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

Pedro Carneiro[a,b,c], Sokbae Lee[d,b,c], Hugo Reis[e,f]

[a] *University College London*
[b] *Institute for Fiscal Studies*
[c] *Centre for Microdata Methods and Practice*
[d] *Columbia University*
[e] *Banco de Portugal*
[f] *Católica Lisbon School of Business and Economics*

Abstract

The majority of immigrants to the United States at the beginning of the 20th century adopted American first names. In this paper we study the economic determinants of name choice, by relating the propensity of immigrants to carry an American first name to the local concentration of their compatriots and local labor market conditions. We find that high concentrations of immigrants of a given nationality discouraged members of that nationality from taking American names, in particular for more recent arrivals. In contrast, labor market conditions for immigrants do not seem to be associated with more frequent name changes among immigrants.

Keywords: Americanization, culture, first name, identity, immigration
JEL: code, J15, N32

⋆We would like to thank the Co-Editor, Albrecht Glitz, two anonymous referees, Christian Dustmann and Imran Rasul for their helpful comments, and Lucena Vieira for very competent research assistance. 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, ERC-2014-CoG-646917-ROMIA, and ERC-2015-CoG-682349).

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, which may not be obvious, but it turns out to be empirically important: the adoption of American first names. Using data from US Censuses collected in the first half of the 20th century,[1] Abramitzky et al. (2017), Biavaschi et al. (2017), and Goldstein and Stecklov (2016) have used a variety of empirical strategies to show that the adoption of American sounding first names led to substantial improvements in labor market outcomes of first and second generation immigrants. Their conjecture is that this is due to a more successful assimilation process by those with American names.

In spite of the benefits to adopting an American name, 20 to 30% of all immigrants in the 1900-1930 Censuses decided to keep a foreign name. This proportion varied considerably across locations, even within groups of immigrants with the same nationality, suggesting that there exists local variation in the costs and benefits of adopting an American name.[2] In this paper we examine how local labor market conditions and the local concentration of immigrants of different nationalities affect this decision. More generally, we seek to understand how different economic and social factors accelerate or hinder the process of immigrant assimilation.

As in Watkins and London (1994) and Lazear (1999), we start by ex-

[1] There is a substantial body of the literature on international migration to the US at the turn of the 20th century. See, for example, Hatton and Williamson (1998), Abramitzky et al. (2012), Abramitzky et al. (2013), Abramitzky et al. (2014) or Bandiera et al. (2013)), among others.

[2] For example, around 16% of all Immigrants decided to keep their foreign name in Pennsylvania, while in California and Texas the number increases to 42% and 52%, respectively. By immigrants' birthplace we observe that, for instance, among German immigrants about 20% keep their name in Pennsylvania and 30% in Texas. Regarding Mexican immigrants, 39% and 64% keep their origin name in Pennsylvania and Texas, respectively.

amining 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 defined as 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 (multiplied by 100), so it is a measure of the extent to which a random immigrant is able to find other immigrants with the same origin residing in the same location as him.[3] However, we go much beyond Watkins and London's (1994) and Lazear's (1999) cross-sectional analysis, by relating the propensity of immigrants to carry an American first name to the local concentration of their compatriots.

Our idea is instead to exploit within county variation to estimate the impact of local immigrant concentrations on naming decisions. This is important since the estimates from purely cross sectional analyses are more prone to contamination by local unobservables which drive both name adoption and local conditions, than estimates relating changes in name adoption to changes in local social and economic conditions over time. 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. This effect is particularly strong for immigrants who have been in the US for less than 5 years.

In addition, a novelty of our paper is to relate the propensity of immigrants to carry an American first name to the local labor market conditions of their compatriots. Local labor market conditions for immigrants could affect the opportunities for trading both with other immigrants and with natives. We do not find strong impacts of the local labor market measures for immigrants on adoption of American names.

We also try to understand to what extent immigrants with different years in the US responded differently to incentives when deciding to adopt American Names. To gain insight into the role of assimilation, we split the sample

[3]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.

by years since arrival and examine the heterogeneity. The results suggest that the longer an immigrant is in the US, the weaker seems to be the average impact of network (concentration index) on American name adoption, while there is not much of a pattern with regards to economic variables. In fact, it is interesting that the impact of the concentration index is very strong early on after arrival.

Finally, we examine the heterogeneity of the main results by the birthplace of immigrants. The results show that the concentration index has a stronger impact on name adoption in immigrants from Asia and Southern countries, a smaller but also significant effect on immigrants from Nordic countries and Mexico but no effect on name adoption of immigrants from Eastern Europe and Germany.

Although we show results based on the standard indices of name foreignness used, for example, in Abramitzky et al. (2017), and Goldstein and Stecklov (2016), our preferred models use an alternative definition of American name. The definition of American name we use takes all the top 100 baby names in the US Social Security records of the 1900s (independently of the etymological origin), plus all the remaining names in that list (top 101 to top 1000) which are etymologically of American origin, i.e. names that are not from a Latin, Greek, German, Eastern European, Scandinavian, Asian, or other non-American/non-British origin.[4] Our results are qualitatively similar regardless of the definition we use.

According to our definition of what is an American name, we document 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 much less common among immigrants at the time of their arrival (for example, less than 1% for Italians according to arrival records from Ellis Island). In addition, the proportion of individuals from different countries (in the 1900-1930 US censuses) who have one of the top three American names (John, William and George) among immigrants from that country is larger for immigrants from 6 countries than it is for US natives. This is an indication that a significant number of immigrants tend to have more stereotypical American names than natives. The percentage of top three American

[4]For example, the name "Frank" is of German origin but its popularity rank is the 8th among male babies born in the US during 1900s. Hence, it is regarded as American first name. The name "Otto", which was the name of kings in Germany, is ranked at the 125th and so is treated as German name.

names ranges from around 2% for Japan to about 25% for Greece. In contrast, only 15% of US natives held any of the three top names. Consistent with the findings of Abramitzky et al. (2017), Biavaschi et al. (2017), and Goldstein and Stecklov (2016), immigrants with American names (according to our preferred definition) are shown to work in occupations with better occupation scores, to be more likely to have a US born spouse, to be more likely to speak english, and to be more likely to be a US citizen.

It is worthwhile noting that Abramitzky et al. (2017), and Goldstein and Stecklov (2016) focus the bulk of their studies on the naming choice of second generation immigrants, who are all born in the US. The relevant actor in this case is the immigrant parent, who must decide how to name her native born child. We focus instead on the decision of the immigrant to change his own first name after he arrives in the US. Our framework, which exploits variation in local labor market conditions experienced by immigrants residing in different counties, is more appropriate to understand this decision, than to understand the decision of how to name one's child, which probably considers more long run forecasts of the future economic prospects of children. Nevertheless, we also show that the naming choice of second generation immigrants responds to the concentration index for sons (but not for daughters).

Beyond the three papers already referred to (Abramitzky et al. (2017), Biavaschi et al. (2017), and Goldstein and Stecklov (2016)), and regarding name changes in broader contexts, this paper is related to Arai and Thoursie (2009), Algan et al. (2013), 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. In a different context, Algan et al. (2013) studied parental naming decisions between Arabic and non-Arabic first names to newborn babies in France over the 2003-2007 period. Using exogenous allocation of public housings dwellings as an identification strategy, they found evidence for the significant economic factors affecting parental naming choices. 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 et al. (2014), among others).[5] See also Olivetti and Paserman (2015) and Güell et al. (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) and focuses 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 on this issue.

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). It also briefly shows to what extent the adoption of an American name, according to our preferred definition, is associated with better assimilation, as measured by labor market and social outcomes. Section 3 presents a simple economic model of name choice, as a function of labor market and network variables. Section 4 provides estimates of this model, and Section 5 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)).[6] 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 popula-

[5] 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.

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

tion. 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. Defining American Names

There is no unique way to classify names according to how American or foreign they sound. A procedure already used in the literature would be to define an "American Name Index" (ANI) such as the following:

$$\text{ANI}_{name,t} = \frac{\text{Prob}(name|\text{US born}, t)}{\text{Prob}(name|\text{US born}, t) + \text{Prob}(name|\text{Foreign born}, t)} \times 100,$$

where *name* is a particular first name and t reflects the census year. The index ranges from 0 to 100, takes the value 100 if all individuals who have a specific name are US born, and is equal to 0 if only immigrants have a particular name.[7] Abramitzky et al. (2017) construct a very similar foreignness index for each name, equal to

$$\text{Foreignness Index}_{name,t} = \frac{\text{Prob}(name|\text{Foreign born}, t)}{\text{Prob}(name|\text{US born}, t) + \text{Prob}(name|\text{Foreign born}, t)} \times 100,$$

and Goldstein and Stecklov (2016) also construct a similar index for each ethnicity defined by father's birthplace.

Although we present results using ANI, we opted not to use it as our main definition. In order to see why, take for example the name John, which is the most popular name in the US in the early 1900s according to both the Social Security records and the IPUMS, taken by about 6% of all native born in that period. It just happens that John is also an extremely popular name among immigrants, not because they are given this name at birth, but because immigrants often choose to adopt the most popular American names after they arrive in the US. In the 1900-1930 Censuses, around 9% of all immigrants are called John. As a result the average ANI index for John is only 47.4, which is at the 68th percentile of the distribution of the ANI index.

[7] Fryer and Levitt (2004) used such an index in their study of black and white names.

Our main results use a dichotomous classification for each name, which is either American or Foreign, instead of a continuous measure such as the ANI or the Foreignness Index just described. As stated previously, the definition of American name we use takes all the top 100 baby names in the US Social Security records of the 1900s (independently of the etymological origin), plus all the remaining names in that list (top 101 to top 1000) which are etymologically of American origin, i.e. names that are not from a Latin, Greek, German, Eastern European, Scandinavian, Asian, or other non-American/non-British origin. The names classification is described in detail in Online Appendix A.

Nevertheless, we realize that our preferred classification is also not perfect. Therefore we present results with both types of classification. They are qualitatively similar.

2.2. 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[8] includes immigrants, originating from 16 different countries of birth: Germany, Italy, former USSR,[9] Poland, Sweden, Mexico, Norway, Hungary, Czechoslovakia, Denmark, Greece, France, Japan, China, Portugal, and Spain.[10]

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 name across nationalities is always of American/British origin (John, Joseph, Peter and George).[11]

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

[9] 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.

[10] 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.

[11] Our figures in this table and table 3 below are different from those in Biavaschi et al. (2017). The most comparable numbers are the ones where we both use the Census. But

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.

whereas they focus on 1930, we pool together all Censuses between 1900 and 1930. The numbers related to name Americanization are very different because they are based on completely different datasets (Census vs Ancestry.com), concerning very different populations (whole of the US vs New York) and completely different algorithms for defining what is an American name. Still, there is a positive correlation between the figures in our table 3 and the first columns of the first table in the paper we are referring to: immigrants from Russia/USSR and countries in Central Europe have a higher proportion of immigrants with American names than those from countries in Northern or Southern Europe.

