summary statistics and estimation results are weighted by the sample size in each census (which means that observations from the 1900 census have a weight of 0.2 relative to observations in the remaining censuses).

2.1 First Names of Immigrants

A person is classified as an immigrant if he or she was born in a foreign country. All other individuals are classified as natives.

Our sample[4] includes immigrants, originating from 16 different countries of birth: Germany, Italy, former USSR,[5] Poland, Sweden, Mexico, Norway, Hungary, Czechoslovakia, Denmark, Greece, France, Japan, China, Portugal, and Spain.[6]

Table 1 shows the top 10 names for different countries of immigrants' origin. The most popular name in the US, John, is also the most common name among immigrants from Germany, the former USSR, Poland, Sweden, Norway, Hungary, Czechoslovakia and France (half of the countries included in our sample). Moreover, with the exception of a few countries such as Portugal, Spain, Mexico and China, the most common

[3]For further details, see IPUMS website (`http://usa.ipums.org/usa/`).

[4]In the appendix, we provide more information regarding male immigrants. Table B.1 in Online Appendix B presents the distribution by country of origin of male immigrants, for the sample used in the paper.

[5]The census for these years groups all individuals from the former USSR in a single category. Therefore, we use former USSR to define immigrants from a large set of countries because this is the definition available in the IPUMS data.

[6]Both British and Irish immigrants are excluded from our sample since their first names are much closer to American names than those from other countries.

name across nationalities is always of American/British origin (John, Joseph, Peter and George).

It is remarkable how prevalent American names are among immigrants of different nationalities. For example, among the Italians, 8.6% are called Joseph, 6.7% are Frank and 6.4% are John. Among the Portuguese, 12.4% are called John, 12.0% are Joseph and 7.1% are Frank. Interestingly, there are only two American names (Joe and John) among the top 10 names for Mexicans, suggesting that they had little need for a name change. Names such as Joseph, John and Frank are completely non-existent in Italian or Portuguese cultures. These are very distinctively American names.

The naming patterns among female immigrants are similar to those for males. Table 2 shows the top 10 names for different countries of female immigrants' origin. With the exception of a few countries such as Japan, Mexico and China, the most popular name across nationalities is either Mary or Anna.

In what follows, we distinguish the immigrants by type of their first names: American and non-American first names. We use a clear and objective procedure to classify names, which is described in detail in Online Appendix A. This procedure essentially involves comparing names that are common in the US population (from Social Security records) with names that are distinctively from each country and not likely to be American names (which are available from country-specific name databases). To give some examples, for German immigrants, we classify John, Frank and Steven as American names and Otto and Claus as non-American. For Italians, George, Leo and Vincent are classified as American and Antonio and Domenico as non-American. For immigrants from the former USSR, William, Robert and Simon are classified as American and Ivan and Vladimir as non-American.

Table 3 presents for both, males and females, the percentage of American names according to our definition by year and by immigrants' birthplace, respectively. Starting with males, in the pooled sample, the percentage of immigrants with an American name is about 77%, being relatively constant between 1900 and 1930. Nevertheless, there is significant heterogeneity across countries, with this percentage ranging from around 8% for those from Japan, to around 93% for those from the former USSR. For female immigrants, the overall percentage of American names is about 76.9%, which is very similar to male immigrants' percentage (76.5%). The heterogeneity across different origins is also similar between males and females.

Figure 1 displays the proportion of individuals from different countries (in the 1900-1930 US censuses) who have one of the top three American names among immigrants from that country. Notice that this proportion is larger for immigrants from several

countries than it is for US natives, indicating that immigrants tend to have more stereotypical American names than natives. The percentage of top three American names ranges from around 2.5% for Japan to more than 40% for Czechoslovakia.

Although we do not have information on the timing of name change among immigrants, we can see how the proportion of immigrants with American names changes with the amount of time elapsed since the immigrant's arrival to the US. In order to do this, we would like to know the prevalence of American names among immigrants at the time of their arrival, which is not available in the census data. Therefore, we combined data from the US National Archives and Records Administration (NARA) and the Public Use 5% Sample of the 1900 census.

Starting with immigrants from Italy (the second-largest origin group of immigrants in our sample), the specific dataset we use is called "Italians to America Passenger Data File" from the NARA and contains records of 845,368 Italian passengers who arrived in the US between 1855 and 1900, with information on their last and first names, age and gender. The 1900 census sample provides respondents' first and last names, birthplace, years in the US and gender, among other variables. Combining the two data sources, we can construct Figure 2, which shows the percentage of male Italian immigrants with American first names by number of years in the US.

There is clear evidence that male Italian immigrants changed first names soon after their arrival. Using passport information at the US ports for the period 1855-1900, Figure 2 shows only 0.9% of Italian males had an American name at the time of arrival in the US. However, in the 1900 census data, this figure rises to around 50% for those in the US for less than one year, and then reaches about 70% for those in the US for more than 20 years.

Figure 2 shows similar statistics for immigrants from Germany and the former USSR (the two other most represented origin countries of immigrants in the sample).[7] Using passport information at the US ports, around 40% of the German males had an American name upon arrival. This number reaches around 70% for those in the US for less than one year and around 80% for those in the US for more than 20 years. For those from the former USSR, the figures are much higher. About 70% of them had an American name upon arrival, and this proportion rises to around 90% over time. These figures show significant heterogeneity among these three countries of origin.

[7]The dataset used is called "Germans to America Passenger Data File" and "Former USSR to America Passenger Data File."

2.2 Other Characteristics of Immigrants

The information collected in the 1900-1930 censuses includes migration variables, such as country of birth, years in the US and age upon arrival, along with the usual demographic variables, some education variables such as literacy and ability to speak English, and work and occupation variables.

Table B.2 in Online Appendix B gives a brief description and relevant definitions of the data used in the analysis. In particular, we focus our attention on economic and network variables, which can be constructed for each geographical and census year.[8] Throughout the paper, we use county as the geographical unit of interest, since it allows for substantial regional variation, it is likely to constitute a labor market of interest and it is large enough for us to be able to construct reliable economic and network variables with the available data.[9] All the economic and network variables have a -1 subscript, indicating that they use county-level data from the previous census.[10] For example, for a particular immigrant in 1920, $Immigr.\ unemp.\ rate_{-1}$ denotes what the unemployment rate for immigrants was in 1910 in the individual's current county of residence.

Table 4 compares the values of these variables for immigrants who have an American name and for those who do not. Starting with male immigrants, our data shows that immigrants with an American name tend to have been in the US for a longer period and to be younger (in both cases, there is a difference of around half a year) and to live in more populated places. In terms of economic variables, those with American names tend to live in counties where the immigrants' unemployment rate is higher and in counties where both natives and immigrants perform better in terms of their occupational score. In terms of network variables, there is a clear difference between the two groups. Immigrants with an American name tend to live in communities with a lower concentration index - i.e., in communities in which a smaller percentage of residents are from their native country - than those who did not adopt an American name. In particular, the average immigrant with an American name lives in a county in which 7% of the residents were born in their native country. This figure reaches almost 10% for those who kept their non-American name. Furthermore, immigrants with an

[8] All variables are available for each census, with the exception of the unemployment rates for the years 1920 and 1930.

[9] COUNTY identifies the county where the household was enumerated, using the Inter-University Consortium for Political and Social Research (ICPSR) coding scheme. An example of a county is Los Angeles, which belongs to the State of California. County is available for all 1850-1930 samples. For further details, see IPUMS website (`http://usa.ipums.org/usa/`).

[10] The reason for focusing on lagged values instead of contemporaneous values (which could easily be constructed) of these variables will become obvious below.

American name tend to live in places where the literacy level is relatively higher and where the percentage of immigrants from the same country with an American name is higher, compared with immigrants who have a non-American name.

Some of these correlations are interesting and suggest that at least a few of these variables may be strongly associated with the choice of an American name, such as a high local unemployment rate of immigrants or a low local concentration index, both of which may increase the incentives for assimilation. Below, we investigate these hypotheses more rigorously. Table 4 presents also the summary statistics for female immigrants and compares the values for those who have an American Name and for those who do not. In general, the results for females are in line with those for the male immigrants.

2.3 Measures of Immigrants Assimilation

We look at several economic and social-cultural outcomes of immigrants: whether the immigrant was full-year employed last year; log of occupational score; whether the immigrant married a US-born spouse (excluding second-generation immigrant); whether the immigrant speaks English; and whether the immigrant is a US citizen. Variable descriptions are presented in Table B.2 in Online Appendix B.

Table 4 shows values of these variables for male immigrants with and without an American name, for the overall sample. Relative to those without American names, male immigrants with an American name are more likely to: i) have a higher (log of) occupational score; ii) speak English; and iii) become a US citizen.[11]

Table 4 shows also the corresponding results for female immigrants for the whole sample. Since the labor force participation is low among females between 1900 and 1930, we also report the percentage who were in the labor force. As in the case of males, relative to those without American names, female immigrants with an American name are more likely to: i) have a higher (log of) occupational score; ii) be in the labor force; iii) marry a man born in the US; iv) speak English; and v) become a US citizen. The female immigrants with an American name are not more likely to be employed for the full year than those without American names.

[11]Looking at the three most represented groups of immigrants by birthplace, the results show significant heterogeneity. Table B.5 in Online Appendix B shows that for German immigrants, there is hardly any difference between the economic outcomes of those with and without American names. This is not true when we look to the other two groups (immigrants from Italy and former USSR), for whom adopting an American name is associated with better assimilation as measured by the variables in Table B.5.

2.4 American Names among Second-Generation Immigrants

It is also interesting to examine naming patterns for second-generation immigrants. In our sample, about 85% of all boys and 88% of all girls born to immigrants have an American name.[12]

Table 5 relates to naming patterns for fathers and sons and, fathers and daughters. Starting with the whole sample, 86% of boys born of immigrant males with American names also have American names. When we look at children of immigrants with non-American names, the proportion with American names is lower, but it is still substantial, at 78%. When we look across the three main nationalities - Italians, Germans and those from the former USSR - the patterns are quite similar. In general, Table 5 shows similar patterns regarding the relationship between name choices of fathers and daughters. It is remarkable that a very high proportion of second-generation immigrant children have American names, regardless of whether the father did or did not adopt an American name (although the probability is a little higher for fathers who adopted an American name).

3 American Names and Measures of Immigrant Assimilation

There is clear evidence that immigrants, especially Italians, changed their first names immediately after arrival, a clear sign that the Americanization of one's first name could be important. This could be because an American name provides a change in one's social identity, making integration easier. It shows a clear intention to quickly assimilate on the part of immigrants. However, adopting names that are common in the dominant culture may not necessarily imply a change in one's outcomes.

In this section, we examine to what extent the adoption of an American name is associated with different labor market and other social outcomes. The goal of this section is not to argue that a change in one's name necessarily leads to a change in outcomes, but simply to document an association between name change and a variety of variables, which could all be caused by one's underlying desire (and ability) to assimilate. This will tell us the extent to which assimilation through the adoption of an American name is associated with assimilation as measured by several other

[12]Table B.6 in Online Appendix B presents the summary statistics for the children sample (sons and daughters).

variables.[13]

To examine whether the associations between American name and assimilation measures reported in Table 4 survive the inclusion of controls, we estimate the following specification:

$$Y_{ict} = \alpha AmericanName_{ict} + \beta' X_{ict} + \gamma' \phi_{ct} + \psi_t + \epsilon_{ict} \quad (1)$$

where Y_{ict} is the outcome variable for immigrant i living in county c in census year t, X_{ict} is a vector of individual characteristics, ϕ_{ct} is a vector of county characteristics (which includes county fixed effects in one of the specifications), ψ_t are year effects and ϵ_{ict} is an i.i.d. (independent and identically distributed) error term. $AmericanName_{ict}$ is an indicator that takes value 1 if the immigrant has an American name. We report estimates of α, the conditional association between adopting an American name and the outcomes we consider.

We show the main results for male immigrants in Table 6 for all outcomes including as controls the census year, immigrant birthplace, cohort of entry, number of years in the US, age and log of population in the county of residence, and county fixed effects.[14] Our specification is demeaning in that we report estimates of α using only within-county changes over time in the proportion of immigrants with American names and changes in their different outcomes. For male immigrants, we present results not only for the whole sample, but also for the subsamples of immigrants from Germany, Italy and the former USSR. For female immigrants, we present results only for the whole sample, for brevity, in Table 7 but with different specifications of control variables.

[13] In a recent paper, Biavaschi, Giulietti and Siddique (2013) study the economic payoff to name Americanization, by constructing a panel of immigrants observed before and after changing their name, and instrumenting the decision to change name with the degree of complexity of the original name, measured by Scrabble points. Their assumption is that the complexity of one's name relative to the American norm affects the costs but not the benefits of name Americanization (which could potentially be violated if those with the larger Scrabble points discrepancy had different wages in the absence of name change, and different returns to name change, from those with lower Scrabble points discrepancy). Using data from 1930, they find that name Americanization leads to significant increases in occupation scores.

[14] In Online Appendix B (see Tables B.7 to B.11) we present the results in different tables, one for each outcome, with different specifications. Each table has three columns. Column (1) includes as controls only the census year. Column (2) controls also for immigrant birthplace, cohort of entry, number of years in the US, age and log of population in the county of residence. Finally, column (3) includes county fixed effects, and corresponds to our specification reported in Table 6.

3.1 Male Immigrants: Economic Assimilation

Columns (1) and (2) in Table 6 display the results for two economic assimilation measures: log occupational score and full-year employment last year, respectively. The estimates suggest that adopting an American name has a positive and significant association with the log of the occupational score, especially if the immigrant is from the former USSR or Italy. Nevertheless, the association is relatively modest. More specifically, a former-USSR immigrant with an American name would have a log occupational score 0.027 points higher (which represents only 7% of one standard deviation) than one without an American name. For Italian immigrants, this difference is even smaller, at around 0.008 (which represents only 3% of one standard deviation).

Our estimates also suggest that adopting an American name may be negatively associated with full-year employment for Germans. Again, qualitatively speaking, the association is modest.

3.2 Male Immigrants: Cultural Assimilation

Column (3) shows that having an American name is positively associated with marriage with an American native woman. The point estimate again is apparently small, and it is only statistically significant for Italians. Notice, however, that the baseline probability of marrying an American is low to start with. For example, on average, only 1.3% of Italian immigrant males with non-American names ever marry an American wife, but this proportion is 1.3 percentage points higher for those who have an American name (an increase of 100% in this probability).

Column (4) indicates that adopting an American name has a positive association with the ability to speak English, in particular if the immigrant is Italian or from the former USSR. Focusing on column (3), having an American name increases the likelihood of an Italian and a former-USSR immigrant speaking English by around 3 percentage points (from averages of 71.1% and 75.2%, respectively).

3.3 Male Immigrants: Civic Assimilation

In terms of civic assimilation, measured by the acquisition of US citizenship, Column (5) shows that having an American name is associated with an increase in the probability of becoming a US citizen by 1 percentage point (from an average of 44.7%). This association is stronger among immigrants from the former USSR, for whom having an American name increases the likelihood of becoming a US citizen by almost 3

percentage points (from an average of 39.6%). For German and Italian immigrants, this association is not robust to the inclusion of county fixed effects.

3.4 Female Immigrants

Table 7 reports the regression results for female immigrants. Once the county fixed effect is added (column (3), which is our preferred specification), adopting an American name is significantly associated with labor force participation, having an American husband and speaking English. The most notable finding is that, as in the case of males, having an American name is positively associated with marriage with an American. The point estimate of 2.3 percentage points for females is almost 4 times the corresponding figure for males (0.6 percentage points). The baseline probability of marrying an American man is 10.9%, resulting in an increase of around 20% in this probability.