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 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 10% for those from Japan, to around 93% for those from the former USSR. For female immigrants, the overall percentage of American names is about 78%, which is very similar to male immigrants' percentage. The heterogeneity across different origins is also similar between males and females.[12]

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 (John, William and George) among immigrants from that country. Notice that this proportion is larger for immigrants from 6 countries than it is for US natives, indicating that a significant number of immigrants tend to have more stereotypical American names than natives. The percentage of top three American names ranges from around 2% for Japan to about 25% for Greece.

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.

[12]Table D.1 in Online Appendix D presents the same statistics for the ANI index.

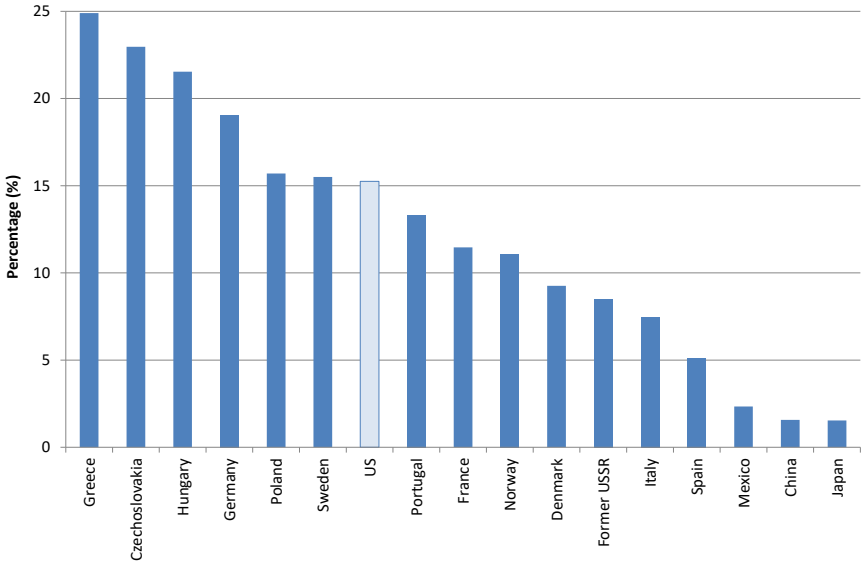
Table 3: Percentage of Female and Male Immigrants with American Names - By Year of Census and Country of Birth

	Female	Male
Pooled Sample	78.2	76.7
By year		
1900	73.2	75.0
1910	77.2	75.8
1920	80.7	78.6
1930	79.9	76.8
By Country of Birth		
Germany	72.1	78.2
Italy	75.0	70.4
Former USSR	92.5	93.1
Poland	90.1	91.1
Sweden	72.6	73.8
Mexico	49.4	41.3
Norway	66.7	58.5
Hungary	92.7	91.7
Czechoslovakia	93.9	92.9
Denmark	71.4	68.2
Greece	72.5	78.2
France	65.8	75.7
Japan	3.7	10.1
China	20.3	25.7
Portugal	81.4	62.2
Spain	42.2	42.5

Notes: Pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%). Immigrants between 16 and 65 years old.

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 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, John, William and George, are reported in Table A.2 - Column (2) from IPUMS.

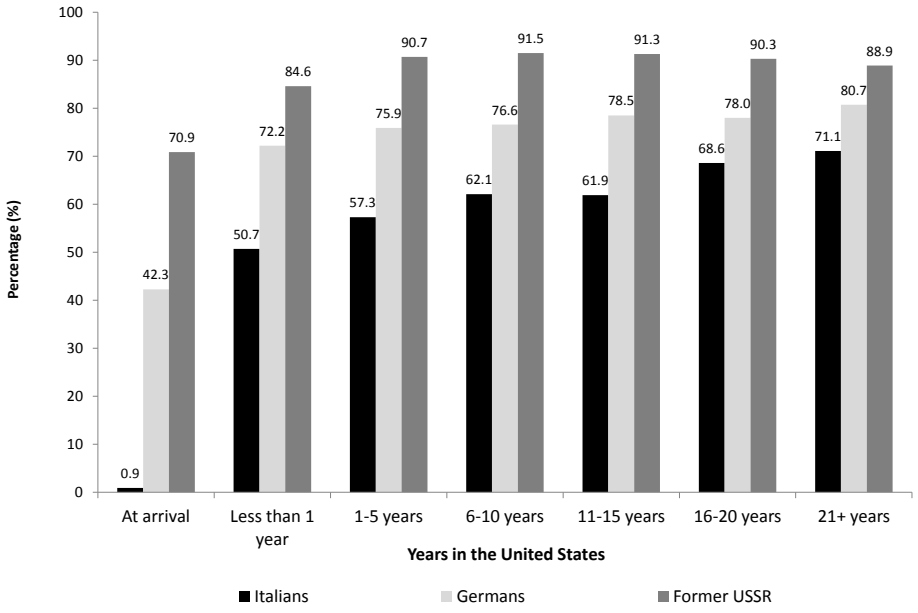
figure rises to around 50% for those in the US for less than one year, and then reaches about 70% for those in the US for more than 20 years.

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).[13] 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

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

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.

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 (National Archives and Records Administration (1977–2002)). 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.

One caveat behind Figure 2 is that since we combine data from different cross-sectional sources, we do not know the magnitudes of the potential biases coming from changes in immigrant cohort quality as well as from selection

due to return migrations. If the quality of immigrants differs by cohort and they return to home countries selectively, then years in the US may not necessarily reflect the assimilation patterns across comparable immigrants.

2.3. 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 C.2 in Online Appendix C 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.[14] 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.[15] All the economic and network variables use county-level data from the contemporaneous census.[16] For example, for a particular immigrant in 1930, $Immigr.\ labor\ force\ participation\ rate$ denotes what the labor force participation rate for immigrants was in 1930 in the individual's 1930 county of residence.

The *occupational score* is a constructed variable that assigns occupational income scores to each occupation. It assigns each occupation in all years a value representing the median total income (in hundreds of 1950 dollars) of all persons with that particular occupation in 1950. This variable thus provides a continuous measure of occupations, according to the economic

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

[15] 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 (url-http://usa.ipums.org/usa/).

[16] In the previous version of the paper, we have used lagged variables to measure economic and network variables. While it may be true that lagged variables are to some extent less prone to endogeneity due to some contemporaneous unobservable shocks, they are also clearly worse measures of the local drivers of contemporaneous name Americanization. We thank the Co-Editor and an anonymous referee for pointing out this issue.

rewards enjoyed by people working at them in 1950. Regarding unemployment, we calculate the *unemployment rate* by 1−(those who were full-year employed)/(those in the labor force). This does not match the modern definition of the unemployment rate which was only established in 1940. See Card (2011) for the detailed historical account of the origin of the modern definition. The *concentration index* is defined as the number of individuals aged 16 to 65 in a county who were born in the given immigrants native country, divided by the total number of individuals residing in the county of residence of each immigrant (multiplied by 100), so it is a measure of the extent to which a random immigrant is able to find other immigrants with the same origin residing in the same location as him.

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 (difference of around one year), to be of the same age 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 more than 6% of the residents were born in their native country. Furthermore, immigrants with an American name tend to live in places where the percentage of immigrants is lower.

Table 4 also presents 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. The noticeable differences are that on average, female immigrants with an American name are younger (almost 2 years difference) and have been in the US for a slightly shorter period (less than half a year).

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: Summary Statistics - American Name vs Non-American Name (1900-1930)

Variables	Mean (standard deviation)				Diff. mean (standard error)	
	Males		Females		Males	Females
	American name	Non-American name	American name	Non-American name		
Years in the US	19.4	18.2	19.6	19.9	1.2***	-0.3***
	(12.2)	(12.3)	(12.1)	(13.0)	(0.1)	(0.1)
Age	39.3	39.3	38.4	40.3	0.0	-1.9***
	(12.3)	(12.4)	(12.6)	(13.1)	(0.1)	(0.1)
Log population	9.3	8.9	9.0	8.5	0.4***	0.4***
	(1.8)	(1.9)	(1.7)	(1.8)	(0.0)	(0.0)
Economic Variables (by geographical unit)						
Immigr. unemp. rate (1900-1910)	20.7	20.0	7.8	7.0	0.7***	0.8***
	(11.2)	(11.4)	(9.6)	(9.6)	(0.1)	(0.1)
Native unemp. rate (1900-1910)	16.9	16.8	8.9	8.4	0.1	0.4***
	(7.2)	(7.8)	(7.6)	(8.0)	(0.1)	(0.1)
Immigr. labor force participation rate	94.6	94.4	21.6	21.5	0.2***	0.1
	(4.8)	(5.1)	(8.8)	(9.0)	(0.0)	(0.1)
Native labor force participation rate	90.3	90.0	31.8	30.0	0.4***	1.8***
	(4.0)	(4.5)	(9.0)	(9.4)	(0.0)	(0.1)
Immigr. log occ. score	3.12	3.06	2.60	2.55	0.06***	0.05***
	(0.18)	(0.22)	(0.28)	(0.30)	(0.001)	(0.002)
Native log occ. score	3.16	3.12	2.84	2.81	0.04***	0.03***
	(0.16)	(0.23)	(0.19)	(0.22)	(0.001)	(0.001)
Network Variables (by geographical unit)						
Concentration index	6.6	8.3	5.9	7.9	-1.8***	-2.0***
	(6.2)	(9.2)	(6.0)	(9.7)	(0.1)	(0.1)
Share of immigrants	34.5	35.2	32.0	32.0	-0.7***	0.0
	(14.4)	(14.7)	(13.3)	(13.7)	(0.1)	(0.1)
Measures of Immigrants Assimilation						
Log occupational score	3.12	3.03	2.67	2.54	0.09***	0.13***
	(0.38)	(0.43)	(0.64)	(0.65)	(0.003)	(0.011)
Full-Year employed last year (%)	82.8	83.6	95.9	96.6	-0.9***	-0.7***
	(37.7)	(37.0)	(19.8)	(18.2)	(0.3)	(0.2)
Labor Force participation (%)	94.1	93.8	19.2	18.7	0.3**	0.5*
	(23.6)	(24.1)	(39.4)	(39.0)	(0.1)	(0.3)
American wife/husband	5.2	5.2	11.5	11.4	0.0	0.1
	(22.1)	(22.2)	(31.8)	(31.8)	(0.2)	(0.3)
Speaks English	86.5	79.0	80.7	74.5	7.5***	6.2***
	(34.2)	(40.7)	(39.5)	(43.6)	(0.2)	(0.3)
US citizenship (%)	53.8	44.7	51.3	47.1	9.1***	4.2***
	(49.9)	(49.7)	(50.0)	(50.0)	(0.3)	(0.5)

Notes: Pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%) with exception of the unemployment variables, which are available only for 1900 and 1910. Geographical unit - county. For further detail, see note from Table C.2. * Significant at 10%; ** significant at 5%; *** significant at 1%.

2.4. Measures of Immigrants Assimilation

We look at several economic and social-cultural outcomes of immigrants: whether the immigrant was full-year employed last year; labor force participation; 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 C.2 in Online Appendix C.

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) be in the labor force; iii) speak English; and iv) become a US citizen. Furthermore, male immigrants with an American name are less likely to be employed for the full year than those without American names.[17]

Table 4 also shows the corresponding results for female immigrants for the whole sample. First, the labor force participation is low among females between 1900 and 1930. Second, 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) speak English; and iv) become a US citizen. Also, as in the case of males, the female immigrants with an American name are not more likely to be employed for the full year than those without American names.

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.

We mentioned above three recent papers by Abramitzky et al. (2017), Bi-

[17]Looking at the three most represented groups of immigrants by birthplace, the results show significant heterogeneity. Table C.5 in Online Appendix C shows that for German immigrants, there is much smaller difference between the economic outcomes of those with and without American names. The differences are significantly stronger 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 C.5.

avaschi et al. (2017), and Goldstein and Stecklov (2016), which have used a variety of empirical strategies to show that the adoption of American sounding first names have indeed led to substantial improvements in labor market outcomes of first and second generation immigrants. The conjecture in these papers is that this is due to a more successful assimilation process by those with American names. In Online Appendix B we present a large range of estimates confirming that the same patterns emerge (at least in a correlational sense) when we use our definition of American name (as well as the ANI index[18]).

2.5. American Names among Second-Generation Immigrants

It is also interesting to examine naming patterns for second-generation immigrants. In our sample, about 86% of all boys and 93% of all girls born to immigrants have an American name.[19]

Table 5 relates to naming patterns for fathers and sons and, fathers and daughters. Starting with the whole sample, 87% 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 79%. 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).

[18]See Tables D.2 and D.3 in Online Appendix D.

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

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	21.1	12.6	Daughter	Non-American name	9.7	5.0
	American name	78.9	87.4		American name	90.3	95.0

German Father

		Non-American name	American name			Non-American name	American name
		Percentage				Percentage	
Son	Non-American name	18.4	12.2	Daughter	Non-American name	4.3	4.1
	American name	81.6	87.8		American name	95.7	95.9

Italian Father

		Non-American name	American name			Non-American name	American name
		Percentage				Percentage	
Son	Non-American name	21.2	12.6	Daughter	Non-American name	9.3	6.0
	American name	78.8	87.4		American name	90.7	94.0

Former-USSR Father

		Non-American name	American name			Non-American name	American name
		Percentage				Percentage	
Son	Non-American name	18.3	14.8	Daughter	Non-American name	7.1	4.2
	American name	81.7	85.2		American name	92.9	95.8

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

3. 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.[20]

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 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.[21] 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)).

[20] 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.

[21] 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.

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.[22]

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, $h(1 - p_N, e_A, e_N) = (1 - p_N)\, e_A - p_N e_N$.[23]

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

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

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

In order to simplify estimation of the model in (1), 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 (1) that the proportion of individuals

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

[23] 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 $\frac{\partial^2 h}{\partial e_N \partial p_N} = -1 < 0$.

with American first names is

$$\Pr(t_i < b_i) = G\left[g_1^{-1}\left\{h(1-p_N, e_A, e_N)\right\} - g_2(q_N)\right], \tag{2}$$

where $G : \mathbb{R} \mapsto \mathbb{R}$ is the cumulative distribution function (CDF) of ε_i. The binary choice model in (2) is the basis of our empirical work.

Although the model in equation (2) is quite simple, it implies some restrictions on the specification of our econometric model. One important prediction is that, under reasonable assumptions,[24] $\Pr(t_i < b_i)$ 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 (2) for p_N), examining the impact of the exogenous variables (q_N, e_A, e_N) on p_N. In particular, we estimate

$$\Pr(t_i < b_i) = \Psi\left(q_N, e_A, e_N\right), \tag{3}$$

where Ψ is a reduced-form function, and we take $\Psi(\cdot)$ to be a linear probability model for convenience in Section 4.

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 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 the endogeneity problems, one could use lagged variables. Nevertheless, while it may be true that these are to some extent less prone to endogeneity due to some contemporaneous unobservable shocks, they are also clearly worse measures of the local drivers of contemporaneous name Americanization. Therefore, we have decided to construct our variables using the contemporaneous census. This means that (q_N, e_A, e_N) are

[24] 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 $\frac{\partial g_2}{\partial q_N} > 0$.

measured contemporaneously and correspond respectively to the concentration of immigrants in a particular county in that census year, the average economic well-being of natives in the county in that census year and the average economic well-being of immigrants in the county in that census year. In addition, instead of relying purely on cross-sectional variation to estimate equation (3), we use the census years available to us to construct a four period panel of counties, 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.[25]

There exist other explanatory variables, say x_i, which we control for when implementing equation (3) 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).

4. Determinants of American First Names

In this section we examine how different economic and social network variables affect the choice of an American name by immigrants. In order to do this, our idea is to exploit within county variation to estimate the impact of local immigrant concentrations on naming decisions. This strategy allows us to account for the role of permanent local (county level) heterogeneity, which can affect both name Americanization by immigrants and the variables we study as potential determinants of name Americanization.

We use a linear probability model for first-name choice. To start with, we focus on male immigrants. Our estimation sample includes 238560 male immigrants from 16 different countries, between 16 and 65 years of age, in the 1900, 1910, 1920 and 1930 censuses.

Our basic specification is

$$\begin{aligned}&\Pr\left(AmericanName_{ibect}=1\right)\\&=\gamma q_{Nbct}+e_{Act}\delta_A+e_{Nct}\delta_N+Z_{bct}\rho+YearsUS_{ibect}\beta_Y+X_{ibect}\beta_X+\alpha_b+\phi_c+\delta_e+\psi_t.\end{aligned}\tag{4}$$

[25] To use the unemployment rate variables we are restricted to a two period panel, as the unemployment variables are only available in the 1900 and 1910 census.

Here, the dependent variable (also defined above), $AmericanName_{ibect}$, is an indicator 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}, we use the concentration index ("Concentration index" 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 census (t). The variables e_{Act} and e_{Nct} are vectors which include, for each county and year, current values of labor force participation for natives and immigrants ("Immigr. labor force participation rate" and "Native labor force participation rate") and average log occupational scores for natives and immigrants ("Immigr. log occ. score" and "Native log occ. score").[26] Z_{bct} includes other county variable, the percentage of immigrants between 16 and 65 years old in a given county ("Share of Immigrants"). This variable is 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 $YearsUS_{ibect}$ includes number of years in the US, its squared and its cubic term, X_{ibect} includes age in years, its squared and its cubic 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 (five year group intervals fixed effects from 1831-1835 up to 1926-1930) 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 models where we include fixed effects for the interactions of birthplace and year, interactions of county and year, and interactions of county and birthplace.

4.1. Main Results - Male Immigrants

Table 6 presents estimates of equation (4). This and subsequent tables regarding the determinants of name choice show the effects in percentage

[26]For robustness purposes, we also consider different specifications including economic and network variables using only individuals that arrived to the US later than the individual (entry after) "Immigr. labor force participation rate (entry after)" , "Immigr. log occ. score (entry after)" and another group of variables using only the individual that arrived to the US before the individual "Immigr. labor force participation (entry before)" , "Immigr. log occ. score (entry before) " . Furthermore, we use also unemployment rates for natives and immigrants ("Immigr. unemp. rate" and "Native unemp. rate"), which are only available for 1900 and 1910 census.

points. Column (1) of Table 6 shows that years in the US and the concentration index are the most important characteristics that determine the adoption of an American name. We fail to reject the joint hypothesis that the coefficients on all economic variables ("Immigr. labor force participation rate", "Native labor force participation rate", "Immigr. log occ. score", "Native log occ. score") are equal to zero. However, we reject the hypothesis that the coefficients on local network variables ("Concentration index", "Share of Immigrants") 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.

There are no robust effects on name adoption of the four economic variables we include in the regression. With regards to individual regressors, the adoption of an American name increases with years in the US.

In terms of magnitudes, our estimates indicate that the likelihood of having an American name increases by 5.4 percentage points (p.p.) for an increase of years in the US from 0 to 10 and decreases by 1.4 p.p. for a one standard deviation increase in the concentration index.

As mentioned above, we also considered different specifications of equation (4), where we included fixed effects interactions of birthplace and year, county and year, and county and birthplace. Columns (2) to (4) of Table 6 present estimates for the different specifications and the results are quite robust to the different specification choices. In columns (2) and (3), interactions between birthplace and year and those between county and year actually increase the effect of the concentration index by less than 1 p.p.. It is only marginally significant and its size is slightly smaller in column (4). However, we regard this as reassuring evidence since including interactions between birthplace and county is a very demanding specification. The effect of years in the US is quite stable across all specifications.

Table 7 replicates the results of Table 6 using the ANI index to classify

Table 6: American Name - Linear Probability Model (LPM) - Male Immigrants

	(1)	(2)	(3)	(4)
Years in the US	0.749***	0.550***	0.607***	0.610***
	(0.100)	(0.103)	(0.105)	(0.103)
(Years in the US)^2 / 10	-0.224***	-0.162***	-0.181***	-0.187***
	(0.029)	(0.030)	(0.003)	(0.003)
(Years in the US)^3 /100	0.019***	0.019***	0.021***	0.022***
	(0.004)	(0.004)	(0.004)	(0.004)
Age	0.131	0.100	0.104	0.0349
	(0.240)	(0.245)	(0.247)	(0.236)
Age^2 / 10	-0.084	-0.074	-0.074	-0.055
	(0.061)	(0.062)	(0.062)	(0.060)
Age^3 / 100	0.008	0.007	0.007	0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Log population	-0.581	-0.419		
	(0.656)	(0.514)		
Immigr. labor force participation rate	-2.939	-3.325		
	(2.976)	(2.935)		
Native labor force participation rate	-1.226	-1.012		
	(4.060)	(3.979)		
Immgr. log occ. score	2.034	2.094		
	(1.664)	(1.617)		
Native log occ. score	-2.200	-1.976		
	(1.955)	(1.940)		
Concentration index	-0.196***	-0.284***	-0.294***	-0.149*
	(0.039)	(0.036)	(0.044)	(0.081)
Share of Immigrants	0.022	0.008		
	(0.030)	(0.032)		
Observations	238,560	238,560	238,560	238,560
R-squared	0.159	0.162	0.187	0.230
Fixed Effects				
Year	YES			
Birthplace	YES			
County	YES	YES		
Cohort of Entry group	YES	YES	YES	YES
Birthplace x Year		YES	YES	YES
County x Year			YES	YES
County x Birthplace				YES
Joint Hypothesis Test				
Economic Variables	0.97	1.05	-	-
(p-value)	(0.423)	(0.380)	-	-
Network Variables	13.29	36.89	-	-
(p-value)	0.000	0.000	-	-

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table 7: American Name Index (ANI) - Male Immigrants

	(1)	(2)	(3)	(4)
Years in the US	0.392***	0.278***	0.284***	0.277***
	(0.044)	(0.044)	(0.044)	(0.043)
(Years in the US)^2 / 10	-0.095***	-0.061***	-0.065***	-0.068***
	(0.014)	(0.014)	(0.014)	(0.014)
(Years in the US)^3 / 100	0.008***	0.006***	0.006***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Age	-0.063	-0.065	-0.050	-0.092
	(0.100)	(0.100)	(0.100)	(0.102)
Age^2 / 10	-0.034	-0.033	-0.036	-0.026
	(0.026)	(0.026)	(0.026)	(0.027)
Age^3 / 100	0.004*	0.004*	0.004*	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Log population	0.030	0.178	-	-
	(0.279)	(0.233)	-	-
Immigr. labor force participation rate	0.384	0.624	-	-
	(1.412)	(1.377)	-	-
Native labor force participation rate	-3.866**	-2.918*	-	-
	(1.762)	(1.679)	-	-
Immgr. log occ. score	1.149	1.074	-	-
	(0.758)	(0.727)	-	-
Native log occ. score	0.186	-0.348	-	-
	(0.882)	(0.833)	-	-
Concentration index	-0.118***	-0.130***	-0.146***	-0.107***
	(0.014)	(0.016)	(0.018)	(0.033)
Share of Immigrants	0.011	-0.002	-	-
	(0.013)	(0.014)	-	-
Observations	238,560	238,560	238,560	238,560
R-squared	0.137	0.143	0.167	0.216
Fixed Effects				
Year	YES	-	-	-
Birthplace	YES	-	-	-
County	YES	YES	-	-
Cohort of Entry group	YES	YES	YES	YES
Birthplace x Year	-	YES	YES	YES
County x Year	-	-	YES	YES
County x Birthplace	-	-	-	YES
Joint Hypothesis Test				
Economic Variables	1.79	1.24	-	-
(p-value)	(0.128)	(0.291)	-	-
Network Variables	40.23	39.85	-	-
(p-value)	0.000	0.000	-	-

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

names, instead of our dichotomous definition. The results are qualitatively similar to the ones just reported. Across all four columns increases in the concentration index decrease the adoption of an American name, and the adoption of an American name increases with years in the US. As before, we fail to reject the joint hypothesis that the coefficients on all economic variables are equal to zero.