4 A Model of First-Name Choice

To motivate the econometric model of our next section, we build on Lazear (1999), who developed a simple model of culture and language. His model is based on the presumption that a common culture and a common language facilitate trade between individuals. American first names can be seen as one component of US culture. When traders negotiate a contract or more, they generally engage in both market and non-market interactions, and their first names will become known to each other. Sharing a common culture through first names could enhance trust between individuals.[15]

To describe this more formally, suppose that there are two types of first names: American and non-American first names, labeled F_A and F_N. Individuals can belong to either one of two cultures in the US: an American culture, labeled A, and a non-American culture, labeled N. For simplicity, assume that individuals of the F_A type belong to culture A and those of the F_N type belong to culture N. An individual can change culture by changing his or her first name. Define p_N as the proportion of individuals who belong to the non-American culture in equilibrium.

We consider the decision problem of an immigrant who is endowed with an F_N-type name and is considering whether or not to adopt an F_A-type name. Trades can

[15]One alternative would be to rely more heavily on the identity aspect of the name, which is only implicit here, and develop a model as in Akerlof and Kranton (2000). Lazear's (1999) model is, however, more suited to our application.

occur between individuals regardless of their cultures (or first names), but there are different probabilities that trade occurs within and across cultures.

Let t_i be the cost of adopting an American first name for individual i with a foreign first name.[16] We assume that this cost t_i depends on two components: a taste term, say ε_i, which varies across individuals, and the proportion of immigrants (living in the area), say q_N. Hence, $t_i = g(\varepsilon_i, q_N)$ for some function g. We expect the partial derivative of g with respect to the second argument to be positive, since it is plausible to assume that it is more costly to adopt an American name if one is surrounded by a high number of individuals of the N culture, because of social interactions or peer pressure (or group identity type reasons, as in Akerlof and Kranton (2000)). For simplicity, q_N is taken as given and cannot be changed. An alternative and slightly different model would set $q_N = p_N$ and solve for it in equilibrium.

There exist gains associated with the adoption of an American first name. As in Lazear (1999), we assume that the net gain, say b_i, associated with the adoption of an American first name depends on three factors: the proportion of those with the American culture $(1 - p_N)$ and the level of economic well-being for individuals of American and non-American cultures, say e_A and e_N, respectively. Hence, $b_i = h(1 - p_N, e_A, e_N)$ for some function h. We expect the derivatives of the function with respect to the first and second arguments to be positive, whereas the derivative with respect to the third argument should be negative.[17]

One simple way to motivate these assumptions on h would be the following. Suppose an individual is only able to trade with other individuals from his or her own culture. In addition, assume that individuals meet at random, and the probability of meeting someone from the N culture is p_N. When a meeting takes place, the value of a trade with someone from the N culture is e_N, while the value of trading with someone from the A culture is e_A (the value of the trade increases with the economic well-being of the trading partner). Then the expected value of income for someone of the N culture is equal to $p_N e_N$ (the probability of finding someone from the same culture times the value of a trade with that person) and the expected value of someone of the A culture is equal to $(1 - p_N) e_A$. In this particular case, $h(1 - p_N, e_A, e_N) = (1 - p_N) e_A - p_N e_N$.[18]

[16] In principle, we could also have immigrants for whom their original given name is already an American name. For these individuals, there is no name switching involved.

[17] Lazear (1999)'s model abstracts from e_A and e_N, but we model them explicitly here to make the model consistent with our empirical work.

[18] Therefore $\frac{\partial h}{\partial p_N} = -(e_A + e_N) < 0$, $\frac{\partial h}{\partial e_A} = (1 - p_N) > 0$, $\frac{\partial h}{\partial e_N} = -p_N < 0$, $\frac{\partial^2 h}{\partial e_A \partial p_N} = -1 < 0$ and $\frac{\partial^2 h}{\partial e_N \partial p_N} = -1 < 0$.

We now describe an immigrant's name choice decision. An immigrant acquires an American first name if and only if

$$t_i < b_i, \text{ or equivalently } g(\varepsilon_i, q_N) < h(1 - p_N, e_A, e_N), \tag{2}$$

which is similar to equation (2) in Lazear (1999).

In order to simplify estimation of the model in (2), assume that $g(\varepsilon_i, q_N) = g_1(\varepsilon_i + g_2(q_N))$, where $g_1 : \mathbb{R} \mapsto \mathbb{R}$ is a strictly increasing function and $g_2 : \mathbb{R} \mapsto \mathbb{R}$ is a flexible function of only q_N. Also, assume that ε_i is independent of (p_N, q_N, e_A, e_N). Then it follows from (2) that the proportion of individuals with American first names is

$$\Pr(t_i < b_i) = G\left[g_1^{-1}\left\{h(1 - p_N, e_A, e_N)\right\} - g_2(q_N)\right], \tag{3}$$

where $G : \mathbb{R} \mapsto \mathbb{R}$ is the cumulative distribution function (CDF) of ε_i. The binary choice model in (3) is the basis of our empirical work.

Although the model in equation (3) is quite simple, it implies some restrictions on the specification of our econometric model. One important prediction is that, under reasonable assumptions,[19] $\Pr(t_i < b_i)$ is decreasing in both p_N and q_N, decreasing in e_N and increasing in e_A. However, verifying these predictions is empirically challenging. The main difficulty is that $1 - p_N$ is the same as the proportion of American first names, i.e., $\Pr(t_i < b_i) = 1 - p_N$, which is determined in equilibrium. In other words, while one could try to argue that (q_N, e_A, e_N) are exogenous variables, p_N is clearly endogenous and determined in equilibrium. In the absence of a convincing strategy to identify this structural model, we choose instead to focus on a reduced-form model (after solving equation (3) for p_N), examining the impact of the exogenous variables (q_N, e_A, e_N) on p_N. In particular, we estimate

$$\Pr(t_i < b_i) = \Psi\left(q_N, e_A, e_N\right), \tag{4}$$

where Ψ is a reduced-form function, and we take $\Psi(\cdot)$ to be the probit model for convenience in Section 5.

Nevertheless, even if we limit ourselves to this more limited objective, it is still difficult to argue that (q_N, e_A, e_N) really are exogenous variables. Once we consider our empirical setting, with multiple locations and multiple nationalities, it is natural to think of local unobservable variables that could simultaneously affect (q_N, e_A, e_N) and p_N. For example, a high degree of industrialization of a county could attract

[19] Specifically, we assume that $\frac{\partial h}{\partial p_N} < 0$, $\frac{\partial h}{\partial e_A} > 0$, $\frac{\partial h}{\partial e_N} < 0$ and $\frac{\partial g_2}{\partial q_N} > 0$.

many immigrants, have an impact on immigrants' and natives' wealth, and make it attractive for an immigrant to acquire an American name.

In order to minimize these problems, we proceed as follows. First, instead of using contemporaneous county measures (q_N, e_A, e_N) on the right-hand side of equation (4), we construct lagged values of these variables for each county using the previous census. This means that (q_N, e_A, e_N) are measured with a 10-year lag and correspond respectively to the concentration of immigrants in a particular county 10 years ago, the average economic well-being of natives in the county 10 years ago and the average economic well-being of immigrants in the county 10 years ago. Second, instead of relying purely on cross-sectional variation to estimate equation (4), we use the census years available to us to construct a three- (and four-) period panel of counties (depending on the specification), allowing us to include both time and county indicators in the model, which control for a time trend in the adoption of an American name by immigrants and for county time-invariant unobservables.

There exist other explanatory variables, say x_i, which we control for when implementing equation (4) in the next section and which we specify below. Since there is no obvious functional form choice for modelling x_i and (q_N, e_A, e_N) together, we consider several specifications (and sometimes include interaction terms between different variables).

5 Determinants of American First Names

We use a probit model for first-name choice. To start with, we focus on male immigrants. Our estimation sample includes 174509 male immigrants from 16 different countries, between 16 and 65 years of age, in the 1900, 1910 and 1920 censuses.[20]

Table 8 presents estimates of equation (4). This and subsequent tables regarding the determinants of name choice show the marginal effects of explanatory variables, multiplied by 100 (hence, in terms of percentage points). Our basic specification is

$$\begin{aligned}&\Pr\left(AmericanName_{ibect}=1\right)\\&=\Phi\left(\gamma q_{Nbct-1}+e_{Act-1}\delta_A+e_{Nct-1}\delta_N+Z_{bct-1}\rho+X_{ibect}\beta+\alpha_b+\phi_c+\delta_e+\psi_t\right). \quad (5)\end{aligned}$$

Here, the dependent variable (also defined above), $AmericanName_{ibect}$, is an indicator

[20]In our basic specification, we did not use observations from the 1930 census since in the 1920 and 1930 censuses there is no information regarding the unemployment status for each individual, thereby preventing us from creating local unemployment rates in lagged levels. To construct lagged variables for the 1900 census, we used the 1880 census (the previous census available before 1900).

variable that has value 1 if the first name of individual i, born in country b, entering the US in year e and living in county c at time t is an American name. To proxy q_{Nbct-1}, we use the concentration index ("Concentration index$_{-1}$" in the tables), also used in Lazear (1999), which is specific to each birth nationality (b) and county (c), and which is measured at the time of the previous census ($t-1$). The variables e_{Act-1} and e_{Nct-1} are vectors which include, for each county and year, lagged values of unemployment rates for natives and immigrants ("Immigr. unemp. rate$_{-1}$" and "Native unemp. rate$_{-1}$") and average log occupational scores for natives and immigrants ("Immigr. log occ. score$_{-1}$" and "Native log occ. score$_{-1}$").[21] Z_{bct-1} includes other lagged county variables: namely, the logarithm of the number of male immigrants between 16 and 65 years old ("Log immigr. population$_{-1}$"), the percentage of male immigrants between 16 and 65 years old from the same country being literate ("Own immigr. literacy percentage$_{-1}$") and the percentage of male immigrants between 16 and 65 years old from the same country with an American name ("Own immigr. American name percentage$_{-1}$"). These variables are included in the model to capture the impact of the characteristics of the local network of immigrants, so they complement the concentration index just described. The vector X_{ibect} includes number of years in the US and its squared term, age in years and its squared term, and the log of the number of respondents in the sample for each geographical unit ("Log population"). The fixed effects α_b, ϕ_c, δ_e and ψ_t are birthplace, county, cohort of entry and year dummies, respectively. All standard errors are clustered at the county level in order to capture cross-sectional and time-series dependence in county-level shocks.

We also present results from more flexible models with interactions between some of these variables. In one model, we interact the number of years an individual has spent in the US (one of the variables in X_{ibect}) with the variables in $(e_{Act-1}, e_{Nct-1}, Z_{bct-1})$. In another model, we interact q_{Nbct-1} with these same variables.

5.1 Main Results - Male Immigrants

Column (1) of Table 8 presents estimates from the basic specification of equation (5). This column shows that years in the US, age, lagged immigrants' log of occupational score and lagged concentration index are the most important characteristics that determine the adoption of an American name. We fail to reject the joint hypothesis

[21] For robustness purposes, we also consider different specifications including "Own immigr. unemp. rate$_{-1}$", "Exc. own immigr. unemp. rate$_{-1}$", "Own immigr. log occ. score$_{-1}$" and "Exc. own immigr. log occ. score$_{-1}$", which are variables similar to the economic variables just listed for immigrants, but which are specific to each country of birth.

that the coefficients on all economic variables ("Immigr. unemp. rate$_{-1}$", "Native unemp. rate$_{-1}$", "Immigr. log occ. score$_{-1}$", "Native log occ. score$_{-1}$") are equal to zero. However, we reject the hypothesis that the coefficients on all local network variables ("Concentration index$_{-1}$", "Log immigr. population$_{-1}$", "Own immigr. literacy percentage$_{-1}$", "Own immigr. American name percentage$_{-1}$") are jointly equal to zero.

Our results suggest that increases in the concentration index are likely to reduce the probability of an immigrant adopting an American name. This would be a natural prediction of our model if an increase in q_N increased the costs of changing one's name, because of social pressure. It is also plausible that, in a slightly richer model than the one specified above, q_N would affect trading probability along with p_N. If, along with an increase in p_N, an increase in q_N led to an increase in the probability of trading with individuals of one's ethnicity, then it would reduce the net value of changing one's name (b). This would be an additional channel through which an increase in q_N could have a negative impact on the probability of adopting an American name.

We also see that an increase in the occupation score of immigrants, which is a proxy for their economic well-being, results in a decline in the probability of having an American name. Again, this is a natural prediction of the model presented above: in settings where the average economic well-being of immigrants is high, it is profitable to trade within this group, which means that the adoption of an American name may not be especially valuable. There are no robust effects on name adoption of the three other economic variables we include in the regression. With regards to individual regressors, the adoption of an American name increases with years in the US and decreases with age.

In terms of magnitudes, our estimates indicate that the likelihood of having an American name increases by 2.3 percentage points (p.p.) for a one standard deviation (s.d.) increase in the number of years in the US and decreases by 1.7 p.p. for a one s.d. increase in immigrants' age. The same probability decreases by 1.3 p.p. and 1.4 p.p. for a one s.d. increase in lagged immigrants' log of occupational score and lagged concentration index, respectively.

As mentioned above, we also considered richer specifications of equation (5), where we allowed some variables to interact with each other. We consider two types of interactions which are of particular interest. First, we interacted the four economic variables and the four network variables in the model with years in the US. Our idea is that the longer one is in the US, the more assimilated one is likely to be, and that should influence how likely these variables are to affect American name adoption

(our original hypothesis was that it should dampen their effects on American name adoption). Second, we interact these same economic and network variables with the concentration index, which, as stated above, is specific to each county and country of birth. Again, our question is whether a higher concentration of immigrants sharing the same place of origin is a substitute for or complementary to other economic and network variables in the regression, with regards to their impact on name change. When we look at the impacts of network and economic variables at the average, the main results (in the simpler specification in column (1) of Table 8) are robust to the different specifications presented in columns (2) and (3) of the same table, where all economic and network variables were interacted with years in the US and lagged concentration index, respectively.[22]

Table 9 shows the impact of different economic and network variables on the choice of American name, at different percentiles of the distribution of years in the US. The results suggest that the longer one is in the US, the stronger seems to be the impact of network (concentration index) variables on American name adoption, while there is not much of a pattern with regards to economic variables.

It is interesting that the impact of the concentration index increases with years in the US, contrary to our original hypothesis. One reason may be that, if individuals are heterogeneous, the characteristics of immigrants who are at the margin between adopting and not adopting an American name may change with years in the US. In particular, individuals who resist changing their name after many years in the US could be those with particularly high tastes for their native name and their native culture. If that is the case, it is plausible that they will also be more responsive to changes in the concentration index than the average individual. These individuals are more likely to be the ones at the margin for high values of years in the US, which means that the coefficient in the interaction between each economic and network variable and years in the US is capturing not only the substitutability between these two sets of variables with years in the US (with regards to name change), but also the change in the composition of immigrants who retain their native name.

Table 10 shows, for different percentiles of the lagged concentration index, the impact of all economic and the remaining network variables on the choice of American name. In this case, a higher concentration of immigrants from the same origin is complementary to the percentage of immigrants with the same origin who have an American name, regarding their impact on name adoption (see the last row of the

[22]When we include these interactions, we report the marginal effects for the variables of interest, evaluated at the average values of the other variables.

table). This means that when there are many immigrants of an individual's nationality in a given location, the individual will be particularly sensitive to name choice, especially if they pressure the individual to adopt the American culture, or if they only trade with him if he adopts that culture (a model where p_N and q_N are essentially the same). This is a plausible story, although it would be possible for the concentration of immigrants from the same origin and the percentage of immigrants with the same origin who have an American name to be substitutes instead, since a high concentration index reduces the need to trade with natives and, therefore, could reduce the payoff to a name change.