We now examine to what extent immigrants with different years in the US responded differently to incentives when deciding to adopt American Names. 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). Table 8 shows the impact of the different variables on the choice of American name, at five different subsamples defined by immigrant's time since arrival: 0-5, 6-10, 11-15, 16-20, and more than 21 years in the US. The results suggest that the longer an immigrant is in the US, the weaker seems to be the average impact of network (concentration index) on American name adoption, while there is not much of a pattern with regards to economic variables. In fact, it is interesting that the impact of the concentration index is very strong early on after arrival, in line to our original hypothesis.

Finally, we examine the heterogeneity of the main results by the birthplace of immigrants. In column (1) of Table 9, we reproduce column (1) of Table 6 for comparison. In column (2) of Table 9, we present results interacting the concentration index with aggregated country groups, namely, Southern (Italy, Greece, France, Portugal, and Spain), Nordic (Norway, Sweden, and Denmark), Eastern (Czechoslovakia, Hungary, Poland, and former USSR), Asian (Japan and China), and Mexico. Immigrants from Germany are in the omitted category in this specification. The results show significant heterogeneity among groups of countries of origin. In particular, the concentration index has a stronger impact on name adoption in immigrants from Asia and Southern countries, a smaller but also significant effect on immigrants from Nordic countries and Mexico but no effect on name adoption of immigrants from Eastern Europe and Germany.

Table 8: American Name (LPM) - Male Immigrants - Years in the US

	Years in the US				
	[0-5]	[6-10]	[11-15]	[16-20]	+21
Years in the US	-3.431	-67.81	555.0**	238.3	-0.212
	(2.315)	(60.39)	(235.2)	(384.4)	(0.750)
(Years in the US)^2 / 10	30.55**	94.42	-443.6**	-136.1	0.081
	(13.30)	(81.33)	(189.3)	(219.7)	(0.197)
(Years in the US)^3 / 100	-49.1**	-42.9	117.7**	25.9	-0.006
	(21.4)	(36.1)	(50.6)	(41.8)	(0.017)
Age	-1.710**	-0.366	0.284	1.368**	0.153
	(0.857)	(0.628)	(0.620)	(0.598)	(0.703)
Age^2 / 10	0.446*	0.045	-0.156	-0.446***	-0.100
	(0.254)	(0.178)	(0.172)	(0.159)	(0.160)
Age^3 / 100	-0.037	-0.002	0.016	0.040***	0.009
	(0.023)	(0.016)	(0.015)	(0.013)	(0.012)
Log population	-1.523	-1.571	-0.364	1.592	-1.413*
	(1.905)	(1.674)	(1.437)	(1.111)	(0.840)
Immigr. labor force participation rate	-5.921	4.083	-12.09	-5.329	0.555
	(16.40)	(10.39)	(13.51)	(9.132)	(4.265)
Native labor force participation rate	21.65	-11.27	-4.427	-7.961	-6.294
	(15.42)	(13.06)	(14.36)	(9.910)	(5.873)
Immgr. log occ. score	-7.149	0.565	13.29**	6.699	1.192
	(7.766)	(5.949)	(6.114)	(4.914)	(2.284)
Native log occ. score	3.856	1.214	-8.613	-9.455*	-0.090
	(7.435)	(6.545)	(7.366)	(5.037)	(2.927)
Concentration index	-0.293**	-0.036*	-0.078***	-0.041***	-0.051***
	(0.130)	(0.019)	(0.021)	(0.015)	(0.012)
Share of Immigrants	-0.126	0.140**	0.0410	-0.100*	-0.081**
	(0.082)	(0.066)	(0.089)	(0.052)	(0.038)
Observations	29,512	42,519	33,973	48,760	83,748
R-squared	0.250	0.192	0.236	0.180	0.151
Fixed Effects					
Year	YES	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES	YES
County	YES	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES	YES

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table 9: American Name (LPM) - Male Immigrants - by Birthplace

	(1)	(2)
Years in the US	0.749***	0.727***
	(0.100)	(0.100)
(Years in the US)^2 / 10	-0.224***	-0.218***
	(0.029)	(0.029)
(Years in the US)^3 / 100	0.019***	0.022***
	(0.004)	(0.004)
Age	0.131	0.118
	(0.240)	(0.240)
Age^2 / 10	-0.084	-0.081
	(0.061)	(0.061)
Age^3 / 100	0.008	0.007
	(0.005)	(0.005)
Log population	-0.581	-0.634
	(0.656)	(0.725)
Immigr. labor force participation rate	-2.939	-2.339
	(2.976)	(3.002)
Native labor force participation rate	-1.226	-1.011
	(4.060)	(4.090)
Immgr. log occ. score	2.034	2.042
	(1.664)	(1.684)
Native log occ. score	-2.200	-2.180
	(1.955)	(1.986)
Concentration index	-0.196***	
	(0.039)	
Share of Immigrants	0.022	0.034
	(0.030)	(0.031)
Concentration index$^+$		-0.006
		(0.055)
Concentration index × Southern		-0.466***
		(0.126)
Concentration index × Nordic		-0.283***
		(0.078)
Concentration index × Eastern		-0.069
		(0.072)
Concentration index × Asian		-0.499***
		(0.180)
Concentration index × Mexico		-0.303***
		(0.099)
Observations	238,560	238,560
R-squared	0.159	0.160
Fixed Effects		
Year	YES	YES
Birthplace	YES	YES
County	YES	YES
Cohort of Entry group	YES	YES

Notes: The table reports the effects in percentage points. $^+$ Omitted group: Germany. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930). Group of countries: Nordic (Norway, Sweden, and Denmark), Southern (Italy, Greece, France, Portugal, and Spain), Eastern (Czechoslovakia, Hungary, Poland, and former USSR), and Asian (Japan and China).

4.2. Robustness Checks - Male Immigrants

In this subsection, we examine the sensitivity of the main results presented in Section 4.1. First, we use the main specification using unemployment rates, which is only available for 1900 and 1910. Second, we use different specifications regarding economic and network variables, namely using different cohorts since arrival in the US in the construction of the variables. Overall, this subsection suggests that the results presented in the previous subsection are quite robust to different samples and specification choices.[27]

As an alternative specification, in Table 10, we controlled for unemployment rates for natives and immigrants, which is available only in 1900 and 1910 census years. The estimation results are similar to columns (1) and (2) in Table 6, and the main conclusions remain the same; years in the US and the concentration index are still the only significant variables.

In Table 11 we repeat exactly the same exercise as in Table 6, but changing the economic and network variables included in the regression models. In particular, we construct new sets of economic and network variables using different cohorts of arrival in the US: (i) entry after - for each immigrant we use the same economic and network variables but using only the immigrants that arrived after that specific immigrant; (ii) entry before - for each immigrant we use the same economic and network variables but using only immigrants that arrived before that specific immigrant. These additional results support that our main results, that is, the most important determinants of American name are years in the US and the concentration index, are robust with respect to alternative specifications economic and network variables. The only difference is that the likelihood of adopting an American name decreases with higher labor force participation among immigrants that arrived before in the US. Moreover, in this case, we reject the joint hypothesis that the coefficients on all economic variables are equal to zero (with a p-value of 0.04) in the baseline specification (column 1).

[27]In general, our main results are also robust to a specification where we use lagged network and economic variables instead of the current variables (results are available upon request), with the exception of the importance of the lagged economic variables. When using lagged network and economic variables, our results show that poor lagged labor market conditions for immigrant lead to more frequent name changes. Nevertheless, as stated before, while it may be true that the lagged variables are to some extent less prone to endogeneity due to some contemporaneous unobservable shocks, they are also clearly worse measures of the local drivers of contemporaneous name Americanization.

Table 10: American Name (LPM) - Robustness Check (Unemployment Rate) - Male Immigrants

	(1)	(2)
Years in the US	0.895***	0.797***
	(0.155)	(0.147)
(Years in the US)^2 / 10	-0.303***	-0.265***
	(0.068)	(0.060)
(Years in the US)^3 / 100	0.030***	0.028***
	(0.009)	(0.008)
Age	0.109	0.132
	(0.321)	(0.325)
Age^2 / 10	-0.048	-0.053
	(0.084)	(0.085)
Age^3 / 100	0.003	0.003
	(0.007)	(0.007)
Log population	0.863	0.521
	(1.411)	(1.410)
Immigr. unemp. rate	1.019	0.991
	(2.386)	(2.374)
Native unemp. rate	-2.794	-3.860
	(4.021)	(3.954)
Immgr. log occ. score	2.799	3.056
	(3.346)	(3.353)
Native log occ. score	-1.545	-1.043
	(3.261)	(3.282)
Concentration index	-0.228***	-0.258***
	(0.055)	(0.053)
Share of Immigrants	-0.034	-0.032
	(0.050)	(0.055)
Observations	159,305	159,305
R-squared	0.139	0.140
Fixed Effects		
Year	YES	
Birthplace	YES	
County	YES	YES
Cohort of Entry group	YES	YES
Birthplace x Year		YES
Joint Hypothesis Test		
Economic Variables	0.34	0.46
(p-value)	(0.848)	(0.763)
Network Variables	11.51	14.83
(p-value)	(0.000)	(0.000)

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table 11: American Name (LPM) - Robustness Check (Entry in the US: Before and After) - Male Immigrants

	(1)	(2)	(3)	(4)
Years in the US	0.749***	0.560***	0.606***	0.570***
	(0.123)	(0.128)	(0.133)	(0.133)
(Years in the US)^2 / 10	-0.228***	-0.168***	-0.182***	-0.177***
	(0.038)	(0.040)	(0.040)	(0.040)
(Years in the US)^3 / 100	0.023***	0.019***	0.021***	0.020***
	(0.005)	(0.005)	(0.005)	(0.005)
Age	0.097	0.050	0.020	0.014
	(0.245)	(0.250)	(0.250)	(0.242)
Age^2 / 10	-0.078	-0.064	-0.055	-0.051
	(0.062)	(0.063)	(0.063)	(0.061)
Age^3 / 100	0.007	0.006	0.005	0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Log population	-0.361	-0.112		
	(0.648)	(0.527)		
Immigr. labor force participation rate (entry after)	-0.240	-0.0853	-0.229	-0.333
	(1.297)	(1.261)	(2.025)	(2.116)
Immigr. labor force participation rate (entry before)	-3.476***	-2.299*	-3.258	-1.952
	(1.332)	(1.378)	(2.302)	(2.443)
Native labor participation force rate	-3.892	-3.319		
	(4.413)	(4.334)		
Immgr. log occ. score (entry after)	1.039	1.082	0.0180	0.222
	(0.979)	(0.965)	(1.507)	(1.615)
Immgr. log occ. score (entry before)	-0.995	-0.877	-2.331	-1.963
	(1.031)	(1.021)	(1.420)	(1.523)
Native log occ. score	-1.857	-1.269		
	(2.191)	(2.205)		
Concentration index (entry after)	-0.232***	-0.280***	-0.304***	-0.138
	(0.041)	(0.042)	(0.049)	(0.088)
Concentration index (entry before)	-0.131**	-0.277***	-0.286***	-0.126
	(0.060)	(0.056)	(0.063)	(0.096)
Share of Immigrants (entry after)	0.014	-0.016	0.131	0.094
	(0.031)	(0.033)	(0.121)	(0.133)
Share of Immigrants (entry before)	-0.002	-0.012	0.151	0.103
	(0.042)	(0.043)	(0.121)	(0.132)
Observations	226,259	226,259	226,259	226,259
R-squared	0.155	0.157	0.178	0.218
Fixed Effects				
Year	YES			
Birthplace	YES			
County	YES	YES		
Cohort of Entry group	YES	YES	YES	YES
Birthplace x Year		YES	YES	YES
County x Year			YES	YES
County x Birthplace				YES
Joint Hypothesis Test				
Economic Variables	2.21	1.09	-	-
(p-value)	(0.040)	(0.363)	-	-
Network Variables	8.83	17.82	-	-
(p-value)	(0.000)	(0.000)	-	-

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

4.3. Determinants of American Name Choice for Female Immigrants

In this subsection, we report results for female immigrants. Table 12 shows estimates with specifications similar to those in Table 6. Immigrants (natives) economic and network variables in Table 12 are computed using only female immigrants (natives). We find that in our main specification (column (1)), only years in the US and age 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 2.9 p.p. for an increase of years in the US from 0 to 10 and decreases by 2.0 p.p. for an increase in immigrants' age from 30 to 40.

We also find that the impact of concentration index is not robust, unlike what we document for males. The effect of the concentration index is insignificant in the baseline specification in column (1), it is significantly negative in columns (2) and (3), and it becomes almost zero in column (4). Compared to Table 6, the effect seems weaker.

As in male immigrants, no economic variable seems to affect the adoption of an American name. This is not surprising given that women were much less attached to the labor market in the early 20th century than now.

4.4. 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 that it responds 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 parental decision about whether or not to choose an American name for their children.

We essentially re-estimate equation (4) with a new dependent variable: an indicator for whether the child of the immigrant born in the US has an American name or not. One difference relative to our regressions above is that, across specifications, we add child-specific variables to the variables already used in equation (4). 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 13. In column (1) and (3), we use exactly the same specification as in equation (4), respectively for sons and daughters. In column (2) and (4), for sons and daughters, respectively, we add four child-specific variables: child's age, an indicator for whether the

Table 12: American Name (LPM) - Female Immigrants

	(1)	(2)	(3)	(4)
Years in the US	0.389***	0.277**	0.282**	0.258*
	(0.127)	(0.129)	(0.131)	(0.135)
(Years in the US)^2 / 10	-0.109***	-0.073*	-0.082**	-0.072*
	(0.039)	(0.040)	(0.041)	(0.042)
(Years in the US)^3 / 100	0.011**	0.009*	0.010**	0.009*
	(0.005)	(0.005)	(0.005)	(0.005)
Age	-0.756**	-0.793***	-0.744**	-0.744**
	(0.302)	(0.302)	(0.302)	(0.302)
(Age)^2 / 10	0.154**	0.161**	0.149*	0.148*
	(0.077)	(0.077)	(0.077)	(0.077)
(Age)^3 / 100	-0.014**	-0.015**	-0.014**	-0.014**
	(0.006)	(0.006)	(0.006)	(0.006)
Log population	-0.576	0.154		
	(0.607)	(0.576)		
Immigr. labor force participation rate	-4.206	-2.161		
	(3.097)	(2.981)		
Native labor force participation rate	-1.159	-3.829		
	(4.758)	(4.518)		
Immgr. log occ. score	-1.255	-0.864		
	(0.773)	(0.746)		
Native log occ. score	1.598	1.499		
	(1.506)	(1.475)		
Concentration index	-0.074	-0.122***	-0.092**	0.020
	(0.048)	(0.044)	(0.046)	(0.108)
Share of Immigrants	-0.021	-0.064		
	(0.047)	(0.046)		
Observations	168,631	168,631	168,631	168,631
R-squared	0.118	0.120	0.139	0.178
Fixed Effects				
Year	YES			
Birthplace	YES			
County	YES	YES		
Cohort of Entry group	YES	YES	YES	YES
Birthplace x Year		YES	YES	YES
County x Year			YES	YES
County x Birthplace				YES
Joint Hypothesis Test				
Economic Variables	1.32	0.92	-	-
(p-value)	(0.261)	(0.452)	-	-
Network Variables	1.87	6.50	-	-
(p-value)	(0.154)	(0.002)	-	-

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

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 have substantial impacts on the probability of a son and a daughter having an American name. The likelihood of adopting an American name increases by around 6 and 2 percentage points, respectively. Furthermore, having an immigrant mother decreases the likelihood of having an American name by around 1.2 p.p. for sons, and 0.7 p.p. for daughters. In terms of parental variables, years in the US is a statistically important determinant of American name adoption for daughters (not for sons), but the magnitude is relatively small.

The estimates of the effects of the economic variables are not significant for sons and puzzling for daughters. The native log occupational score has a negative impact on the adoption of an American name. However, we fail to reject that the economic variables are jointly statistically equal to zero.

Finally, in counties with a higher concentration index, immigrants are more likely to give their sons 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. Regarding daughters the point estimate of the concentration index is small and not statistically significant.

We present results for the subsamples by father's nationality in Tables C.9 – C.11 in Online Appendix C, 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 and girls with a father from Italy.

Finally, we redo all our results using the ANI index instead of our American name classification. In Online Appendix D, Tables D.4, D.5, D.6, D.7, D.8 and D.9 report estimation results using the name index approach by repeating the exercises reported in Tables B.1, B.2, 8, 9, 10, 11, 12 and 13, respectively. We find that our main qualitative results remain intact.

Table 13: American Name (LPM) - Sons and Daughters

	Sons		Daughters	
Father with American Name		6.116***		1.699***
		(0.405)		(0.274)
Immigrant Mother		-1.239***		-0.667***
		(0.404)		(0.258)
Siblings (aged 0-10)		0.0620		0.0203
		(0.127)		(0.0732)
Child's Age		-0.399		0.774***
		(0.358)		(0.266)
Child's Age^2		0.088		-0.125**
		(0.086)		(0.060)
Child's Age^3		-0.005		0.007*
		(0.006)		(0.004)
Years in the US (father)	0.194	0.167	0.261**	0.203*
	(0.199)	(0.196)	(0.121)	(0.122)
Years in the US^2 /10 (father)	0.017	0.017	-0.010**	-0.083*
	(0.073)	(0.072)	(0.043)	(0.044)
Years in the US^3 /100 (father)	-0.006	-0.006	0.013**	0.012**
	(0.010)	(0.010)	(0.005)	(0.005)
Age (father)	-0.874	-0.922	-0.117	-0.344
	(0.896)	(0.846)	(0.533)	(0.544)
Age^2 (father) /10	0.169	0.186	0.026	0.074
	(0.222)	(0.210)	(0.132)	(0.133)
Age^3 (father) /100	-0.010	-0.012	-0.002	-0.005
	(0.018)	(0.017)	(0.011)	(0.011)
Log population	-0.915	-0.905	0.833	0.875
	(0.858)	(0.868)	(0.571)	(0.568)
Immigr. labor force participation rate	-3.032	-2.664	5.153	5.196
	(4.543)	(4.533)	(3.233)	(3.243)
Native labor force participation rate	-3.504	-3.498	-2.281	-2.405
	(5.118)	(5.072)	(4.128)	(4.150)
Immgr. log occ. score	-1.895	-2.134	1.561	1.254
	(2.618)	(2.616)	(1.643)	(1.641)
Native log occ. score	-0.074	0.0069	-5.860***	-5.706***
	(2.409)	(2.410)	(2.221)	(2.213)
Concentration Index	-0.170***	-0.155***	-0.034	-0.030
	(0.041)	(0.042)	(0.026)	(0.026)
Share of Immigrants	0.053	0.053	-0.036	-0.035
	(0.037)	(0.036)	(0.032)	(0.031)
Observations	94,440	94,440	90,953	90,953
R-squared	0.058	0.062	0.093	0.094
Fixed Effects				
Year	YES	YES	YES	YES
Birthplace (father)	YES	YES	YES	YES
County (father)	YES	YES	YES	YES
Cohort of Entry group (father)	YES	YES	YES	YES

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

5. 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. 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 social pressures but not to economic variables. Male 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. 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., Eriksson, K., 2012. Europe's tired, poor, huddled masses: self-selection and economic outcomes in the Age of Mass Migration. American Economic Review 102, 1832–1856.

Abramitzky, R., Boustan, L., 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., Eriksson, K., 2014. A nation of immigrants: Assimilation and economic outcomes in the Age of Mass Migration. Journal of Political Economy 122, 467–506.

Abramitzky, R., Boustan, L.P., Eriksson, K., 2017. Cultural Assimilation during the Age of Mass Migration. NBER Working Papers 22381. National Bureau of Economic Research, Inc.

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

Algan, Y., Mayer, T., Thoenig, M., 2013. The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France. CEPR Discussion Paper 9416. Centre for Economic Policy Research.

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

Bandiera, O., Rasul, I., 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., Tirole, J., 2006. Incentives and prosocial behavior. American Economic Review 96, 1652–1678.

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

Biavaschi, C., Giulietti, C., Siddique, Z., 2017. The economic payoff of name Americanization. Journal of Labor Economics 35, 1089–1116.

Card, D., 2011. Origins of the unemployment rate: The lasting legacy of measurement without theory. American Economic Review 101, 552–557.

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

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

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

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

Goldstein, J.R., Stecklov, G., 2016. From Patrick to John F.: Ethnic names and occupational success in the last era of mass migration. American Sociological Review 81, 85–106.