At the same time, the concentration index appears to be a substitute for immigrants' and natives' economic performance. It is possible that a high concentration index results in a high level of pressure to keep the original culture, and therefore reduces the importance of economic incentives for name change, which would explain our results. But we could also have found that an increase in the probability of encountering other similar immigrants increased the probability of trade with them, which would mean that the economic conditions of immigrants would be especially important in those circumstances, which would explain an exactly opposite result. In addition, if there are changes in the composition of immigrants at the margin when the concentration index increases, along the lines of the argument made above for years in the US, then the potential stories we have just described would have to be modified.

5.2 Robustness Checks - Male Immigrants

In this subsection, we examine the sensitivity of the main results presented in Section 5.1. First, we exclude from the sample the three most represented groups of immigrants by birthplace separately. Second, we use different specifications regarding economic variables. Finally, we add an additional year of data. Overall, this subsection suggests that the results presented above are quite robust to different samples and specification choices.[23]

Table 11 shows the results for the baseline specification when we exclude the three most represented groups of immigrants in the sample (one at a time), as defined by their birthplace. In general, our main conclusions are qualitatively unchanged, but at the same time the results show significant heterogeneity among these three groups.

As before, the most important determinants of American name are years in the US, age, lagged log of occupational score and lagged concentration index. When

[23]In general, our main results are also robust to a specification where we use current network and economic variables instead of the lagged variables (results available on request).

excluding Germans and, to a lesser extent, when excluding immigrants from the former USSR, the impact of the economic and network variables becomes greater than in the baseline specification, in particular the impact of the lagged concentration index and the lagged log of occupational score. Interestingly, when excluding Germans, the economic variables as a whole are statistically significant. Furthermore, the economic effect becomes smaller and not significant when we exclude Italians from the sample, while the impact of the lagged concentration index remains relatively the same.

Table 12 changes the economic variables included in the regression models. In particular, we construct two new sets of economic variables where: 1) we use as the reference group only those immigrants with the same birthplace; and 2) we use as the reference group all immigrants other than those with the same birthplace. In column (1), we add to the baseline specification the variables "Own immigr. unemp. rate$_{-1}$" and "Own immigr. log occ. score$_{-1}$"; in column (2), we use only the "own" immigrant economic variables; and in the last column, we have a specification with "own" and "excluding own" immigrant economic variables. The impact of the lagged concentration index remains relatively stable across the different specifications. Regarding the occupational score variables, the impact of the lagged immigrant occupational score becomes smaller and not significant when we add "own immigrant" variables (column (1)), while in the other two specifications the different occupational score variables show up as statistically significant.

Our main results suggest that immigrants' log of occupational score is the main relevant economic variable, while unemployment variables are not statistically important. This allows us to perform an additional robustness exercise which excludes the unemployment variables from our specification. In doing so, we are able to add one year to our data, 1930, and examine whether the results are robust to this new sample and specification. Table 13 presents the results of this new specification. In columns (1) and (2), we use only lagged log of occupational score as economic variable for the 1900-1920 and 1900-1930 samples, respectively. In both specifications, the main conclusions remain the same: lagged immigrants' log of occupational score and lagged concentration index are still the only significant economic and network variables, respectively. Nevertheless, for the 1900-1930 sample, the results are not as strong as in the baseline specification.

5.3 Determinants of American Name Choice for Female Immigrants

In this subsection, we report results for female immigrants. Table 14 shows estimates with specifications similar to those in Table 8. The only difference is that labor force participation rates are included instead of unemployment rates. As shown in Section 3.4, in the case of females, labor force participation is more related to Americanized names than unemployment status is. Across all three columns in Table 14, we find that only age and lagged natives' labor force participation rate are statistically significant determinants of the adoption of an American name for females. In terms of magnitudes, our estimates in column (1) indicate that the likelihood of having an American name increases by 2.2 percentage points for a one standard deviation increase in the lagged natives' labor force participation rate and decreases by 3.9 p.p. for a one s.d. increase in immigrants' age. When we use unemployment rates instead of labor force participation rates, the resulting coefficients are insignificant. We also find that increases in the lagged concentration index do not reduce the probability of a female immigrant adopting an American name, unlike what we document for males.

5.4 Determinants of American Name Choice for Second-Generation Immigrants

So far, we have seen in this section that male immigrants' decision to adopt an American name depends on how long the immigrant has been in the US and on his age and that it responds to the average occupational status of other immigrants in the same county and to the concentration in the same county of other immigrants who have the same country of origin. In this subsection, we examine to what extent these same variables influence the male immigrants' decision about whether or not to choose an American name for their children.

We essentially re-estimate equation (5) with a new dependent variable: an indicator for whether the child of the immigrant has an One difference relative to our regressions above is that, across specifications, we add up to five child-specific variables to the variables already used in equation (5): an indicator for whether the immigrant father has an American name, an indicator for whether the mother was also an immigrant, the number of other siblings between the ages of 0 and 10, the child's age and an indicator for whether the child was born in the US. As above, the model includes dummies for time, county of residence, the father's cohort of entry into the US and

the father's country of origin. Standard errors are clustered at the county level.

The results of this exercise are shown in Table 15. In column (1) and (3), we use exactly the same specification as in equation (5), without any interaction terms, respectively for sons and daughters. In column (2) and (4), for sons and daughters, respectively, we add five child-specific variables: child's age, an indicator for whether the child was US born, an indicator for whether the father has an American name, the number of siblings and an indicator for whether the mother was an immigrant.

Starting with the child-specific variables, having a father with an American name and being US born have substantial impacts on the probability of a son having an American name. The likelihood of adopting an American name increases by around 6 and 3 percentage points, respectively. Furthermore, having an immigrant mother decreases the likelihood of having an American name by around 1.5 p.p. In terms of parental variables, years in the US and father's age are consistently statistically important determinants of American name adoption, but the magnitude of the latter coefficient is very small.

The estimates of the effects of the economic variables are puzzling. The unemployment rate of immigrants has a negative impact on the adoption of an American name, while the unemployment rate of natives has precisely the opposite effect. However, the magnitudes of the coefficients are small, so these results do not appear to be quantitatively important.

Finally, in counties with a higher concentration index, immigrants are more likely to give their children non-American names. This result is consistent with what we found above for the immigrant's choice of his own name and is of similar magnitude.

We present results for the subsamples by father's nationality in Tables B.12 – B.14 in Online Appendix B, for fathers from Germany, Italy and the former USSR, respectively. The only variable that is consistently a statistically important determinant of American name adoption is having a father with an American name, with a bigger effect among boys with a father from Italy.

Regarding daughters, Table 15 shows similar patterns despite the different magnitudes. The main example is the impact of being US born, which increases the probability of having an American name by 42 percentage points. The impacts of having an immigrant mother and a father with an American name are smaller than for boys. By father's nationality, the only variable that is consistently a statistically important determinant of American name adoption is having been born in the US, with a bigger effect among girls with a father from Italy (see Tables B.15 – B.17).

6 Conclusion

This paper shows that the large majority of male immigrants to the United States at the turn of the 20th century adopted American names. This adoption was done soon after arrival in the country. There are however, substantial differences across nationalities in the degree to which American names were adopted. Our data also shows that American names were very common among female immigrants as well as among children of immigrants.

We then show that immigrants who adopted an American name were also more likely to be better assimilated in several other dimensions. Male immigrants had better labor market prospects and were more likely to become US citizens, and female immigrants were more likely to be in the labor force and to speak English. Both male and female immigrants were more likely to have a US-born spouse. These associations survived the inclusion of a large set of controls in the model, including indicators for the county of residence.

Finally, for male immigrants, we show that the adoption of an American name responded to economic and social pressures. Male immigrants living in counties where other immigrants had good labor market conditions found it less profitable to adopt an American name, presumably because they could develop high-valued economic trades with other immigrants living nearby. In addition, immigrants living in counties with a large concentration of other immigrants from the same nationality were less likely to Americanize their names than other immigrants who were more isolated in their county of residence. This may be related to the possibility that a large concentration of individuals of the same culture exerts pressure on each of them to preserve their culture, and therefore their name. Economic and social pressures were less important for female immigrants and also when it came to an immigrant's decision to give an American name to a child, in particular regarding daughters.

References

Abramitzky, R., Boustan, L. and Eriksson, K. (2012). Europe's tired, poor, huddled masses: self-selection and economic outcomes in the Age of Mass Migration. *American Economic Review*, 102(5), 1832–1856.

Abramitzky, R., Boustan, L. and Eriksson, K. (2013). Have the poor always been less likely to migrate? Evidence from inheritance practices during the Age of Mass Migration. *Journal of Development Economics*, 102, 2–14.

Abramitzky, R., Boustan, L. and Eriksson, K. (2014). A nation of immigrants:

assimilation and economic outcomes in the Age of Mass Migration. *Journal of Political Economy*, 122(3), 467–506.

Akerlof, G. and Kranton, R. (2000). Economics and identity. *Quarterly Journal of Economics*, 115(3), 715–753.

Arai, M. and Thoursie, P. (2009). Renouncing personal names: an empirical examination of surname change and earnings. *Journal of Labor Economics*, 27(1), 127–147.

Bandiera, O., Rasul, I. and Viarengo, M. (2013). The making of modern America: migratory flows in the Age of Mass Migration. *Journal of Development Economics*, 102, 23–47.

Benabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(3), 1652–1678.

Bertrand, M. and Mullainathan S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4), 991–1013.

Biavaschi, C., Giulietti, C. and Siddique, Z. (2013). The economic payoff of name Americanization. *IZA Discussion Paper* 7725.

Casey, T. and Dustmann, C. (2010). Immigrants' identity, economic outcomes and the transmission of identity across generations. *Economic Journal*, 120(542), F31–F50.

Charness, G. and Gneezy, U. (2008) What's in a name? Anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization*, 68(1), 29–35.

Cook, L. D., Logan, T. D. and Parman, J. M. (2014). Distinctively black names in the American past. *Explorations in Economic History*, 53, 64–82.

Fryer, R. and Levitt, S. (2004). The causes and consequences of distinctively black names. *Quarterly Journal of Economics*, 119(3), 767–805.

Güell, M., Mora, J. V. R. and Telmer, C. I. (2015). The informational content of surnames, the evolution of intergenerational mobility, and assortative mating. *Review of Economic Studies*, 82(2), 693–735.

Hatton, T. J. and Williamson, J. G. (1998). *The Age of Mass Migration: Causes and Economic Impact*. New York: Oxford University Press.

Lazear, E. (1999). Culture and language. *Journal of Political Economy*, 107(6), S95–S126.

Manning, A. and Roy, S. (2010). Culture clash or culture club? National identity in Britain. *Economic Journal*, 120(542), F72–F100.

National Archives and Records Administration (NARA). Italians to America Passenger Data File, 1855 - 1900. Data files relating to the immigration of Italians to the United States, created ca. 1977 - 2002, documenting the period 1855 - 1900, available at `http://aad.archives.gov/aad/`.

Olivetti, C. and Paserman, M. D. (2013). In the name of the son (and the daughter): intergenerational mobility in the United States, 1850-1930. *NBER Working Paper* 18822.

Rubinstein, Y. and Brenner, D. (2014). Pride and prejudice: using ethnic-sounding names and inter-ethnic marriages to identify labour market discrimination. *Review of Economic Studies*, 81(1), 389–425.

Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R. Schroeder, M. B. and Sobek, M. (2010). Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota.

Watkins, S. and London, A. (1994). Personal names and cultural change: a study of the naming patterns of Italians and Jews in the United States in 1910. *Social Science History*, 18(2), 169–209.

Table 1: Top 10 Popular Male Names - by Country of Birth

Germany		Italy		Former USSR		Poland	
Name	%	Name	%	Name	%	Name	%
JOHN	9.9	JOSEPH	8.6	JOHN	4.8	JOHN	12.2
HENRY	6.1	FRANK	6.7	SAMUEL	4.8	JOSEPH	8.9
WILLIAM	6.0	JOHN	6.4	JOSEPH	4.7	FRANK	5.8
CHARLES	4.3	ANTONIO	2.8	JACOB	4.4	STANLEY	3.7
FRED	3.7	LOUIS	2.8	MORRIS	4.2	MICHAEL	2.5
AUGUST	3.5	TONY	2.6	HARRY	4.1	PETER	2.4
JOSEPH	3.3	JAMES	2.5	LOUIS	4.1	WALTER	2.4
GEORGE	3.1	ANTHONY	2.5	MAX	3.4	JACOB	2.2
HERMAN	3.1	ANGELO	2.2	ABRAHAM	3.2	WILLIAM	2.1
FRANK	3.0	PETER	2.2	SAM	2.6	ANDREW	1.9

Sweden		Mexico		Norway		Hungary	
Name	%	Name	%	Name	%	Name	%
JOHN	13.8	JOSE	7.1	JOHN	9.3	JOHN	16.7
CHARLES	6.8	JUAN	4.9	OLE	8.4	JOSEPH	9.0
CARL	6.1	MANUEL	4.0	ANDREW	4.6	FRANK	4.1
ANDREW	4.5	JESUS	3.9	PETER	4.0	GEORGE	3.9
PETER	3.2	PEDRO	2.9	HANS	3.2	ANDREW	3.4
OSCAR	3.0	FRANCISCO	2.8	MARTIN	2.3	STEVE	3.3
AUGUST	2.7	ANTONIO	2.5	OLAF	1.8	MIKE	3.1
NELS	2.0	JOE	1.8	NELS	1.7	LOUIS	2.9
FRANK	1.9	JOHN	1.6	CARL	1.7	MICHAEL	2.7
AXEL	1.9	RAMON	1.3	THOMAS	1.6	STEPHEN	2.6

Czechoslovakia		Denmark		Greece		France	
Name	%	Name	%	Name	%	Name	%
JOHN	18.6	PETER	7.3	GEORGE	11.8	JOHN	7.2
JOSEPH	12.7	HANS	6.0	JOHN	10.6	JOSEPH	7.0
FRANK	10.9	JOHN	6.0	PETER	6.9	CHARLES	4.1
GEORGE	3.3	CHRIS	3.6	JAMES	6.6	LOUIS	4.0
ANDREW	3.2	ANDREW	3.5	NICK	3.6	HENRY	3.5
JAMES	3.0	CARL	3.2	THOMAS	2.7	GEORGE	3.5
MIKE	2.7	NELS	3.0	NICHOLAS	2.5	FRANK	2.7
ANTON	2.5	JAMES	2.9	WILLIAM	2.5	PETER	2.5
MICHAEL	2.3	CHRISTIAN	2.7	HARRY	2.3	AUGUST	2.2
CHARLES	2.2	JENS	2.1	LOUIS	2.2	PAUL	2.0

Japan		China		Portugal		Spain	
Name	%	Name	%	Name	%	Name	%
GEORGE	1.1	LEE	5.2	MANUEL	21.9	MANUEL	8.0
HARRY	0.7	AH	4.4	JOHN	12.4	JOSE	5.7
FRANK	0.6	SING	2.7	JOSEPH	12.0	JOSEPH	5.5
TOM	0.5	WONG	2.3	FRANK	7.1	JOHN	4.4
HENRY	0.3	SAM	2.1	ANTONE	5.8	FRANK	4.0
JOHN	0.3	YEE	1.7	ANTONIO	4.1	ANTONIO	3.5
KAMA	0.3	FONG	1.5	JOE	3.7	JOE	2.1
CHARLES	0.3	CHIN	1.4	JOSE	2.3	PEDRO	1.9
SAM	0.3	WAH	1.4	TONY	1.8	FRANCISCO	1.5
JOE	0.2	CHARLIE	1.3	ANTHONY	1.4	RAMON	1.4

Note:
IPUMS pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%).