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

Hatton, T.J., 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, S95–S126.

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

National Archives and Records Administration, 1977–2002. 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. NARA.

Olivetti, C., Paserman, M.D., 2015. In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850-1940. American Economic Review 105, 2695–2724.

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

Ruggles, S., Alexander, J.T., Genadek, K., Goeken, R., Schroeder, M.B., Sobek, M., 2010. Integrated public use microdata series: Version 5.0 [machine-readable database. Minneapolis: University of Minnesota.

Watkins, S., 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, 169–209.

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

Appendix A. Data Appendix: Classification of Name Types

In the analysis, we distinguish the immigrants by the type of their first name: American names and Non-American names. In general terms, the definition of American name we use takes all the top 100 baby names in the US Social Security records of the 1900s (independently of the etymological origin), plus all the remaining names in that list (top 101 to top 1000) which are etymologically of American origin, i.e. names that are not from a Latin, Greek, German, Eastern Europe, Scandinavian, Asian, or other non-American/non-British origin.

This appendix gives a detailed description of the classification of the name types. 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 Non-US 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 of origin of the immigrants included in the study (urlhttp://www.behindthename.com/).
- Step 1b For US names, we used the US Social Security Administration database (urlhttp://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 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 Non-US 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 Non-US 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 as follows: 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 Non-US names list. The Common names include all the names included in both lists - the US Social Security database and the Non-US names list. The Origin names include only those names in the Non-US 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 a US name nor a name from his birthplace (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 Finally, we classified, as American names, 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 Non-US list and the British names. All remaining names were classified as Non-American names.

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.[28] Therefore, from Steps 1 and 2, we obtained a classification of types for around 90% of first names for most countries.[29] The Other names never represented more than 2%, except in the cases of Mexico and China.

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 US and Non-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 US names list as an American name and all the remaining Common names as Non-American names. As highlighted before, in Germany and France an American name

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

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

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.6 reports top 10 popular baby names between 1880-1929 by decades and gender, from social security records. Tables A.7 and A.8 show top 10 popular names in our census data by region, census year, and gender. Overall, there are some changes across time and region; however, the overall pattern looks stable. For example, in almost all cases, John and William, and Mary and Anna, are the most common names for males and females, respectively.

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

SS data*		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 the list of US and Non-US names, depending on whether we used TOP 1000, TOP 200 or TOP 100 from the US Social Security Administration database (url-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 Non-US names. In addition, it also includes the British names. The Common names include all the names included in both the US Social Security (top 100) and Non-US list of names. The Origin names include those names in the Non-US list that are not in the US Social Security database 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.

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

	Germany		Italy		Former USSR	
	Name	Rank*	Name	Rank*	Name	Rank*
American name						
US name (names in the top 100 US list that are not in the Non-US list)	JOHN	1	GEORGE	4	WILLIAM	2
Common name (names in the top 100 US list that are also in the Non-US list)	FRANK	8	LEO	44	ROBERT	6
US name (names in the top 101-1000 US list that are not in the Non-US list)	STEVEN	367	VINCENT	114	SIMON	221
Non-American name						
Common name (names in the Non-US list that are also in the top 101-1000 US list)	OTTO	125	ANTONIO	205	IVAN	210
Origin name (names in the Non-US list that are not in the top 1000 US list)	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.

Table A.6: Popularity of names by cohort

Males										
	1880 - 1889		**1890 - 1899**		**1900 - 1909**		**1910 - 1919**		**1920 - 1929**	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	John	7.6	John	6.6	John	5.8	John	5.4	Robert	5.1
2	William	7.2	William	5.9	William	4.7	William	4.4	John	5.0
3	James	4.6	James	4.1	James	4.2	James	4.0	James	4.5
4	George	4.0	George	3.5	George	3.0	Robert	3.4	William	4.5
5	Charles	4.0	Charles	3.0	Charles	2.5	Joseph	2.6	Charles	2.6
6	Frank	2.6	Joseph	2.4	Robert	2.4	George	2.5	George	2.3
7	Joseph	2.2	Frank	2.3	Joseph	2.4	Charles	2.5	Joseph	2.1
8	Henry	2.1	Robert	2.1	Frank	2.0	Edward	1.8	Richard	2.0
9	Robert	2.0	Edward	1.8	Edward	1.7	Frank	1.7	Edward	1.7
10	Thomas	2.0	Henry	1.8	Thomas	1.5	Thomas	1.3	Donald	1.7
	TOP 10	38.4		33.4		30.1		29.6		31.5

Females										
	1880 - 1889		**1890 - 1899**		**1900 - 1909**		**1910 - 1919**		**1920 - 1929**	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	Mary	6.5	Mary	5.6	Mary	5.2	Mary	5.6	Mary	5.7
2	Anna	2.7	Anna	2.3	Helen	2.2	Helen	2.9	Dorothy	3.0
3	Emma	1.8	Margaret	1.6	Margaret	1.9	Dorothy	2.4	Helen	2.3
4	Elizabeth	1.8	Helen	1.6	Anna	1.8	Margaret	2.2	Betty	2.3
5	Margaret	1.6	Elizabeth	1.4	Ruth	1.6	Ruth	2.0	Margaret	2.0
6	Minnie	1.6	Ruth	1.4	Elizabeth	1.3	Mildred	1.5	Ruth	1.8
7	Ida	1.3	Florence	1.4	Dorothy	1.3	Anna	1.4	Virginia	1.4
8	Bertha	1.3	Ethel	1.3	Marie	1.2	Elizabeth	1.4	Doris	1.2
9	Clara	1.3	Emma	1.2	Florence	1.2	Frances	1.2	Mildred	1.2
10	Alice	1.2	Marie	1.1	Mildred	1.2	Virginia	1.1	Frances	1.1
	TOP 10	21.1		19.0		18.8		21.8		21.9

Source:
US Social Security Administration database
(https://www.ssa.gov/oact/babynames/decades/index.html).
Notes:
Popular names by decade are determined taking into account the number of occurrences of each name during the decade, and then rank them in order of decreasing frequency. Rank 1 is assigned to the name with the highest number of occurrences, rank 2 to the name with the next highest, and so on.

Table A.7: Popularity of names by region and census year - Males

Males

Northeast Region

	1900		1910		1920		1930	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	JOHN	10.5	JOHN	9.4	JOHN	8.7	JOHN	8.5
2	WILLIAM	7.5	WILLIAM	6.5	JOSEPH	5.6	JOSEPH	6.1
3	JAMES	4.7	JOSEPH	4.9	WILLIAM	5.2	WILLIAM	4.4
4	GEORGE	4.4	GEORGE	4.0	GEORGE	3.5	FRANK	3.1
5	JOSEPH	4.3	JAMES	3.9	CHARLES	3.2	GEORGE	3.1
6	CHARLES	4.2	CHARLES	3.6	FRANK	3.2	EDWARD	2.9
7	THOMAS	3.3	FRANK	3.1	JAMES	3.1	JAMES	2.7
8	FRANK	3.2	EDWARD	3.0	EDWARD	2.8	CHARLES	2.6
9	EDWARD	3.1	THOMAS	2.5	THOMAS	2.1	THOMAS	2.0
10	HENRY	2.6	HENRY	2.1	HENRY	1.7	ROBERT	1.9
	TOP 10	47.7		42.9		39.2		37.3

Midwest (formerly North Central) Region

	1900		1910		1920		1930	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	JOHN	8.1	JOHN	7.2	JOHN	7.2	JOHN	6.7
2	WILLIAM	5.8	WILLIAM	5.2	WILLIAM	4.4	WILLIAM	3.9
3	GEORGE	4.0	GEORGE	3.8	JOSEPH	3.4	JOSEPH	3.7
4	FRANK	3.8	FRANK	3.4	FRANK	3.3	FRANK	3.2
5	HENRY	3.5	JOSEPH	2.9	GEORGE	3.3	GEORGE	3.0
6	CHARLES	3.2	HENRY	2.8	EDWARD	2.8	EDWARD	2.8
7	JOSEPH	3.0	EDWARD	2.8	HENRY	2.4	ROBERT	2.4
8	EDWARD	2.8	CHARLES	2.6	CHARLES	2.4	CHARLES	2.1
9	JAMES	2.3	ARTHUR	2.0	WALTER	2.0	JAMES	2.0
10	ALBERT	2.0	JAMES	1.9	ARTHUR	1.8	WALTER	1.9
	TOP 10	38.5		34.6		32.8		31.7

South Region

	1900		1910		1920		1930	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	JOHN	8.3	JOHN	7.2	JOHN	6.6	JOHN	5.4
2	WILLIAM	5.6	WILLIAM	5.0	WILLIAM	3.7	WILLIAM	3.5
3	GEORGE	3.7	JOSEPH	3.5	JOSEPH	3.1	JAMES	2.5
4	JOSEPH	3.5	GEORGE	3.1	GEORGE	3.1	JOSEPH	2.4
5	HENRY	3.3	CHARLES	3.0	FRANK	2.9	FRANK	2.3
6	CHARLES	3.2	HENRY	2.6	HENRY	2.5	GEORGE	2.3
7	JAMES	3.1	FRANK	2.6	JAMES	2.4	CHARLES	2.1
8	FRANK	2.7	JAMES	2.4	CHARLES	2.3	ROBERT	1.7
9	EDWARD	2.2	LOUIS	2.0	ROBERT	2.0	HENRY	1.6
10	THOMAS	2.1	ROBERT	1.8	EDWARD	1.8	JOSE	1.6
	TOP 10	37.7		33.1		30.3		25.3

West Region

	1900		1910		1920		1930	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	JOHN	7.5	JOHN	6.8	JOHN	6.2	JOHN	5.7
2	WILLIAM	6.1	WILLIAM	5.4	WILLIAM	3.9	WILLIAM	3.9
3	GEORGE	4.2	GEORGE	4.0	GEORGE	3.6	GEORGE	3.1
4	JAMES	3.7	FRANK	3.1	FRANK	3.0	FRANK	2.9
5	CHARLES	3.5	CHARLES	2.9	JOSEPH	2.7	ROBERT	2.8
6	FRANK	3.3	JAMES	2.7	JAMES	2.6	JAMES	2.4
7	JOSEPH	3.0	JOSEPH	2.5	CHARLES	2.4	JOSEPH	2.2
8	HENRY	2.6	EDWARD	2.3	ROBERT	1.8	CHARLES	2.1
9	EDWARD	2.2	HENRY	2.1	HENRY	1.8	EDWARD	1.8
10	THOMAS	2.2	ALBERT	1.8	ALBERT	1.7	HENRY	1.7
	TOP 10	38.2		33.5		29.6		28.5

Notes: IPUMS: 1900 (5%) and 1910, 1920, 1930 (1%).
Region variable is defined into the following 1990 regional and divisional classification system:
Northeast Region (New England Division: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont; Middle Atlantic Division: New Jersey, New York, Pennsylvania). Midwest Region (East North Central Division: Illinois, Indiana, Michigan, Ohio, Wisconsin; West North Central Division: Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota). South Region (South Atlantic Division: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia; East South Central Division: Alabama, Kentucky, Mississippi, Tennessee; West South Central Division: Arkansas, Louisiana, Oklahoma, Texas). West Region (Mountain Division: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming; Pacific Division: Alaska, California, Hawaii, Oregon, Washington).

Table A.8: Popularity of names by region and census year - Females

Females

Northeast Region

	1900		1910		1920		1930	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	MARY	12.2	MARY	10.4	MARY	9.5	MARY	8.7
2	ANNIE	4.4	ANNA	3.6	ANNA	4.7	ANNA	4.9
3	MARGARET	2.8	MARGARET	3.4	HELEN	3.7	HELEN	4.0
4	ELIZABETH	2.6	ANNIE	3.2	MARGARET	3.2	MARGARET	2.7
5	ANNA	2.5	ELIZABETH	2.7	ELIZABETH	2.4	ROSE	2.5
6	SARAH	1.9	HELEN	2.5	ROSE	2.2	ELIZABETH	2.1
7	EMMA	1.9	CATHERINE	1.6	ANNIE	1.8	JOSEPHINE	1.7
8	CATHERINE	1.5	SARAH	1.5	CATHERINE	1.7	FLORENCE	1.5
9	ALICE	1.5	ALICE	1.5	JOSEPHINE	1.6	DOROTHY	1.5
10	LIZZIE	1.4	ROSE	1.5	FLORENCE	1.5	CATHERINE	1.5
	TOP 10	32.6		31.8		32.3		31.1

Midwest (formerly North Central) Region

	1900		1910		1920		1930	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	MARY	8.3	MARY	6.8	MARY	6.7	MARY	6.2
2	ANNA	4.1	ANNA	4.4	ANNA	4.3	ANNA	4.1
3	EMMA	3.2	EMMA	2.4	HELEN	3.2	HELEN	3.5
4	ANNIE	2.6	CLARA	2.1	MARGARET	2.1	MARGARET	2.2
5	CLARA	2.2	ANNIE	1.8	EMMA	1.7	ROSE	1.8
6	MINNIE	1.9	MARGARET	1.7	ELIZABETH	1.6	DOROTHY	1.7
7	IDA	1.6	HELEN	1.7	CLARA	1.6	RUTH	1.6
8	ELIZABETH	1.5	ELIZABETH	1.7	RUTH	1.4	ELIZABETH	1.6
9	LIZZIE	1.4	MINNIE	1.5	FLORENCE	1.3	MARIE	1.5
10	MARTHA	1.3	MARTHA	1.4	ALICE	1.3	LILLIAN	1.4
	TOP 10	28.1		25.4		25.2		25.5

South Region

	1900		1910		1920		1930	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	MARY	9.7	MARY	8.8	MARY	8.0	MARY	6.4
2	ANNIE	4.0	ANNIE	3.2	ANNIE	3.1	HELEN	2.2
3	ANNA	2.2	ANNA	2.1	ANNA	2.3	ANNA	2.1
4	EMMA	2.2	EMMA	1.9	HELEN	1.9	MARIA	2.1
5	ELIZABETH	1.6	ELIZABETH	1.8	MARGARET	1.6	ELIZABETH	1.5
6	MAGGIE	1.5	MINNIE	1.5	ELIZABETH	1.5	MARGARET	1.5
7	SARAH	1.5	MARGARET	1.4	EMMA	1.5	MARIE	1.4
8	MINNIE	1.4	ROSA	1.3	MINNIE	1.4	ANNIE	1.2
9	KATIE	1.4	MARIE	1.2	JOSEPHINE	1.3	JOSEPHINE	1.1
10	MARGARET	1.4	IDA	1.1	MARIE	1.3	ROSE	1.1
	TOP 10	26.8		24.3		23.9		20.5

West Region

	1900		1910		1920		1930	
Rank	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**	**Name**	**Perc.**
1	MARY	8.3	MARY	6.4	MARY	6.1	MARY	6.1
2	ANNIE	2.9	ANNA	2.5	HELEN	2.4	MARGARET	2.3
3	ANNA	2.1	ANNIE	2.0	ANNA	2.2	ANNA	1.9
4	EMMA	2.0	ALICE	1.9	MARGARET	2.2	HELEN	1.8
5	MARGARET	1.6	ELIZABETH	1.9	ALICE	1.7	DOROTHY	1.6
6	ALICE	1.6	MARGARET	1.8	RUTH	1.6	RUTH	1.6
7	ELIZABETH	1.4	CLARA	1.6	ELIZABETH	1.5	JOSEPHINE	1.3
8	SARAH	1.4	EMMA	1.6	ANNIE	1.3	ROSE	1.2
9	CLARA	1.3	HELEN	1.5	EMMA	1.3	ELIZABETH	1.2
10	MINNIE	1.2	RUTH	1.3	DOROTHY	1.2	ALICE	1.1
	TOP 10	23.7		22.4		21.5		20.1

Notes: IPUMS: 1900 (5%) and 1910, 1920, 1930 (1%).
Region variable is defined into the following 1990 regional and divisional classification system:
Northeast Region (New England Division: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont; Middle Atlantic Division: New Jersey, New York, Pennsylvania). Midwest Region (East North Central Division: Illinois, Indiana, Michigan, Ohio, Wisconsin; West North Central Division: Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota). South Region (South Atlantic Division: Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia; East South Central Division: Alabama, Kentucky, Mississippi, Tennessee; West South Central Division: Arkansas, Louisiana, Oklahoma, Texas). West Region (Mountain Division: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming; Pacific Division: Alaska, California, Hawaii, Oregon, Washington).

Appendix B. American Names and Measures of Immigrant Assimilation

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 variables.[30]

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:

$$Y_{ibect} = \alpha AmericanName_{ibect} + \beta' X_{ibect} + \gamma' \phi_{ct} + \psi_t + \lambda_b + \delta_e + \epsilon_{ibect} \quad \text{(B.1)}$$

where Y_{ibect} is the outcome variable for immigrant i born in country b, entering in the US in year e, living in county c in census year t, X_{ibect} is a vector of individual characteristics, ϕ_{ct} is a vector of county characteristics (which includes county fixed effects in one of the specifications), λ_b are the birthplace fixed effects, δ_e are the cohort of entry fixed effects (five year interval fixed effects from 1831-1935 up to 1926-1930), ψ_t are year effects and ϵ_{ibct} is an i.i.d. (independent and identically distributed) error term. $AmericanName_{ibect}$ 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.

[30] In a recent paper, Biavaschi, Giulietti and Siddique (2017) 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.

We show the main results for male immigrants in Table B.1 (for female immigrants in Table B.2) 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. 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. We present results not only for the whole sample, but also for the subsamples of immigrants from Germany, Italy and the former USSR.

Appendix B.1. Economic Assimilation

Columns (1) to (3) in Table B.1 display the results for male immigrants for three economic assimilation measures: log occupational score, labor force participation, 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 0.028 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 2% of one standard deviation). Our estimates also suggest that adopting an American name may be negatively associated with full-year employment for Germans. The estimates also suggest that adopting an American name has a positive and significant association with the labor force participation, especially if the immigrant is from the former USSR. Again, qualitatively speaking, the association is modest.

Table B.2 shows the results for female immigrants, where adopting an American name is not significantly associated with the three economic assimilation measures.

Appendix B.2. Cultural Assimilation

In table B.1, Column (4) shows that for males 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 2.1 percentage points higher for those who have an American name (an increase of around 60% in this probability).

For females, the point estimate of 1.8 percentage points for females is more than 3 times the corresponding figure for males (0.5 percentage points). The baseline probability of marrying an American man is 11.4%, resulting in an increase of around 15% in this probability.

In table B.1, Column (5) indicates that adopting an American name has a positive association with the ability to speak English, in particular if the male immigrant is Italian or from the former USSR. 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 71.0% and 75.1%, respectively).

For females, the point estimate is very similar also reaching almost 3 percentage points if the female immigrant is Italian or from the former USSR.

Appendix B.3. Civic Assimilation

In terms of civic assimilation, measured by the acquisition of US citizenship, Column (6) shows that having an American name is associated with an increase in the probability of a male immigrant 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 female immigrants, this association is not robust to the inclusion of county fixed effects.

Table B.1: Assimilation Outcomes - Male Immigrants

	Log Occupational Score	Labor Force Participation	Full-year Employed Last Year	American Wife (exc. Second generation)	Speaks English	US Citizenship
Total						
American Name	0.005*	0.412**	-0.075	0.53***	2.39***	0.938***
	(0.003)	(0.160)	(0.003)	(0.189)	(0.236)	(0.253)
Observations	227,866	238,560	154,555	146,247	236,972	227,469
R-squared	0.244	0.065	0.069	0.137	0.315	0.399
Germany						
American Name	-0.008*	0.945***	-0.982**	0.113	0.133	0.414
	(0.004)	(0.282)	(0.004)	(0.371)	(0.338)	(0.451)
Observations	75,049	79,035	64,542	55,407	78,524	75,840
R-squared	0.271	0.085	0.069	0.190	0.147	0.355
Italy						
American Name	0.008**	0.364	0.507	1.290***	3.0***	0.337
	(0.004)	(0.306)	(1.000)	(0.244)	(0.642)	(0.598)
Observations	31,940	33,329	16,691	18,669	33,074	31,605
R-squared	0.158	0.106	0.113	0.175	0.322	0.291
Former USSR						
American Name	0.028***	1.780**	1.440	0.263	2.89***	2.93***
	(0.008)	(0.791)	(1.400)	(0.291)	(1.06)	(0.95)
Observations	26,159	27,485	15,163	17,348	27,368	25,482
R-squared	0.293	0.119	0.110	0.191	0.338	0.381
Fixed Effects						
Year	YES	YES	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES	YES	YES
County	YES	YES	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES	YES	YES

Notes:
Besides the reported fixed effects, all regressions include as controls: Years in the US in cubic form, Age in cubic form, and Log Population.
Robust standard errors in parentheses clustered by county. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.2: Assimilation Outcomes - Female Immigrants

	Log Occupational Score	Labor Force Participation	Full-year Employed Last Year	American Husband (exc. Second generation)	Speaks English	US Citizenship
Total						
American Name	-0.007	0.117	0.120	1.76***	2.49***	0.399
	(0.012)	(0.223)	(0.193)	(0.279)	(0.313)	(0.533)
Observations	31,260	175,802	38,846	128,070	173,598	59,772
R-squared	0.298	0.146	0.101	0.192	0.301	0.332
Germany						
American Name	-0.013	-0.337	-0.181	1.62***	0.578	-0.809
	(0.022)	(0.399)	(0.243)	(0.526)	(0.361)	(0.845)
Observations	10,194	65,402	15,354	46,375	64,536	10,273
R-squared	0.205	0.156	0.118	0.227	0.179	0.443
Italy						
American Name	-0.004	0.537	0.452	0.384	2.90***	0.707
	(0.021)	(0.613)	(0.601)	(0.312)	(0.992)	(1.16)
Observations	2,651	18,721	3,558	15,167	18,460	10,533
R-squared	0.254	0.151	0.134	0.104	0.279	0.183
Former USSR						
American Name	0.059	1.09	0.146	1.04	2.49*	0.767
	(0.036)	(0.862)	(0.953)	(0.642)	(1.35)	(1.82)
Observations	4,359	22,175	6,308	16,151	21,949	10,016
R-squared	0.291	0.263	0.154	0.186	0.329	0.255
Fixed Effects						
Year	YES	YES	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES	YES	YES
County	YES	YES	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES	YES	YES

Notes:
Besides the reported fixed effects, all regressions include as controls: Years in the US in cubic form, Age in cubic form, and Log Population.
Robust standard errors in parentheses clustered by county. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix C. Appendix Tables

In this part of the appendix, we provide additional tables for more details. Table C.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 C.2. Table C.3 presents the summary statistics for the pooled sample and for Italians, Germans and immigrants from the former USSR, for the period 1900-1930. This table shows only the results for male immigrants. Table C.4 shows the summary statistics for the period 1900-1910. The results are in line with Table C.3.