Table 2: Top 10 Popular Female Names - by Country of Birth

Germany		Italy		Former USSR		Poland	
Name	%	Name	%	Name	%	Name	%
MARY	9.4	MARY	15.3	MARY	5.8	MARY	15.1
ANNA	7.1	ROSE	4.8	SARAH	5.6	ANNA	7.1
ELIZABETH	3.4	JOSEPHINE	4.6	ANNA	5.3	JOSEPHINE	2.9
MINNIE	2.9	ANNA	3.2	ROSE	4.8	ROSE	2.6
AUGUSTA	2.8	MARIA	3.0	IDA	4.3	HELEN	2.6
BERTHA	2.7	JENNIE	2.7	ANNIE	4.2	ANNIE	2.6
ANNIE	2.2	MARIE	2.0	FANNIE	2.9	STELLA	2.0
MARGARET	2.2	ANGELINA	1.8	LENA	2.8	JULIA	1.9
EMMA	2.1	ROSA	1.5	BESSIE	2.6	AGNES	1.7
MARIE	2.1	ANNIE	1.5	DORA	2.3	FRANCES	1.7

Sweden		Mexico		Norway		Hungary	
Name	%	Name	%	Name	%	Name	%
ANNA	10.4	MARIA	8.8	ANNA	8.1	MARY	16.7
EMMA	4.7	JUANA	2.8	MARY	5.2	ANNA	7.7
MARY	4.1	GUADALUPE	2.4	MARTHA	3.2	JULIA	5.2
IDA	3.4	MARY	3.2	BERTHA	3.0	ELIZABETH	4.9
AUGUSTA	2.9	CARMEN	1.8	ANNIE	2.7	ANNIE	4.2
HILDA	2.8	ANTONIA	1.5	MARIE	2.6	ROSE	3.2
ANNIE	2.6	DOLORES	1.5	CARRIE	2.4	HELEN	2.7
HANNAH	2.4	FRANCISCA	1.4	LENA	1.9	KATIE	1.7
MATILDA	2.2	PETRA	1.4	JULIA	1.7	BERTHA	1.6
ALMA	2.1	JOSEFA	1.3	INGEBORG	1.5	LIZZIE	1.5

Czechoslovakia		Denmark		Greece		France	
Name	%	Name	%	Name	%	Name	%
MARY	24.9	ANNA	11.0	MARY	12.6	MARY	10.1
ANNA	14.8	MARY	9.4	HELEN	6.5	MARIE	6.3
ANNIE	4.1	MARIE	5.7	ANNA	3.6	JOSEPHINE	3.0
JOSEPHINE	2.8	CHRISTINA	3.8	BESSIE	2.8	LOUISE	3.0
BARBARA	2.7	ANNIE	3.7	CATHERINE	2.3	ANNA	2.7
ELIZABETH	2.6	CARRIE	2.1	IRENE	2.2	MARGARET	2.2
MARIE	2.6	CHRISTINE	1.9	STELLA	2.2	ROSE	1.7
JULIA	2.4	JOHANNA	1.8	KATHERINE	2.2	JULIA	1.7
ROSE	1.8	HANNAH	1.5	GEORGIA	1.5	JEANNE	1.6
FRANCES	1.4	CAROLINE	1.4	ANNIE	1.5	JENNIE	1.3

Japan		China		Portugal		Spain	
Name	%	Name	%	Name	%	Name	%
TOMI	1.1	SHI	30.9	MARY	37.4	MARY	10.1
TOYO	0.8	LEE	6.8	MARIA	6.9	MARIA	5.9
KAME	0.7	YONG	4.3	ROSE	2.8	CARMEN	4.6
HARU	0.7	SU	1.9	ANNIE	2.3	JOSEPHINE	2.5
TOME	0.7	AH	1.6	MARIE	1.9	MARIE	2.3
YOSHI	0.6	CHIN	1.5	ROSA	1.9	DOLORES	1.9
TAKA	0.6	MARIE	1.4	ANNA	1.8	ANTONIA	1.6
MITSU	0.6	MAYME	1.4	FRANCES	1.3	ANNA	1.4
SUYE	0.6	EDITH	1.4	EMILY	1.2	FRANCES	1.3
KAMA	0.5	MARY	0.9	AMELIA	1.0	MERCEDES	1.3

Note:
IPUMS pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%).

Table 3: Percentage of Female and Male Immigrants with American Names - By Year of Census and Country of Birth

	Female	Male
Pooled Sample	76.9	76.5
By year		
1900	72.7	74.9
1910	76.6	75.7
1920	80.1	77.8
1930	79.4	77.1
By Country of Birth		
Germany	71.7	77.9
Italy	73.2	70.1
Former USSR	92.3	93.1
Poland	88.5	91
Sweden	72.8	73.5
Mexico	49.5	39.6
Norway	67.2	58.1
Hungary	92.6	91.6
Czechoslovakia	90.9	92.7
Denmark	72.1	67.8
Greece	66.9	77.2
France	68.7	74.9
Japan	1.7	7.5
China	6.4	23.9
Portugal	82.2	62.4
Spain	26.6	42.3

Notes:

Pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%).

Immigrants between 16 and 65 years old.

Table 4: Summary Statistics - American Name vs Non-American Name (1900-1920)

Variables	Mean (standard deviation)			
	Males		Females	
	American name	Non-American name	American name	Non-American name
Years in the US	18.1	17.6	18.4	20.0
	(12.1)	(12.2)	(12.0)	(13.0)
Age	38.4	38.9	37.4	40.6
	(12.4)	(12.5)	(12.8)	(13.2)
Log population	9.2	8.8	8.8	8.3
	(1.8)	(1.8)	(1.8)	(1.8)
Economic Variables (by geographical unit)				
Immigr. unemp. rate$_{-1}$	18.8	17.5	10.7	9.9
	(10.5)	(11.0)	(9.3)	(10.6)
Native unemp. rate$_{-1}$	14.8	14.6	12.6	12.2
	(7.3)	(8.0)	(8.6)	(10.1)
Immigr. labor force rate$_{-1}$	93.9	93.8	21.2	20.1
	(3.9)	(4.2)	(8.8)	(9.4)
Native labor rate$_{-1}$	89.2	88.8	26.8	24.4
	(4.0)	(4.4)	(8.8)	(9.2)
Immigr. log occ. score$_{-1}$	3.10	3.04	2.41	2.37
	(0.17)	(0.20)	(0.22)	(0.25)
Native log occ. score$_{-1}$	3.13	3.09	2.68	2.64
	(0.19)	(0.22)	(0.21)	(0.24)
Network Variables (by geographical unit)				
Concentration index$_{-1}$	7.3	9.8	6.8	9.7
	(8.0)	(11.7)	(7.8)	(11.3)
Log immigr. population$_{-1}$	7.58	7.21	7.7	7.2
	(1.92)	(1.96)	(2.0)	(2.0)
Own immigr. literacy percentage$_{-1}$	86.5	83.4	81.4	82.7
	(17.9)	(21.1)	(21.1)	(23.1)
Own immigr. American name percentage$_{-1}$	80.2	66.8	76.6	67.8
	(17.3)	(26.2)	(16.8)	(19.4)
Measures of Immigrants Assimilation				
Log occupational score	3.12	3.03	2.66	2.54
	(0.39)	(0.43)	(0.64)	(0.65)
Full-Year employed last year (%)	82.8	83.6	96.2	96.8
	(37.7)	(37.0)	(19.2)	(17.7)
Labor Force participation (%)	94.0	93.8	18.9	17.0
	(23.7)	(24.2)	(39.1)	(37.6)
American wife/husband	5.2	5.2	11.7	10.9
	(22.2)	(22.3)	(32.2)	(31.2)
Speaks English	86.5	79.0	80.8	74.5
	(34.5)	(40.7)	(39.4)	(43.6)
US citizenship (%)	53.8	44.7	52.5	48.6
	(49.9)	(49.7)	(49.9)	(50.0)

Notes:
Pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920 (1%).

Geographical unit - county. For further detail, see note from Table B.2.

Table 5: Fathers' and Sons'/Daughters' Naming Patterns

Whole Sample

		Father				Father	
		Non-American name	American name			Non-American name	American name
		Percentage				Percentage	
Son	Non-American name	21.9	13.6	Daughter	Non-American name	16.5	10.9
	American name	78.1	86.4		American name	83.5	89.1

German Father

		Non-American name	American name			Non-American name	American name
		Percentage				Percentage	
Son	Non-American name	18.9%	12.9%	Daughter	Non-American name	8.0	6.8
	American name	81.1%	87.1%		American name	92.0	93.2

Italian Father

		Non-American name	American name			Non-American name	American name
		Percentage				Percentage	
Son	Non-American name	22.6%	13.1%	Daughter	Non-American name	19.9	13.6
	American name	77.4%	86.9%		American name	80.1	86.4

Former-USSR Father

		Non-American name	American name			Non-American name	American name
		Percentage				Percentage	
Son	Non-American name	20.4%	15.3%	Daughter	Non-American name	17.0	14.3
	American name	79.6%	84.7%		American name	83.0	85.7

Note:
IPUMS pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920 (1%).

Table 6: Assimilation Outcomes - Male Immigrants

	Log Occupational Score	Full-year Employed Last Year	American Wife (exc. Second generation)	Speaks English	US Citizenship
Total					
American Name	0.005	-0.001	0.006	0.026	0.010
	(0.002)*	(0.003)	(0.002)***	(0.002)***	(0.002)***
Observations	229415	155910	147282	238619	229051
Germany					
American Name	-0.007	-0.010	0.001	0.002	0.005
	(0.004)*	(0.004)**	(0.004)	(0.003)	(0.005)
Observations	75738	65184	55940	79265	76555
Italy					
American Name	0.008	0.005	0.013	0.031	0.003
	(0.004)**	(0.010)	(0.002)***	(0.007)***	(0.006)
Observations	32083	16807	18735	33219	31739
Former USSR					
American Name	0.027	0.014	0.003	0.032	0.029
	(0.008)***	(0.014)	(0.003)	(0.011)***	(0.010)***
Observations	26269	15255	17422	27485	25593

Notes:
All regression include as controls: Census Year, Birthplace, Years in the US, Age, Log Population, Cohort of entry and County.
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
Years in the US, and age enter in quadratic form.

Table 7: Assimilation Outcomes - Female Immigrants

	(1)	(2)	(3)
Log occupational score	0.118	-0.004	-0.007
	(0.017)***	(0.012)	(0.012)
Full-year employed last year	-0.006	0.001	0.001
	(0.004)*	(0.002)	(0.002)
Labor force participation	0.018	0.005	0.006
	(0.005)***	(0.003)	(0.003)**
American husband	0.008	0.024	0.023
	(0.004)**	(0.003)***	(0.003)***
Speaks English	0.062	0.032	0.026
	(0.010)***	(0.004)***	(0.003)***
US citizenship	0.039	0.003	0.004
	(0.014)***	(0.005)	(0.005)
Controls			
Census Year	✓	✓	✓
Birthplace	×	✓	✓
Years in the US	×	✓	✓
Age	×	✓	✓
Log population	×	✓	✓
Cohort of entry	×	✓	✓
County	×	×	✓

Notes:
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
Years in the US, and age enter in quadratic form.

Table 8: American Name (Probit - Marginal Effects) - Male Immigrants

	(1)	(2)	(3)
Years in the US^{+}	0.190	0.199	0.193
	(0.068)***	(0.068)***	(0.068)***
Age^{+}	-0.136	-0.136	-0.136
	(0.015)***	(0.015)***	(0.015)***
Log population	-2.040	-1.995	-2.281
	(1.454)	(1.389)	(1.484)
Immigr. unemp. rate$_{-1}$	-0.007	0.009	-0.005
	(0.024)	(0.025)	(0.025)
Native unemp. rate$_{-1}$	0.005	-0.001	0.006
	(0.036)	(0.037)	(0.037)
Immigr. log occ. score$_{-1}$	-7.093	-6.569	-8.785
	(2.929)**	(2.955)**	(3.030)***
Native log occ. score$_{-1}$	3.551	2.990	6.005
	(2.733)	(2.822)	(2.943)**
Concentration index$_{-1}{}^{+}$	-0.150	-0.156	-0.142
	(0.046)***	(0.044)***	(0.042)***
Log immigr. population$_{-1}$	0.146	0.118	0.201
	(0.397)	(0.398)	(0.392)
Own immigr. literacy percentage$_{-1}$	1.858	0.761	1.778
	(1.433)	(1.467)	(1.476)
Own immigr. American name percentage$_{-1}$	0.009	0.009	0.026
	(0.013)	(0.013)	(0.015)
1910	-0.265	-0.248	-0.496
	(1.531)	(1.516)	(1.542)
1920	4.812	4.580	4.771
	(1.590)***	(1.576)***	(1.593)***
Observations	174509	174509	174509
Other controls	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County
Joint hypothesis test			
Economic variables	6.45	-	-
(p-value)	0.168	-	-
Network variables	21.11	-	-
(p-value)	0.001	-	-

Notes:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
+ Years in the US, age and concentration index are always included in quadratic form.
In column (2), all economic and network variables are interacted with "Years in the US" . In column (3), all economic and network variables are interacted with "Concentration index$_{-1}$" . There are no interactions with economic and network variables in column (1).

Table 9: Years in the US interacted with Economic and Network Variables - Male Immigrants

	Years in the US				
	0	25th	Median	75th	90th
Economic Variables					
Immigr. unemp. rate$_{-1}$	0.078	0.040	0.015	-0.017	-0.046
	(0.047)*	(0.033)	(0.025)	(0.022)	(0.028)*
Native unemp. rate$_{-1}$	-0.089	-0.047	-0.019	0.017	0.048
	(0.065)	(0.046)	(0.037)	(0.034)	(0.042)
Immigr. log occ. score$_{-1}$	-6.289	-6.269	-6.315	-6.553	-7.005
	(4.401)	(3.403)*	(2.938)**	(2.795)**	(3.166)**
Native log occ. score$_{-1}$	3.730	3.256	2.963	2.664	2.503
	(4.510)	(3.391)	(2.835)	(2.592)	(2.927)
Network Variables					
Concentration index$_{-1}$	-0.079	-0.116	-0.142	-0.181	-0.221
	(0.058)	(0.048)**	(0.044)***	(0.040)***	(0.042)***
Log immigr. population$_{-1}$	-0.150	0.010	0.086	0.215	0.334
	(0.498)	(0.425)	(0.391)	(0.376)	(0.396)
Own immigr. literacy percentage$_{-1}$	4.346	2.41	1.104	-0.568	-2.043
	(1.896)**	(1.478)	(1.400)	(1.625)	(2.071)
Own immigr. American name percentage$_{-1}$	0.029	0.018	0.011	0.000	-0.007
	(0.021)	(0.016)	(0.013)	(0.012)	(0.015)

Notes:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points) evaluated at different values of years in the US.

Robust standard errors in parentheses clustered by county.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 10: Lagged Concentration Index interacted with Economic and Network Variables - Male Immigrants

	Lagged Concentration Index				
	0	25th	Median	75th	90th
Economic Variables					
Immigr. unemp. rate$_{-1}$	0.009	0.007	-0.000	-0.016	-0.028
	(0.034)	(0.032)	(0.027)	(0.029)	(0.042)
Native unemp. rate$_{-1}$	-0.002	-0.001	0.003	0.012	0.019
	(0.051)	(0.048)	(0.041)	(0.039)	(0.054)
Immigr. log occ. score$_{-1}$	-13.401	-12.779	-10.455	-5.219	-1.060
	(3.404)***	(3.330)***	(3.128)***	(3.155)*	(3.594)
Native log occ. score$_{-1}$	9.776	9.265	7.358	3.085	0.293
	(3.607)***	(3.500)**	(3.149)**	(2.785)	(2.986)
Network Variables					
Log immigr. population$_{-1}$	-0.0376	-0.004	0.121	0.389	0.593
	(0.407)	(0.404)	(0.396)	(0.413)	(0.457)
Own immigr. literacy percentage$_{-1}$	1.509	1.552	1.702	1.997	2.199
	(1.440)	(1.398)	(1.382)	(1.993)	(2.745)
Own immigr. American name percentage$_{-1}$	-0.010	-0.004	0.014	0.054	0.085
	(0.013)	(0.013)	(0.014)	(0.022)**	(0.030)***

Note:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points) evaluated at different values of the lagged concentration index.

Robust standard errors in parentheses clustered by county.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 11: American Name (Probit - Marginal Effects) - Excluding Various Countries - Male Immigrants

	Excluding Germans	Excluding Italians	Excluding former USSR
Years in the US^{+}	0.293	0.074	0.216
	(0.094)***	(0.068)	(0.078)***
Age^{+}	-0.169	-0.095	-0.144
	(0.017)***	(0.014)***	(0.017)***
Log population	-1.720	-1.143	-2.240
	(1.670)	(1.413)	(1.636)
Immigr. unemp. rate$_{-1}$	2.234	0.119	-1.651
	(3.363)	(2.342)	(2.674)
Native unemp. rate$_{-1}$	0.694	-0.544	3.521
	(4.584)	(3.483)	(4.047)
Immigr. log occ. score$_{-1}$	-11.535	-4.070	-6.434
	(3.924)***	(2.790)	(3.252)**
Native log occ. score$_{-1}$	7.973	2.179	2.845
	(3.278)**	(2.576)	(3.105)
Concentration index$_{-1}{}^{+}$	-0.246	-0.132	-0.245
	(0.064)***	(0.041)***	(0.045)***
Log immigr. population$_{-1}$	-0.168	-0.002	0.422
	(0.576)	(0.379)	(0.430)
Own immigr. literacy percentage$_{-1}$	1.243	0.379	2.417
	(1.485)	(1.612)	(1.677)
Own immigr. American name percentage$_{-1}$	-0.001	-0.003	-0.002
	(0.015)	(0.014)	(0.014)
1910	0.725	0.382	-1.020
	(1.846)	(1.515)	(1.742)
1920	5.885	3.689	4.121
	(1.991)***	(1.558)**	(1.872)**
Observations	103018	152948	155657
Other controls	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County
Joint hypothesis test			
Economic variables	11.69	2.30	5.17
(p-value)	0.020	0.680	0.270
Network variables	21.45	15.70	37.96
(p-value)	0.001	0.008	0.000

Notes:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
+ Years in the US, age and concentration index are always included in quadratic form.
The specification is the one in column (1) of Table 8, where there are no interactions with economic and network variables.

Table 12: American Name (Probit - Marginal Effects) - Robustness Check (Economic Variables) - Male Immigrants

	(1)	(2)	(3)
Years in the US^{+}	0.199	0.201	0.191
	(0.067)***	(0.067)***	(0.068)**
Age^{+}	-0.136	-0.136	-0.136
	(0.015)***	(0.015)***	(0.016)***
Log population	-2.239	-2.343	-2.664
	(1.487)	(1.476)	(1.476)*
Immigr. unemp. rate$_{-1}$	-1.186	-	-
	(2.884)	-	-
Own immigr. unemp. rate$_{-1}$	-0.128	-0.327	0.824
	(1.902)	(1.752)	(2.078)
Exc. own immigr. unemp. rate$_{-1}$	-	-	-2.166
	-	-	(1.944)
Native unemp. rate$_{-1}$	2.070	1.055	2.147
	(3.727)	(3.391)	(3.470)
Immigr. log occ. score$_{-1}$	-5.081	-	-
	(3.283)	-	-
Own immigr. log occ. score$_{-1}$	-1.497	-2.006	-2.237
	(1.202)	(1.128)*	(1.149)**
Exc. own immigr. log occ. score$_{-1}$	-	-	-4.206
	-	-	(1.693)***
Native log occ. score$_{-1}$	1.424	0.238	0.174
	(2.831)	(2.678)	(2.842)
Concentration index$_{-1}{}^{+}$	-0.151	-0.149	-0.127
	(0.046)***	(0.046)***	(0.045)***
Log immigr. population$_{-1}$	0.403	0.459	0.654
	(0.411)	(0.415)	(0.433)*
Own immigr. literacy percentage$_{-1}$	2.231	2.247	1.797
	(1.564)	(1.563)	(1.561)
Own immigr. American name percentage$_{-1}$	0.020	0.020	0.030
	(0.014)	(0.014)	(0.014)**
1910	-0.567	-0.775	-0.938
	(1.569)	(1.550)	(1.560)
1920	4.775	4.675	5.203
	(1.599)***	(1.597)***	(1.617)**
Observations	172893	172893	170647
Other controls	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County
Joint hypothesis test			
Economic variables	6.81	3.43	10.61
(p-value)	0.339	0.488	0.101
Network variables	29.44	28.55	33.20
(p-value)	0.000	0.000	0.000

Notes:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
+ Years in the US, age and concentration index are always included in quadratic form.
The specification is the one in column (1) of Table 8, where there are no interactions with economic and network variables.

Table 13: American Name (Probit - Marginal Effects) - Robustness Check (1920 vs 1930) - Male Immigrants

	1900-1920	1900-1930
Years in the US^{+}	0.190	0.214
	(0.068)***	(0.067)***
Age^{+}	-0.137	-0.161
	(0.015)***	(0.013)***
Log population	-2.032	-0.437
	(1.454)	(0.730)
Immigr. log occ. score$_{-1}$	-7.129	-5.887
	(2.924)**	(2.117)***
Native log occ. score$_{-1}$	3.603	1.521
	(2.685)	(2.021)
Concentration index$_{-1}{}^{+}$	-0.149	-0.065
	(0.046)***	(0.039)*
Log immigr. population$_{-1}$	0.146	0.090
	(0.397)	(0.352)
Own immigr. literacy percentage$_{-1}$	1.853	1.859
	(1.431)	(1.192)
Own immigr. American name percentage$_{-1}$	0.009	0.010
	(0.013)	(0.011)
1910	-0.283	0.610
	(1.482)	(1.078)
1920	4.780	3.526
	(1.546)***	(1.481)**
1930	-	2.886
	-	(2.094)
Observations	174509	210586
Other controls	Birthplace Cohort of entry County	Birthplace Cohort of entry County
Joint hypothesis test		
Economic variables	6.30	7.77
(p-value)	0.043	0.021
Network variables	21.19	22.4
(p-value)	0.001	0.000

Notes:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
+ Years in the US, age and concentration index are always included in quadratic form.
The specification is the one in column (1) of Table 8, where there are no interactions with economic and network variables.

Table 14: American Name (Probit - Marginal Effects) - Female Immigrants

	(1)	(2)	(3)
Years in the US^{+}	0.046	0.076	0.045
	(0.090)	(0.091)	(0.090)
Age^{+}	-0.302	-0.301	-0.302
	(0.021)***	(0.021)***	(0.021)***
Log population	-1.407	-1.304	-0.896
	(1.926)	(1.760)	(2.014)
Immigr. labor force rate$_{-1}$	-0.011	-0.0004	-0.006
	(0.050)	(0.049)	(0.049)
Native labor force rate$_{-1}$	0.250	0.243	0.231
	(0.068)***	(0.069)***	(0.072)***
Immigr. log occ. score$_{-1}$	-0.184	-0.348	-0.191
	(1.119)	(1.139)	(1.119)
Native log occ. score$_{-1}$	-1.429	-1.996	-0.671
	(1.643)	(1.647)	(1.729)
Concentration index$_{-1}{}^{+}$	0.054	0.030	0.085
	(0.050)	(0.047)	(0.050)*
Log immigr. population$_{-1}$	-0.095	-0.252	-0.107
	(0.641)	(0.630)	(0.664)
Own immigr. literacy percentage$_{-1}$	1.719	-1.028	-0.118
	(1.450)	(1.520)	(1.618)
Own immigr. American name percentage$_{-1}$	-0.008	-0.007	-0.020
	(0.014)	(0.014)	(0.016)
1910	2.356	2.552	2.632
	(2.114)	(2.004)	(2.242)
1920	6.690	6.206	6.408
	(1.981)***	(1.930)***	(1.951)***
Observations	126136	126136	126136
Other controls	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County
Joint hypothesis test			
Economic variables	15.64	-	-
(p-value)	0.004	-	-
Network variables	10.25	-	-
(p-value)	0.068	-	-

Notes:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
+ Years in the US, age and concentration index are always included in quadratic form. In column (2), all economic and network variables are interacted with "Years in the US" . In column (3), all economic and network variables are interacted with "Concentration index$_{-1}$" . There are no interactions with economic and network variables in column (1).

Table 15: American Name (Probit - Marginal Effects) - Sons and Daughters

	Sons		Daughters	
	(1)	(2)	(3)	(4)
Father with American Name	-	5.826	-	1.667
	-	(0.368)***	-	(0.218)***
Immigrant Mother	-	-1.461	-	-0.197
	-	(0.584)**	-	(0.329)
Siblings (aged 0-10)	-	0.017	-	0.016
	-	(0.164)	-	(0.072)
Child's Age	-	0.073	-	0.082
	-	(0.070)	-	(0.041)**
Child US born	-	2.935	-	41.970
	-	(0.795)***	-	(0.768)***
Years in the US^{+}	0.259	0.208	1.209	-0.044
	(0.101)***	(0.100)**	(0.093)***	(0.054)
Age^{+}	-0.056	-0.031	-0.463	-0.044
	(0.022)**	(0.024)	(0.025)***	(0.016)***
Log population	-1.546	-1.303	-4.215	-0.707
	(1.367)	(1.381)	(1.665)**	(0.968)
Immigr. unemp. rate$_{-1}$	-0.066	-0.067	-0.034	-0.033
	(0.038)*	(0.037)*	(0.033)	(0.024)
Native unemp. rate$_{-1}$	0.099	0.098	-0.044	-0.005
	(0.053)*	(0.053)*	(0.050)	(0.036)
Immigr. log occ. score$_{-1}$	1.080	1.139	-4.037	-4.284
	(4.376)	(4.323)	(3.893)	(2.681)
Native log occ. score$_{-1}$	-2.088	-1.794	1.703	1.543
	(3.898)	(3.839)	(3.438)	(2.377)
Concentration index$_{-1}{}^{+}$	-0.112	-0.105	0.012	-0.031
	(0.038)***	(0.037)***	(0.035)	(0.021)
Log immigr. Population$_{-1}$	0.654	0.672	-0.234	-0.199
	(0.613)	(0.619)	(0.603)	(0.355)
Own immigr. Literacy percentage$_{-1}$	0.364	0.113	0.573	-0.456
	(1.675)	(1.653)	(1.362)	(0.991)
Own immigr. American name percentage$_{-1}$	0.001	0.001	0.017	0.012
	(0.015)	(0.015)	(0.013)	(0.010)
1910	-1.022	-1.051	-0.937	1.984
	(1.613)	(1.621)	(1.842)	(1.116)*
1920	0.903	0.095	8.379	4.610
	(2.071)	(2.002)	(1.868)***	(1.181)***
Observations	74724	74724	70309	70309
Other controls (father)	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County

Notes:

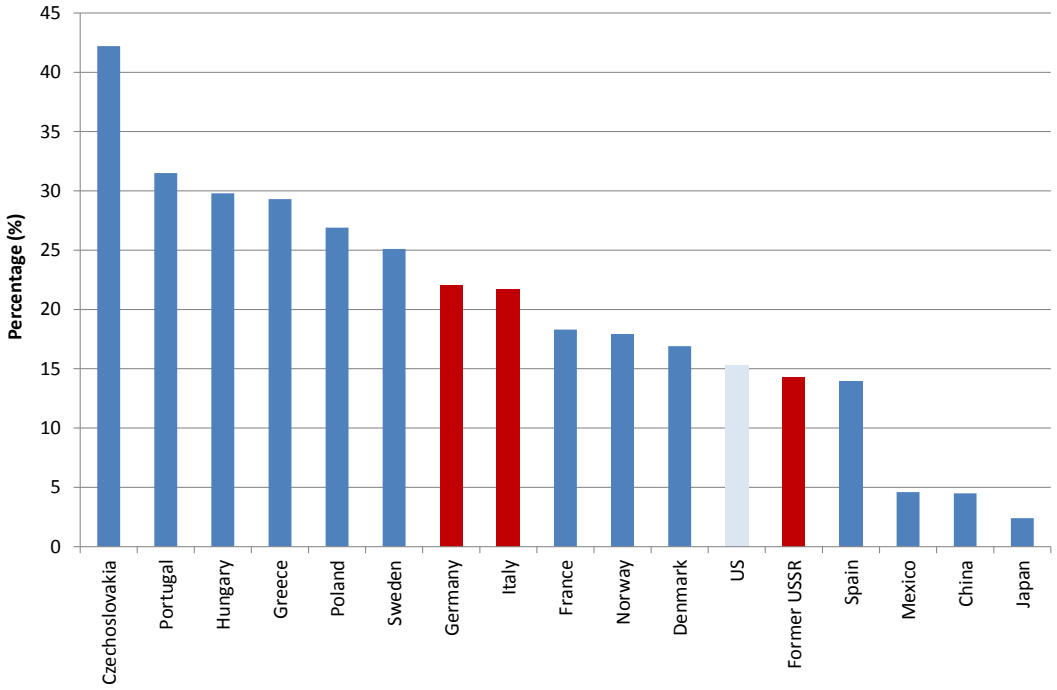
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).

Robust standard errors in parentheses clustered by county.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

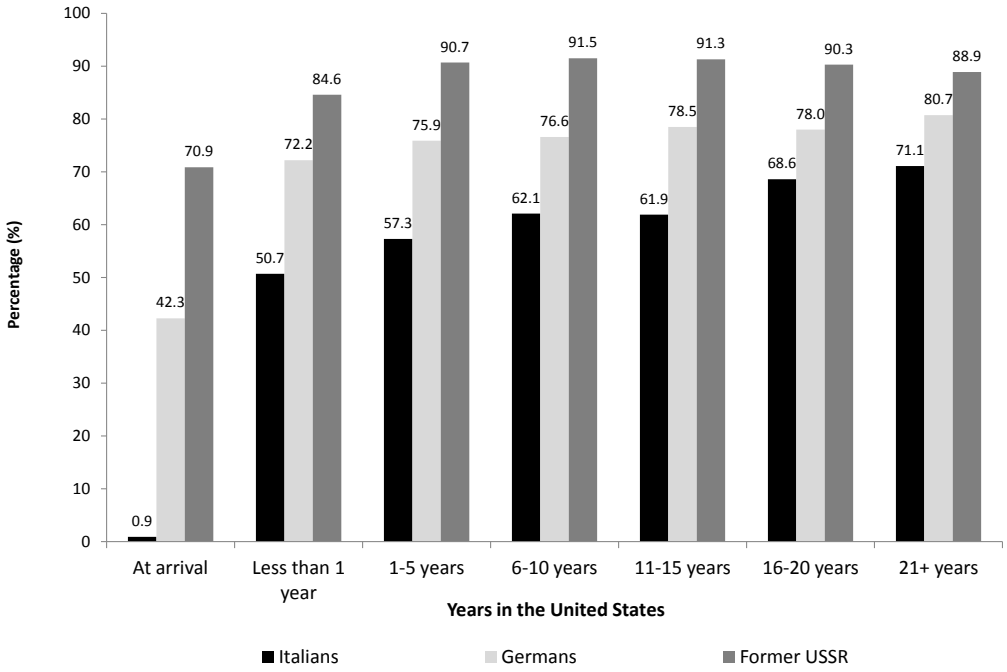
+ Years in the US, age and concentration index are always included in quadratic form.

Figure 1: Male Immigrants - Top Three American Names



Notes:
IPUMS pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%). Top three American names in this figure include only American names. As a result, they are different from those in Table 1, which includes all names. For example, Manuel, the most popular name among Portuguese immigrants, is excluded from this classification. Therefore, for Portuguese immigrants, the top three American names are John, Joseph and Frank. As we can see in Table 1, the top three American names are not necessarily the same across countries.

Figure 2: Male Immigrants with American First Names



Notes:
At arrival: US National Archive data - Italians to America Passenger Data File, 1855-1900; Germans to America Passenger Data File, 1850-1897; Former USSR immigrants to America Passenger Data File, 1834-1897. For each country we used a random sample of around 1000 males between 16 and 65 years old. For further details, see http://aad.archives.gov/aad/ .
All other figures are from IPUMS - the 1900 census.
In this figure, the names definition follows the steps presented in Online Appendix A.

Online Appendices to "Please Call Me John: Name Choice and the Assimilation of Immigrants in the United States, 1900-1930"

A Data Appendix: Classification of Name Types

This appendix gives a detailed description of the classification of the name types. In the analysis, we distinguish the immigrants by the type of their first name: "Origin" names and "American" names. The procedure for attributing the type to each name was as follows:

Step 1 Each name in the IPUMS census was matched with two different sources to obtain each type.

Step 1a For Origin names, we used an internet site regarding etymology and history of first names, which includes current names but also old-fashioned ones for all countries included in the study (`http://www.behindthename.com/`).

Step 1b For US names, we used the US Social Security Administration database (`http://www.ssa.gov/OACT/babynames/`) with the most popular names of the 1900s, which includes a list with the rank, the name and the number of occurrences of the (top 1000) most popular given names for male babies born during 1900-1909. All names are from Social Security card applications for births that occurred in the US. As an example, the list of the top 100 names is reported in Table A.1.

Step 2 We searched for misspelling of names (e.g. JOHHN for JOHN or ANONIO for ANTONIO).

Step 3 From the lists of Origin and US names, we identified names in both lists, which we defined as "Common" names. This variable indicates whether a person's first name is included in the US Social Security database and is also present in the Origin names list (for example, LEO is an Italian name and is in rank 44 of the US Social Security database).

Step 4a We grouped the identified names according to type - US names, Origin names, Common names and Other names. The US names include a person's first name if it is in the US Social Security database and is not present in the Origin

names list. The Common names include all the names included in both lists - the US Social Security database and the Origin names list. The Origin names include only those names in the Origin list that are not in the US Social Security database. Finally, the last category refers to the case when a person's first name is neither an American name nor from a country of origin (for example, an Italian immigrant with a French name (JEROME)).

Step 4b The Other names were additionally classified as Group, British and Other. The Other names "Group" includes a person's first name if it belongs to their group of countries: Latin (Portugal, Spain, France and Italy), Scandinavian (Sweden, Norway and Denmark), Eastern European (Czechoslovakia, Hungary, Poland and former USSR) and Asian (Japan and China). For Greeks and Mexicans, we used the Latin group; for Germany, we used the Scandinavian group. The Other names "British" includes a person's first name if it belongs to a British country (e.g. England). The Other names "Other" are those remaining after this division. Also, for some names it was impossible to obtain classification by the rules described so far. The Other (Other), unclassified and missing names were treated as "Missing" names.

Step 5a We classified, as American names, all US names and all Common names in the top 100 most popular names in the Social Security database. All remaining names were classified as non-American names. In other words, the list of American names includes all names in the top 100 in the US Social Security database (see Table A.1), plus all the remaining names in that list (101-1000) that were not present in the Origin list. Additionally, it includes the British names included in the Other (British) list.

Step 5b Observations with Missing names were deleted from the sample.

Since the classification of first names is important for our analysis, we now make detailed comments on name classification. In what follows, for the brevity of the paper, we focus on the case of males. After Step 1, around 81% of the names were identified. Step 2 accounted for around 10% of the names.[24] Therefore, from Steps 1 and 2, we obtained a classification of types for around 90% of first names for most countries.[25] The Other names never represented more than 2%, except in the cases of Mexico and China.

[24]This figure depends on the census, but it never exceeded 11%.

[25]By country, Japan and Mexico present most difficulties in identifying the type of first names.

Table A.2 presents the top 10 names in the US. The first pair of columns show the list of top 10 names identified from the US Social Security Administration database, which provides the most popular given names for male babies born between 1900 and 1909. The second pair of columns correspond to the list of the top 10 names constructed by us using data from the US censuses between 1900 and 1930. The only difference between the two lists is that the first one includes Thomas while the second one includes Henry, so we should be able to use the two lists interchangeably. In the paper, we define American names based on the Social Security list, which presumably reflects more accurately the names of natives, since it includes the whole population of infants applying for a Social Security number at the beginning of the century.

Table A.3 presents the percentage of names that are common to the list of Origin and US names, depending on whether we used TOP 1000, TOP 200 or TOP 100 from the Social Security database. As we can see from the table, in general, the percentage of immigrants with common names in TOP 1000 shows considerable heterogeneity. On the one hand, we have countries with high percentages such as Germany and France reaching around 50%. Other Latin countries apart from Italy and the Scandinavian countries present lower figures but still above 30%. On the other hand, we find the Eastern European countries and Asian countries with figures below 10%. Interestingly, as we move to TOP 200 and in particular to TOP 100, the percentage is reduced dramatically in most countries, especially Latin countries, where the percentage falls below 5%. These percentages remain high in countries such as France and Germany, with figures above 35%. Nevertheless, it is reasonable to consider a Common name in the top 100 as an American name. All the remaining Common names are considered as Origin names (non-American). Therefore, our measure of American name, and consequently of name changers, comprises the American names and the Common top 100. As highlighted before, in Germany and France an American name does not necessarily represent someone who changed their name.

Table A.4 reports the proportions of each type of name across different birthplaces. This table shows significant heterogeneity in the sample. Table A.5 shows some examples of name classification (into American and non-American names) for the top three countries in terms of percentage of immigrants - Germany, Italy and the former USSR.

Table A.1: Most Popular Names in the 1900s - Males

Rank	**Name**	**Number**	Percentage	**Rank**	**Name**	**Number**	Percentage
1	John	84602	5.7644	51	Floyd	4968	0.3385
2	William	69331	4.7239	52	Donald	4639	0.3161
3	James	62174	4.2363	53	Kenneth	4582	0.3122
4	George	43592	2.9702	54	Jesse	4559	0.3106
5	Charles	36193	2.4660	55	Russell	4552	0.3102
6	Robert	35871	2.4441	56	Clyde	4527	0.3085
7	Joseph	35176	2.3967	57	Oscar	4517	0.3078
8	Frank	29054	1.9796	58	Peter	4376	0.2982
9	Edward	24519	1.6706	59	Lester	4361	0.2971
10	Thomas	21789	1.4846	60	Leroy	4301	0.2931
11	Henry	21772	1.4835	61	Ray	4258	0.2901
12	Walter	19996	1.3624	62	Stanley	4155	0.2831
13	Harry	17908	1.2202	63	Clifford	4054	0.2762
14	Willie	17749	1.2093	64	Lewis	4043	0.2755
15	Arthur	15407	1.0498	65	Benjamin	4009	0.2732
16	Albert	15320	1.0438	66	Edwin	3959	0.2697
17	Clarence	13573	0.9248	67	Frederick	3900	0.2657
18	Fred	13072	0.8907	68	Chester	3789	0.2582
19	Harold	12791	0.8715	69	Claude	3704	0.2524
20	Paul	12640	0.8612	70	Eddie	3697	0.2519
21	Raymond	11648	0.7936	71	Cecil	3610	0.2460
22	Richard	10693	0.7286	72	Lloyd	3506	0.2389
23	Roy	10633	0.7245	73	Jessie	3377	0.2301
24	Joe	10466	0.7131	74	Martin	3344	0.2278
25	Louis	10377	0.7070	75	Bernard	3297	0.2246
26	Carl	9865	0.6722	76	Tom	3293	0.2244
27	Ralph	9302	0.6338	77	Will	3247	0.2212
28	Earl	9253	0.6305	78	Norman	3177	0.2165
29	Jack	9225	0.6286	79	Edgar	3154	0.2149
30	Ernest	8658	0.5899	80	Harvey	3105	0.2116
31	David	8636	0.5884	81	Ben	3028	0.2063
32	Samuel	8150	0.5553	82	Homer	2915	0.1986
33	Howard	8020	0.5464	83	Luther	2866	0.1953
34	Charlie	7927	0.5401	84	Leon	2863	0.1951
35	Francis	7020	0.4783	85	Melvin	2841	0.1936
36	Herbert	6958	0.4741	86	Philip	2816	0.1919
37	Lawrence	6849	0.4667	87	Johnnie	2777	0.1892
38	Theodore	6491	0.4423	88	Jim	2717	0.1851
39	Alfred	6356	0.4331	89	Milton	2694	0.1836
40	Andrew	6281	0.4280	90	Everett	2596	0.1769
41	Sam	6161	0.4198	91	Allen	2559	0.1744
42	Elmer	6160	0.4197	92	Leslie	2531	0.1725
43	Eugene	5910	0.4027	93	Alvin	2484	0.1692
44	Leo	5632	0.3837	94	Victor	2445	0.1666
45	Michael	5230	0.3564	95	Marvin	2324	0.1583
46	Lee	5158	0.3514	96	Stephen	2249	0.1532
47	Herman	5065	0.3451	97	Alexander	2231	0.1520
48	Daniel	5042	0.3435	98	Jacob	2191	0.1493
49	Anthony	5042	0.3435	99	Hugh	2176	0.1483
50	Leonard	5021	0.3421	100	Patrick	2143	0.1460

Source: US Social Security Administration database.
Notes:
This table includes a list with the rank, the name and the number of occurrences of the most popular given names for male babies born during 1900-1909. All names are from Social Security card applications for births that occurred in the United States. The database also includes the top 1000 but for simplicity and space constraints we provide only the top 100.

Table A.2: Top 10 Popular Male Names in the US

US (SS data)*		US (IPUMS)**	
Name	%	Name	%
JOHN	5.8	JOHN	6.6
WILLIAM	4.7	WILLIAM	5.3
JAMES	4.2	GEORGE	3.4
GEORGE	3.0	JAMES	3.4
CHARLES	2.5	CHARLES	2.9
ROBERT	2.4	FRANK	2.4
JOSEPH	2.4	JOSEPH	2.4
FRANK	2.0	ROBERT	2.0
EDWARD	1.7	EDWARD	1.9
THOMAS	1.5	HENRY	1.9

Notes:
* US Social Security Administration database - the most popular given names for male babies born during 1900-1909; takes into account the number of occurrences of all names from Social Security card applications for births that occurred in the United States.
** IPUMS data, pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%).

Table A.3: Common Names - Males

	Top 1000	Top 200	Top 100
Germany	53.4	47.7	38.3
Italy	10.7	0.3	0.3
Former USSR	4.3	3.5	3.0
Poland	8.6	8.0	6.2
Sweden	33.0	24.2	19.4
Mexico	21.9	10.1	0.8
Norway	32.8	16.1	14.1
Hungary	8.2	7.5	6.6
Czechoslovakia	7.8	7.2	5.6
Denmark	31.2	19.8	17.8
Greece	26.3	19.3	16.2
France	46.8	38.7	35.1
Japan	0.0	0.0	0.0
China	4.8	4.5	4.5
Portugal	29.6	24.2	0.8
Spain	33.4	15.6	1.9

Notes:
This table presents the percentage of names that are common to Origin and American names, depending on whether we used TOP 1000, TOP 200 or TOP 100 from the US Social Security Administration database (`http://www.ssa.gov/OACT/babynames/`). The most popular given names for male babies born during 1900-1909 take into account the number of occurrences of all names from Social Security card applications for births that occurred in the United States.

Table A.4: Male Immigrant Names by Type (Pooled Sample)

	US name	Common name (top 100)	Origin name	Missing	Observations
Germany	35.1	38.3	18.6	8.0	103245
Italy	60.9	0.6	23.6	15.0	43724
Former USSR	82.4	3.0	2.8	11.8	34751
Poland	77.6	6.3	5.5	10.6	24335
Sweden	48.1	19.4	21.1	11.4	27317
Mexico	29.1	2.4	38.1	30.5	10698
Norway	36.4	14.1	33.4	16.1	16105
Hungary	78.0	6.7	4.6	10.7	10368
Czechoslovakia	82.6	5.7	4.7	7.0	8836
Denmark	42.6	17.8	25.3	14.3	7961
Greece	51.6	16.5	17.0	14.9	3780
France	33.3	35.3	18.4	13.0	5135
Japan	6.7	0.0	66.7	26.6	3580
China	18.7	4.5	55.6	21.1	5566
Portugal	58.5	1.5	34.4	5.6	2963
Spain	34.3	2.4	45.9	17.4	1176
All Sample	54.1	14.7	18.6	12.6	309540

Notes:
Pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%).
The US names include a person's first name if it is in the US Social Security database and it is not present in the list of Origin names. In addition, it includes the British names included in Other (British). The Common names include all the names included in both the US Social Security (top 100) and Origin list of names. The Origin names include those names in the Origin list that are not in the US Social Security database (top 100) and the names included in Other (Group). The Other (Group) names include a person's first name if it belongs to their group of countries: Latin (Portugal, Spain, France and Italy), Scandinavian (Sweden, Norway and Denmark), Eastern European (Czechoslovakia, Hungary, Poland and former USSR) and Asian (Japan and China). For Greeks and Mexicans, we used the Latin group; for Germany, we used the Scandinavian group. The Missing names include those names that were impossible to classify in this way, the Other names that were not from the group of origin and missing names.

Table A.5: Some Examples of Name Classification - Males

	Germany		Italy		Former USSR	
	Name	Rank*	Name	Rank*	Name	Rank*
American name						
US (not Common) - top 100	JOHN	1	GEORGE	4	WILLIAM	2
Common top 100	FRANK	8	LEO	44	ROBERT	6
US (not Common) - top 100-1000	STEVEN	367	VINCENT	114	SIMON	221
Non-American name						
Common (top 100-1000)	OTTO	125	ANTONIO	205	IVAN	210
Origin (not Common)	CLAUS	-	DOMENICO	-	VLADIMIR	-

Source: IPUMS.
Notes:
* Rank of the US Social Security database.
The names presented in the table were chosen randomly, i.e., without any specific order.

B Appendix Tables

In this part of the appendix, we provide additional tables for more details. Table B.1 presents the distribution by country of origin of male immigrants, for the sample used in the paper. Variable descriptions are presented in detail in Table B.2. Table B.3 presents the summary statistics for the pooled sample and for Italians, Germans and immigrants from the former USSR, for the period 1900-1920. This table shows only the results for male immigrants. Table B.4 shows the summary statistics for the period 1900-1930, The results are in line with Table B.3.

Table B.5 gives the summary statistics of assimilation outcomes for the sample as well as for three most represented groups of immigrants by birthplace. Table B.6 presents the summary statistics for the children sample (sons and daughters). Recall that in Table 6 in the main text, we show the regression results for male immigrants for all assimilation outcomes including as controls the census year, immigrant birthplace, cohort of entry, number of years in the US, age and log of population in the county of residence, and county fixed effects. In Tables B.7 – B.11, we present the estimation results in different tables, one for each outcome, with different specifications. Tables B.12 – B.17 present estimation results for children's name choice for the subsamples by father's nationality for fathers from Germany, Italy and the former USSR, respectively.

Table B.1: Male Immigrants' Birthplace - Sample Distribution

Country	Observations*	Percentage**
Germany	103245	24.9
Italy	43724	17.5
Former USSR	34751	14.1
Poland	24335	9.0
Sweden	27317	7.7
Mexico	10698	4.7
Norway	16105	4.6
Hungary	10368	3.9
Czechoslovakia	8836	3.1
Denmark	7961	2.4
Greece	3780	2.0
France	5135	1.5
Japan	3580	1.5
China	5566	1.4
Portugal	2963	1.1
Spain	1176	0.6
Total***	309540	100.0

Notes:
* IPUMS pooled sample - unweighted by census sample
** IPUMS pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%).
*** Total represents 63% of total immigrants. The sample distribution of males between 16 and 65 years old is very similar and that age group represents 88% of total male immigrants.

Table B.2: Other Characteristics - Definition

Explanatory variable	Description
Current Census Variables	
Years in the US	Reports how long a person who was born in a foreign country or US outlying area had been living in the United States
Age	Reports the person's age in years
Log population	Reports the log of sample population for geographical unit (county)
Previous Census Variables	
Economic Variables (by geographical unit)	
Immigr. unemp. rate$_{-1}$	Immigrants' unemployment rate by geographical unit (county)
Own immigr. unemp. rate$_{-1}$	Unemployment rate for immigrants with the same birthplace by geographical unit (county)
Exc. own immigr. unemp. rate$_{-1}$	Unemployment rate for immigrants excluding those with the same birthplace by geographical unit (county)
Native unemp. rate$_{-1}$	US-born unemployment rate by geographical unit (county)
Immigr. log occ. score$_{-1}$	Immigrants' average log of occupational score by geographical unit (county)
Own immigr. log occ. score$_{-1}$	Average log of occupational score for immigrants with the same birthplace by geographical unit (county)
Exc. own immigr. log occ. score$_{-1}$	Average log of occupational score for immigrants excluding those with the same birthplace by geographical unit (county)
Native log occ. score$_{-1}$	US-born average log of occupational score by geographical unit (county)
Network Variables (by geographical unit)	
Concentration index$_{-1}$	100 times the number of persons between 16 and 65 in the county who were born in the given immigrant's native country divided by the number of persons in the county
Log immigr. population$_{-1}$	Log of number of male immigrants between 16 and 65 by geographical unit (county)
Own immigr. literacy percentage$_{-1}$	Percentage of male immigrants with the same birthplace (between 16 and 65) being literate by geographical unit (county)
Own immigr. American name percentage$_{-1}$	Percentage of male immigrants with the same birthplace (between 16 and 65) with an American name by geographical unit (county)
Other Outcome Variables	
Log occupational score	Log of occupational score
Full-year employed last year	Dichotomous variable indicating whether the immigrant was employed last year.
American wife/husband	Dichotomous variable indicating whether the immigrant married a woman/man born in the US excluding second generation (those women/men with immigrant parents)
Speaks English	Dichotomous variable indicating whether the immigrant speaks English
US citizenship	Dichotomous variable indicating whether the immigrant has US citizenship

Note:
COUNTY identifies the county where the household was enumerated, using the Inter-University Consortium for Political and Social Research (ICPSR) coding scheme. An example of a county is Los Angeles, which belongs to the State of California. County is available for all 1850-1930 samples. For further details, see IPUMS website (http://usa.ipums.org/usa/).

Table B.3: Summary Statistics - Male Immigrants (1900-1920)

Variable	Mean (standard deviation)			
	Pooled sample	Germany	Italy	Former USSR
Explanatory Variables				
Years in the US	18.0	24.8	12.5	12.4
	(12.1)	(12.7)	(8.7)	(8.6)
Age	38.5	43.5	34.5	33.6
	(12.5)	(12.5)	(11.1)	(11.0)
Log population	9.14	8.82	9.55	9.96
	(1.80)	(1.77)	(1.56)	(1.51)
Economic Variables (by geographical unit)				
Immigr. unemp. rate$_{-1}$	18.5	15.8	21.2	21.9
	(10.6)	(10.6)	(9.1)	(9.7)
Native unemp. rate$_{-1}$	14.8	13.2	16.4	15.7
	(7.4)	(7.4)	(6.3)	(6.0)
Number of observations	*174509*	*70962*	*21458*	*18826*
Own immigr. unemp. rate$_{-1}$	19.5	13.9	23.6	23.1
	(18.5)	(11.9)	(17.2)	(16.3)
Exc. own immigr. unemp. rate$_{-1}$	16.2	15.8	17.9	19.5
	(12.9)	(13.9)	(11.8)	(11.4)
Number of observations	*171533*	*70220*	*21202*	*18591*
Immigr. log occ. score$_{-1}$	3.08	3.06	3.15	3.16
	(0.18)	(0.18)	(0.10)	(0.11)
Own immigr. log occ. score$_{-1}$	3.07	3.07	3.09	3.19
	(0.23)	(0.21)	(0.14)	(0.19)
Exc. own immigr. log occ. score$_{-1}$	3.07	3.03	3.15	3.15
	(0.19)	(0.21)	(0.13)	(0.11)
Native log occ. score$_{-1}$	3.12	3.07	3.19	3.21
	(0.20)	(0.21)	(0.13)	(0.13)
Network Variables (by geographical unit)				
Concentration index$_{-1}$	7.9	12.3	5.0	5.5
	(9.1)	(8.0)	(4.2)	(6.6)
Log immigr. population$_{-1}$	7.49	7.11	7.93	8.52
	(1.94)	(1.99)	(1.74)	(1.66)
Own immigr. literacy percentage$_{-1}$	85.7	96.9	65.8	83.4
	(18.7)	(4.1)	(17.7)	(13.2)
Own immigr. American name percentage$_{-1}$	77.0	79.6	69.3	91.7
	(20.6)	(9.7)	(17.5)	(8.2)
Number of observations	*174509*	*70962*	*21458*	*18826*

Notes:
Pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920 (1%).
Geographical unit - county. For further detail, see note from Table B.2.

Table B.4: Summary Statistics - Male Immigrants (1900-1930)

Variable	Mean (standard deviation)			
	Pooled sample	Germany	Italy	Former USSR
Explanatory Variables				
Years in the US	19.4	25.1	15.7	15.3
	(12.2)	(13.5)	(9.8)	(9.9)
Age	39.5	43.8	36.8	35.7
	(12.3)	(12.7)	(11.4)	(11.6)
Log population	9.43	9.01	9.82	10.18
	(1.79)	(1.80)	(1.51)	(1.50)
Number of observations	*210612*	*76221*	*29986*	*24031*
Economic Variables (by geographical unit)				
Immigr. unemp. rate$_{-1}$	18.5	15.8	21.2	21.9
	(10.6)	(10.6)	(9.1)	(9.7)
Native unemp. rate$_{-1}$	14.8	13.2	16.4	15.7
	(7.4)	(7.4)	(6.3)	(6.0)
Number of observations	*174509*	*70962*	*21458*	*18826*
Own immigr. unemp. rate$_{-1}$	19.5	13.9	23.6	23.1
	(18.5)	(11.9)	(17.2)	(16.3)
Exc. own immigr. unemp. rate$_{-1}$	16.2	15.8	17.9	19.5
	(12.9)	(13.9)	(11.8)	(11.4)
Number of observations	*171533*	*70220*	*21202*	*18591*
Immigr. log occ. score$_{-1}$	3.10	3.07	3.16	3.17
	(0.17)	(0.18)	(0.11)	(0.12)
Own immigr. log occ. score$_{-1}$	3.09	3.09	3.11	3.20
	(0.23)	(0.22)	(0.14)	(0.19)
Exc. own immigr. log occ. score$_{-1}$	3.10	3.05	3.16	3.16
	(0.19)	(0.21)	(0.13)	(0.12)
Native log occ. score$_{-1}$	3.15	3.09	3.21	3.22
	(0.19)	(0.21)	(0.12)	(0.13)
Network Variables (by geographical unit)				
Concentration index$_{-1}$	7.4	11.0	5.8	6.4
	(8.4)	(8.0)	(4.2)	(6.7)
Log immigr. population$_{-1}$	7.55	7.13	7.91	8.48
	(1.91)	(2.00)	(1.69)	(1.65)
Own immigr. literacy percentage$_{-1}$	85.6	97.1	69.4	85.2
	(17.9)	(4.3)	(16.4)	(12.5)
Own immigr. American name percentage$_{-1}$	77.9	79.5	72.4	92.5
	(20.5)	(10.7)	(15.9)	(7.7)
Number of observations	*210612*	*76221*	*29986*	*24031*

Notes:
Pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%).
Geographical unit - county. For further detail, see note from Table B.2.

Table B.5: Summary Statistics - Male Immigrants - Assimilation Outcomes by Country of Birth

Variables	Mean (standard deviation)							
	Sample		Germany		Italy		Former USSR	
	American name	Non-American name	American name	Non-American name	American name	Non-American name	American name	Non-American name
Log occupational score	3.118	3.025	3.100	3.105	3.123	3.103	3.226	3.149
	(0.384)	(0.428)	(0.413)	(0.425)	(0.319)	(0.320)	(0.380)	(0.385)
Full-year employed last year (%)	82.8	83.6	85.0	86.1	80.5	78.4	81.3	80.7
	(37.7)	(37.0)	(35.7)	(34.6)	(39.6)	(41.2)	(39.0)	(39.5)
American wife (%)	5.2	5.2	9.3	8.2	3.4	1.3	1.3	0.7
	(22.2)	(22.3)	(29.0)	(27.4)	(18.2)	(11.5)	(11.3)	(8.5)
Speaks English (%)	86.5	79.0	95.4	94.8	79.8	71.1	83.7	75.2
	(34.5)	(40.7)	(21.1)	(22.3)	(40.1)	(47.3)	(36.9)	(43.2)
US citizenship (%)	53.8	44.7	76.0	72.9	40.0	34.7	46.6	39.6
	(49.9)	(49.7)	(42.7)	(44.5)	(49.0)	(47.6)	(49.9)	(48.9)

Note:
Pooled sample - weighted by census sample size: 1900(5%) and 1910, 1920, 1930 (1%).

Table B.6: Summary Statistics - Child Immigrants

Variables	Mean (standard deviation)	
	Sons	Daughters
Explanatory Variables		
Father with American Name	78.0	78.2
	(41.4)	(41.3)
Immigrant mother	79.1	80.6
	(40.7)	(39.5)
Siblings (aged 0-10)	2.2	2.2
	(1.4)	(1.4)
Child's age	5.0	5.0
	(3.1)	(3.1)
Child US born	95.3	94.0
	(21.1)	(23.7)
Years in the US (father)	18.5	18.1
	(9.4)	(9.2)
Father's age	39.1	39.0
	(7.9)	(7.8)
Log population	9.09	9.19
	(1.79)	(1.76)
Number of observations	*74724*	*70309*

Notes:
Pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920 (1%).

Table B.7: Log Occupational Score - Male Immigrants

	(1)	(2)	(3)
Total			
American name	0.091	0.014	0.005
	(0.007)***	(0.003)***	(0.002)*
Observations	229415	229415	229415
Germany			
American name	-0.003	-0.001	-0.007
	(0.006)	(0.005)	(0.004)*
Observations	75738	75738	75738
Italy			
American name	0.015	0.013	0.008
	(0.005)***	(0.004)***	(0.004)**
Observations	32083	32083	32083
Former USSR			
American name	0.066	0.043	0.027
	(0.011)***	(0.010)***	(0.008)***
Observations	26269	26269	26269
Controls			
Year	✓	✓	✓
Birthplace	×	✓	✓
Years in the US	×	✓	✓
Age	×	✓	✓
Log population	×	✓	✓
Cohort of entry	×	✓	✓
County	×	×	✓

Notes:
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
Years in the US, and age enter in quadratic form.

Table B.8: Full-year Employed Last Year - Male Immigrants

	(1)	(2)	(3)
Total			
American name	-0.009	-0.005	-0.001
	(0.004)**	(0.003)***	(0.003)
Observations	155910	155910	155910
Germany			
American name	-0.010	-0.013	-0.010
	(0.004)**	(0.004)***	(0.004)**
Observations	65184	65184	65184
Italy			
American name	0.018	0.009	0.005
	(0.010)*	(0.010)	(0.010)
Observations	16807	16807	16807
Former USSR			
American name	0.005	0.007	0.014
	(0.014)	(0.013)	(0.014)
Observations	15255	15255	15255
Controls			
Year	✓	✓	✓
Birthplace	×	✓	✓
Years in the US	×	✓	✓
Age	×	✓	✓
Log population	×	✓	✓
Cohort of entry	×	✓	✓
County	×	×	✓

Notes:
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
Years in the US, and age enter in quadratic form.

Table B.9: American Wife (excluding Second Generation) - Male Immigrants

	(1)	(2)	(3)
Total			
American name	-0.001	0.009	0.006
	(0.002)	(0.002)***	(0.002)***
Observations	147282	147282	147282
Germany			
American name	0.012	0.008	0.001
	(0.004)***	(0.004)**	(0.004)
Observations	55940	55940	55940
Italy			
American name	0.020	0.015	0.013
	(0.003)***	(0.003)***	(0.002)***
Observations	18735	18735	18735
Former USSR			
American name	0.005	0.005	0.003
	(0.003)	(0.003)	(0.003)
Observations	17422	17422	17422
Controls			
Year	✓	✓	✓
Birthplace	×	✓	✓
Years in the US	×	✓	✓
Age	×	✓	✓
Log population	×	✓	✓
Cohort of entry	×	✓	✓
County	×	×	✓

Notes:
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
Years in the US, and age enter in quadratic form.

Table B.10: Speaks English - Male Immigrants

	(1)	(2)	(3)
Total			
American name	0.074	0.029	0.026
	(0.008)***	(0.002)***	(0.002)***
Observations	238619	238619	238619
Germany			
American name	0.007	0.003	0.002
	(0.003)**	(0.003)	(0.003)
Observations	79265	79265	79265
Italy			
American name	0.061	0.034	0.031
	(0.006)***	(0.007)***	(0.007)***
Observations	33219	33219	33219
Former USSR			
American name	0.064	0.041	0.032
	(0.011)***	(0.010)***	(0.011)***
Observations	27485	27485	27485
Controls			
Year	✓	✓	✓
Birthplace	×	✓	✓
Years in the US	×	✓	✓
Age	×	✓	✓
Log population	×	✓	✓
Cohort of entry	×	✓	✓
County	×	×	✓

Notes:
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
Years in the US, and age enter in quadratic form.

Table B.11: US Citizenship - Male Immigrants

	(1)	(2)	(3)
Total			
American name	0.095	0.012	0.010
	(0.011)***	(0.003)***	(0.002)***
Observations	229051	229051	229051
Germany			
American name	0.029	0.009	0.005
	(0.005)***	(0.004)**	(0.005)
Observations	76555	76555	76555
Italy			
American name	0.040	0.010	0.003
	(0.007)***	(0.006)	(0.006)
Observations	31739	31739	31739
Former USSR			
American name	0.050	0.028	0.029
	(0.012)***	(0.010)***	(0.010)***
Observations	25593	25593	25593
Controls			
Year	✓	✓	✓
Birthplace	×	✓	✓
Years in the US	×	✓	✓
Age	×	✓	✓
Log population	×	✓	✓
Cohort of entry	×	✓	✓
County	×	×	✓

Notes:
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
Years in the US, and age enter in quadratic form.

Table B.12: American Name (Probit - Marginal Effects) - Sons (German Father)

	(1)	(2)	(3)	(4)
Father with American name	-	-	5.355	5.356
	-	-	(0.658)***	(0.659)***
Immigrant mother	-	-	-	-0.465
	-	-	-	(0.667)
Siblings (aged 0-10)	-	-	-	-0.030
	-	-	-	(0.211)
Child's age	-	0.219	0.221	0.222
	-	(0.099)**	(0.100)**	(0.101)**
Child US born	-	6.475	6.510	6.493
	-	(1.820)***	(1.861)***	(1.858)***
Years in the US^{+}	0.259	0.220	0.227	0.221
	(0.152)*	(0.154)	(0.156)	(0.158)
Age^{+}	-0.057	-0.075	-0.072	-0.064
	(0.041)	(0.042)*	(0.043)*	(0.043)
Log population	-1.955	-1.862	-0.967	-1.008
	(2.628)	(2.613)	(2.638)	(2.638)
Immigr. unemp. rate$_{-1}$	-0.088	-0.084	-0.081	-0.080
	(0.064)	(0.064)	(0.063)	(0.063)
Native unemp. rate$_{-1}$	0.076	0.072	0.074	0.074
	(0.102)	(0.101)	(0.100)	(0.100)
Immigr. log occ. score$_{-1}$	-3.189	-3.099	-3.013	-2.982
	(7.747)	(7.735)	(7.630)	(7.624)
Native log occ. score$_{-1}$	1.387	1.446	1.078	1.115
	(6.677)	(6.678)	(6.721)	(6.717)
Concentration index$_{-1}{}^{+}$	0.014	0.018	0.016	0.016
	(0.130)	(0.130)	(0.131)	(0.131)
Log immigr. population$_{-1}$	-0.242	-0.263	-0.260	-0.235
	(1.029)	(1.031)	(1.035)	(1.034)
Own immigr. literacy percentage$_{-1}$	3.362	2.672	1.043	1.054
	(10.344)	(10.376)	(10.484)	(10.489)
Own immigr. American name percentage$_{-1}$	-0.025	-0.025	-0.016	-0.016
	(0.052)	(0.052)	(0.051)	(0.051)
1910	-1.342	-1.139	-0.652	-0.757
	(2.972)	(2.967)	(2.987)	(2.965)
1920	1.999	1.770	1.071	0.971
	(3.568)	(3.584)	(3.645)	(3.612)
Observations	31171	31171	31171	31171
Other controls (father)	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County

Notes:

The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).

Robust standard errors in parentheses clustered by county.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

+ Years in the US, age and concentration index are always included in quadratic form.

Table B.13: American Name (Probit - Marginal Effects) - Sons (Italian Father)

	(1)	(2)	(3)	(4)
Father with American name	-	-	7.647	7.484
	-	-	(0.813)***	(0.818)***
Immigrant mother	-	-	-	-7.033
	-	-	-	(2.149)***
Siblings (aged 0-10)	-	-	-	0.091
	-	-	-	(0.445)
Child's age	-	0.128	0.108	0.105
	-	(0.179)	(0.180)	(0.174)
Child US born	-	1.061	1.587	1.520
	-	(1.607)	(1.578)	(1.553)
Years in the US^{+}	0.524	0.508	0.461	0.422
	(0.323)	(0.326)	(0.332)	(0.340)
Age^{+}	-0.044	-0.059	-0.055	-0.003
	(0.063)	(0.067)	(0.069)	(0.069)
Log population	-12.869	-12.899	-11.549	-11.953
	(5.010)***	(4.956)***	(4.927)**	(5.017)**
Immigr. unemp. rate$_{-1}$	-0.350	-0.354	-0.364	-0.358
	(0.129)***	(0.129)***	(0.126)***	(0.125)***
Native unemp. rate$_{-1}$	0.839	0.846	0.845	0.840
	(0.212)***	(0.210)***	(0.206)***	(0.206)***
Immigr. log occ. score$_{-1}$	-4.312	-4.088	-5.726	-5.419
	(21.167)	(21.229)	(20.914)	(20.894)
Native log occ. score$_{-1}$	26.872	26.824	28.266	28.170
	(18.651)	(18.597)	(18.044)	(17.883)
Concentration index$_{-1}{}^{+}$	-0.083	-0.090	-0.062	-0.065
	(0.371)	(0.370)	(0.354)	(0.354)
Log immigr. population$_{-1}$	-2.400	-2.355	-2.060	-1.874
	(3.106)	(3.100)	(3.072)	(3.117)
Own immigr. literacy percentage$_{-1}$	1.749	1.765	1.282	0.946
	(4.724)	(4.725)	(4.663)	(4.661)
Own immigr. American name percentage$_{-1}$	-0.084	-0.085	-0.077	-0.073
	(0.045)*	(0.045)*	(0.047)	(0.047)
1910	-10.599	-10.738	-9.708	-10.127
	(5.977)*	(5.984)*	(6.012)	(5.952)*
1920	2.396	2.267	1.257	1.030
	(6.893)	(6.879)	(7.013)	(7.056)
Observations	8651	8651	8651	8651
Other controls (father)	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County

Notes:

The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).

Robust standard errors in parentheses clustered by county.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

+ Years in the US, age and concentration index are always included in quadratic form.

Table B.14: American Name (Probit - Marginal Effects) - Sons (Former USSR Father)

	(1)	(2)	(3)	(4)
Father with American name	-	-	5.096	5.025
	-	-	(1.719)***	(1.715)***
Immigrant mother	-	-	-	1.099
	-	-	-	(1.804)
Siblings (aged 0-10)	-	-	-	-0.278
	-	-	-	(0.372)
Child's age	-	-0.346	-0.348	-0.352
	-	(0.189)*	(0.189)*	(0.187)*
Child US born	-	-1.678	-1.810	-1.855
	-	(2.032)	(2.010)	(1.998)
Years in the US^{+}	-0.271	-0.203	-0.198	-0.177
	(0.275)	(0.265)	(0.263)	(0.263)
Age^{+}	-0.128	-0.081	-0.077	-0.077
	(0.088)	(0.099)	(0.098)	(0.102)
Log population	-1.394	-1.304	-1.574	-1.559
	(7.607)	(7.606)	(7.482)	(7.465)
Immigr. unemp. rate$_{-1}$	-0.103	-0.106	-0.116	-0.123
	(0.162)	(0.162)	(0.161)	(0.161)
Native unemp. rate$_{-1}$	0.045	0.050	0.067	0.076
	(0.295)	(0.296)	(0.297)	(0.298)
Immigr. log occ. score$_{-1}$	24.488	24.529	24.935	24.626
	(26.315)	(26.385)	(26.365)	(26.585)
Native log occ. score$_{-1}$	-21.355	-20.778	-20.835	-21.487
	(15.909)	(16.078)	(16.006)	(15.802)
Concentration index$_{-1}{}^{+}$	0.464	0.455	0.415	0.409
	(0.211)**	(0.213)**	(0.212)*	(0.214)*
Log immigr. population$_{-1}$	4.453	4.361	4.421	4.540
	(3.459)	(3.472)	(3.492)	(3.488)
Own immigr. literacy percentage$_{-1}$	-17.212	-17.096	-16.621	-16.632
	(6.836)**	(6.757)**	(6.608)**	(6.597)**
Own immigr. American name percentage$_{-1}$	-0.137	-0.135	-0.123	-0.123
	(0.069)**	(0.070)*	(0.069)*	(0.069)*
1910	0.0004	-0.024	-0.367	-0.602
	(7.489)	(7.497)	(7.502)	(7.579)
1920	2.996	2.807	2.845	2.541
	(7.004)	(6.981)	(6.898)	(6.844)
Observations	8219	8219	8219	8219
Other controls (father)	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County

Notes:

The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).

Robust standard errors in parentheses clustered by county.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

+ Years in the US, age and concentration index are always included in quadratic form.

Table B.15: American Name (Probit - Marginal Effects) - Daughters (German Father)

	(1)	(2)	(3)	(4)
Father with American name	-	-	0.103	0.103
	-	-	(0.372)	(0.373)
Immigrant mother	-	-	-	0.061
	-	-	-	(0.496)
Siblings (aged 0-10)	-	-	-	-0.067
	-	-	-	(0.139)
Child's age	-	0.081	0.079	0.078
	-	(0.058)	(0.058)	(0.057)
Child US born	-	40.040	39.944	40.044
	-	(0.866)***	(2.333)***	(2.344)***
Years in the US^{+}	0.012	0.014	0.015	0.015
	(0.080)	(0.080)	(0.080)	(0.079)
Age^{+}	0.007	0.005	0.004	0.005
	(0.025)	(0.025)	(0.026)	(0.026)
Log population	-1.294	-1.359	-1.416	-1.358
	(1.394)	(1.402)	(1.399)	(1.399)
Immigr. unemp. rate$_{-1}$	-0.010	-0.011	-0.010	-0.010
	(0.043)	(0.043)	(0.043)	(0.043)
Native unemp. rate$_{-1}$	-0.062	-0.064	-0.065	-0.065
	(0.066)	(0.066)	(0.066)	(0.066)
Immigr. log occ. score$_{-1}$	-1.269	-1.306	-1.204	-1.217
	(5.389)	(5.397)	(5.401)	(5.405)
Native log occ. score$_{-1}$	2.627	2.520	2.594	2.508
	(4.277)	(4.276)	(4.269)	(4.271)
Concentration index$_{-1}{}^{+}$	0.153	0.121	0.125	0.122
	(0.079)	(0.079)	(0.079)	(0.079)
Log immigr. population$_{-1}$	-0.851	-0.785	-0.713	-0.788
	(0.590)	(0.589)	(0.588)	(0.588)
Own immigr. literacy percentage$_{-1}$	-19.289	-24.274	-24.502	-24.251
	(6.960)***	(6.939)***	(6.872)***	(6.869)***
Own immigr. American name percentage$_{-1}$	0.070	0.070	0.071	0.070
	(0.040)*	(0.040)*	(0.040)*	(0.040)*
1910	1.134	1.336	1.524	1.314
	(1.590)	(1.593)	(1.586)	(1.598)
1920	3.015	3.405	3.064	3.376
	(2.137)	(2.127)	(2.103)	(2.122)
Observations	27056	27056	27056	27056
Other controls (father)	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County

Notes:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
+ Years in the US, age and concentration index are always included in quadratic form.

Table B.16: American Name (Probit - Marginal Effects) - Daughters (Italian Father)

	(1)	(2)	(3)	(4)
Father with American name	-	-	4.075	4.033
	-	-	(0.54)***	(0.542)***
Immigrant mother	-	-	-	-1.405
	-	-	-	(1.090)
Siblings (aged 0-10)	-	-	-	-0.021
	-	-	-	(0.205)
Child's age	-	0.012	0.019	0.021
	-	(0.116)	(0.116)	(0.116)
Child US born	-	53.319	53.513	53.428
	-	(2.991)***	(2.931)***	(2.925)***
Years in the US^{+}	1.508	-0.054	-0.054	-0.063
	(0.288)***	(0.215)	(0.221)	(0.221)
Age^{+}	-0.557	-0.074	-0.053	-0.041
	(0.058)***	(0.045)*	(0.044)	(0.048)
Log population	-14.065	-3.110	-2.342	-2.461
	(4.484)***	(3.236)	(3.101)	(3.134)
Immigr. unemp. rate$_{-1}$	-0.109	-0.174	-0.155	-0.154
	(0.110)	(0.093)*	(0.091)*	(0.091)*
Native unemp. rate$_{-1}$	-0.065	0.128	0.118	0.113
	(0.234)	(0.176)	(0.170)	(0.170)
Immigr. log occ. score$_{-1}$	19.364	15.790	14.736	15.557
	(14.352)	(10.394)	(10.158)	(10.179)
Native log occ. score$_{-1}$	-34.045	-24.057	-22.382	-23.012
	(14.497)**	(10.617)**	(10.492)**	(10.463)**
Concentration index$_{-1}{}^{+}$	-0.255	0.017	0.011	0.015
	(0.395)	(0.344)	(0.338)	(0.341)
Log immigr. population$_{-1}$	6.179	2.297	2.405	2.495
	(2.555)**	(1.948)	(1.899)	(1.899)
Own immigr. literacy percentage$_{-1}$	-6.403	-1.276	-2.010	-2.032
	(3.674)*	(2.698)	(2.622)	(2.619)
Own immigr. American name percentage$_{-1}$	0.043	-0.006	-0.0002	0.0001
	(0.047)	(0.041)	(0.040)	(0.040)
1910	-3.541	1.620	1.632	1.498
	(4.933)	(3.717)	(3.561)	(3.585)
1920	20.644	8.980	7.842	7.835
	(6.971)***	(5.181)*	(5.274)	(5.314)
Observations	8477	8477	8477	8477
Other controls (father)	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County

Notes:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
+ Years in the US, age and concentration index are always included in quadratic form.

Table B.17: American Name (Probit - Marginal Effects) - Daughters (Former USSR Father)

	(1)	(2)	(3)	(4)
Father with American name	-	-	0.563	0.532
	-	-	(0.710)	(0.706)
Immigrant mother	-	-	-	0.669
	-	-	-	(0.983)
Siblings (aged 0-10)	-	-	-	-0.071
	-	-	-	(0.193)
Child's age	-	0.175	0.175	0.174
	-	(0.107)	(0.108)	(0.104)*
Child US born	-	32.555	32.556	32.537
	-	(0.850)***	(0.847)***	(0.842)***
Years in the US^{+}	2.312	-0.070	-0.071	-0.066
	(0.328)***	(0.149)	(0.149)	(0.152)
Age^{+}	-0.884	-0.030	-0.029	-0.030
	(0.049)***	(0.040)	(0.040)	(0.044)
Log population	-9.474	0.201	0.252	0.309
	(4.393)**	(1.761)	(1.756)	(1.736)
Immigr. unemp. rate$_{-1}$	-0.094	-0.058	-0.061	-0.062
	(0.094)	(0.072)	(0.072)	(0.072)
Native unemp. rate$_{-1}$	0.222	0.062	0.066	0.071
	(0.157)	(0.102)	(0.103)	(0.103)
Immigr. log occ. score$_{-1}$	6.198	6.332	6.381	5.991
	(21.080)	(9.760)	(9.805)	(9.767)
Native log occ. score$_{-1}$	29.387	11.147	10.916	10.684
	(16.797)*	(6.587)*	(6.618)*	(6.664)
Concentration index$_{-1}{}^{+}$	-0.317	-0.055	-0.058	-0.057
	(0.205)	(0.097)	(0.097)	(0.096)
Log immigr. population$_{-1}$	-4.977	-3.076	-3.062	-3.098
	(3.122)	(1.948)	(1.940)	(1.919)
Own immigr. literacy percentage$_{-1}$	10.599	4.996	5.088	5.059
	(6.044)*	(2.740)*	(2.731)*	(2.727)*
Own immigr. American name percentage$_{-1}$	-0.030	0.014	0.016	0.016
	(0.065)	(0.044)	(0.043)	(0.042)
1910	-1.9142	2.993	3.030	3.005
	(5.727)	(2.422)	(2.425)	(2.410)
1920	16.149	4.317	4.311	4.208
	(7.035)**	(3.068)	(3.060)	(3.026)
Observations	8217	8217	8217	8217
Other controls (father)	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County	Birthplace Cohort of entry County

Notes:
The table reports marginal effects multiplied by 100 (hence, in terms of percentage points).
Robust standard errors in parentheses clustered by county.
* Significant at 10%; ** significant at 5%; *** significant at 1%.
+ Years in the US, age and concentration index are always included in quadratic form.