Table C.5 gives the summary statistics of assimilation outcomes for the sample as well as for three most represented groups of immigrants by birthplace. Table C.6 presents the summary statistics for the children sample (sons and daughters) for the second generation, i.e. only children born in the US.

In Table C.7 we add a dummy variable for each enumeration district, which is an area covered by a single enumerator, to control for the effects of enumerators.

In Table C.8, we examine to what extent urban and rural immigrants responded to different incentives when deciding to adopt American names. The labor market opportunities, the social networks, and the nature of trade are different between farming rural communities and industrial urban areas. In Table C.8, we repeat the exactly the same exercise as in Table 6, but we run separate regressions for urban and rural immigrants. An immigrant is defined to be an urban immigrant if he lives in a city or a place of more than 2,500 inhabitants. In some counties, there are both urban and rural immigrants. To calculate county-level aggregate variables, we use only the urban inhabitants of a county when we consider urban immigrants and only the rural inhabitants of the county when we look at rural immigrants. In both cases the most important determinants of American name are years in the US and the concentration index. Nevertheless, there are some differences in the way social networks matter for name adoption between rural areas (where 72% of immigrants adopt American first names) and urban areas (where 79% of immigrants adopt American first names). The concentration index has a strong impact on name adoption in rural but, perhaps surprisingly, smaller in urban areas. This indicates that in rural areas, where immigrants are perhaps more isolated, the presence of an immigrant community with the same origin may delay assimilation through the choice of an American

name. Concerning rural immigrants, the coefficient on the native labor force participation rate is statistically significant, and its magnitude is much larger than the corresponding estimate for the whole sample, reported in Table 6. One caveat behind our findings is that the decision whether to live in a rural area or an urban area is potential endogenous.

Tables C.9 – C.11 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 C.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 C.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)
Economic Variables (by geographical unit)	
Immigr. unemp. rate	Immigrants' unemployment rate by geographical unit (county)
Native unemp. rate	US-born unemployment rate by geographical unit (county)
Immigr. labor force participation rate	Immigrants Labor force participation rate by geographical unit (county).
Native labor force participation rate	US born Labor force participation rate by geographical unit (county).
Immigr. log occ. score	Immigrants' average log of occupational score by geographical unit (county)
Native log occ. score	US-born average log of occupational score by geographical unit (county)
Network Variables (by geographical unit)	
Concentration index	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
Share of immigrants	Share of immigrants between 16 and 65 in total population aged 16-65 by geographical unit (county)
Other Outcome Variables	
Log occupational score	Log of occupational score
Labor Force Participation	Dichotomous variable indicating whether the immigrant is in the labor force.
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 C.3: Summary Statistics - Male Immigrants (1900-1930)

Variable	Mean (standard deviation)			
	Pooled sample	Germany	Italy	Former USSR
Explanatory Variables				
Years in the US	19.2	25.3	15.5	15.4
	(12.2)	(13.5)	(9.9)	(10.0)
Age	39.3	43.9	36.7	35.8
	(12.3)	(12.6)	(11.4)	(11.6)
Log population	9.20	8.89	9.65	9.94
	(1.87)	(1.85)	(1.62)	(1.69)
Economic Variables (by geographical unit)				
Immigr. labor force participation rate	94.6	94.3	94.9	94.8
	(4.9)	(5.5)	(3.9)	(4.0)
Native labor force participation rate	90.3	90.5	90.3	90.6
	(4.2)	(3.9)	(3.7)	(3.8)
Immigr. log occ. score	3.11	3.09	3.17	3.17
	(0.19)	(0.19)	(0.12)	(0.15)
Native log occ. score	3.15	3.11	3.22	3.21
	(0.21)	(0.23)	(0.13)	(0.17)
Network Variables (by geographical unit)				
Concentration index	7.0	7.9	7.1	7.7
	(7.0)	(6.0)	(4.3)	(6.6)
Share of immigrants	34.7	31.7	36.5	39.1
	(14.5)	(15.0)	(13.8)	(14.5)

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 C.2.

Table C.4: Summary Statistics - Male Immigrants (1900-1910)

Variable	Mean (standard deviation)			
	Pooled sample	Germany	Italy	Former USSR
Explanatory Variables				
Years in the US	16.8	23.4	9.4	10.21
	(12.3)	(12.4)	(8.1)	(8.0)
Age	37.6	42.5	32.9	32.2
	(12.5)	(12.6)	(10.8)	(10.6)
Log population	8.51	8.57	8.87	9.28
	(1.83)	(1.79)	(1.69)	(1.68)
Economic Variables (by geographical unit)				
Immigr. unemp. rate	20.5	19.8	21.9	22.4
	(11.3)	(11.1)	(10.3)	(9.5)
Native unemp. rate	16.9	16.6	17.6	16.2
	(7.3)	(6.6)	(6.9)	(6.0)
Immigr. labor force participation rate	95.1	94.7	95.7	95.4
	(4.2)	(4.6)	(3.0)	(3.5)
Native labor force participation rate	90.9	90.9	91.4	91.3
	(3.7)	(3.6)	(3.1)	(3.2)
Immigr. log occ. score	3.07	3.07	3.13	3.14
	(0.19)	(0.19)	(0.13)	(0.15)
Native log occ. score	3.10	3.08	3.18	3.19
	(0.22)	(0.23)	(0.15)	(0.17)
Network Variables (by geographical unit)				
Concentration index	8.5	10.0	7.4	8.4
	(7.7)	(6.1)	(5.4)	(7.3)
Share of immigrants	37.8	33.9	40.6	42.2
	(15.0)	(15.2)	(15.0)	(14.7)

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

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

Variables	Mean (standard deviation)					
	Germany		Italy		Former USSR	
	American name	Non-American name	American name	Non-American name	American name	Non-American name
Log occupational score	3.100	3.106	3.123	3.103	3.226	3.149
	(0.412)	(0.425)	(0.318)	(0.320)	(0.380)	(0.385)
Full-year employed last year (%)	85.0	86.1	80.5	78.4	81.2	80.6
	(35.7)	(34.6)	(39.6)	(41.1)	(39.0)	(39.6)
Labor Force Participation (%)	92.8	92.1	94.7	94.4	94.1	92.8
	(25.8)	(27.1)	(22.3)	(23.0)	(23.5)	(25.9)
American wife (%)	9.2	8.2	3.4	1.3	1.3	0.7
	(28.9)	(27.4)	(18.1)	(11.5)	(11.3)	(8.5)
Speaks English (%)	95.3	94.7	79.8	71.0	83.6	75.1
	(21.1)	(22.3)	(40.1)	(45.4)	(37.0)	(43.2)
US citizenship (%)	75.9	72.8	40.0	34.6	46.5	39.6
	(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 C.6: Summary Statistics - Child Immigrants - Second Generation

Variables	Mean (standard deviation)	
	Sons	Daughters
Explanatory Variables		
Father with American Name	79.4	79.5
	(40.4)	(40.4)
Immigrant mother	77.5	77.8
	(41.8)	(41.6)
Siblings (aged 0-10)	2.0	2.0
	(1.4)	(1.4)
Child's age	5.0	5.0
	(3.1)	(3.1)
Years in the US (father)	19.7	19.7
	(9.1)	(9.0)
Father's age	39.2	39.3
	(7.9)	(7.8)
Log population	8.50	8.52
	(1.82)	(1.82)

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

Table C.7: American Name (LPM) - Robustness Check - Enumerator District

	(1)	(2)
Years in the US	0.749***	0.732***
	(0.100)	(0.105)
(Years in the US)^2 /10	-0.224***	-0.215***
	(0.029)	(0.033)
(Years in the US)^3 /100	0.023***	0.023***
	(0.004)	(0.004)
Age	0.131	0.057
	(0.240)	(0.270)
Age^2 /10	-0.084	-0.058
	(0.061)	(0.069)
Age^3 /100	0.008	0.005
	(0.005)	(0.006)
Log population	-0.581	-0.564
	(0.656)	(0.732)
Immigr. labor force participation rate	-2.939	-4.452
	(2.976)	(3.879)
Native labor force participation rate	-1.226	-2.377
	(4.060)	(5.231)
Immgr. log occ. score	2.034	2.616
	(1.664)	(2.222)
Native log occ. score	-2.200	-0.955
	(1.955)	(2.672)
Concentration index	-0.196***	-0.194***
	(0.039)	(0.040)
Share of Immigrants	0.022	0.012
	(0.030)	(0.040)
Observations	238,560	238,560
R-squared	0.159	0.350
Fixed Effects		
Year	YES	YES
Birthplace	YES	YES
County	YES	YES
Cohort of Entry group	YES	YES
Enumerator District	-	YES
Joint Hypothesis Test		
Economic Variables	0.97	0.81
(p-value)	(0.423)	(0.518)
Network Variables	13.29	12.29
(p-value)	0.000	0.000

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930). The Enumerator District comprises 29539 categories.

Table C.8: American Name (LPM) - Male Immigrants - Urban vs. Rural Areas

	Urban	Rural
Years in the US	0.725***	0.683***
	(0.117)	(0.181)
(Years in the US)^2 / 10	-0.241***	-0.164***
	(0.033)	(0.055)
(Years in the US)^3 / 100	0.025***	0.017**
	(0.004)	(0.007)
Age	0.081	0.148
	(0.291)	(0.438)
Age^2 / 10	-0.081	-0.062
	(0.074)	(0.114)
Age^3 / 100	0.008	0.005
	(0.006)	(0.009)
Log population	-1.059*	0.085
	(0.548)	(0.768)
Immigr. labor force participation rate	-0.228	-2.058
	(4.517)	(3.384)
Native labor force participation rate	-11.39**	-1.131
	(5.360)	(4.378)
Immgr. log occ. score	-3.471	2.550
	(2.793)	(1.786)
Native log occ. score	-0.235	-1.028
	(3.860)	(2.009)
Concentration index	-0.140***	-0.208***
	(0.051)	(0.039)
Share of Immigrants	0.0001	0.0103
	(0.043)	(0.033)
Observations	158,958	78,958
R-squared	0.138	0.211
Fixed Effects		
Year	YES	YES
Birthplace	YES	YES
County	YES	YES
Cohort of Entry group	YES	YES

Notes:
The table reports the effects in percentage points.
Robust standard errors in parentheses clustered by county.
*** Significant at 1%; ** Significant at 5%; * Significant at 10%.
Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table C.9: American Name (LPM) - Sons and Daughters (German Father)

	Sons		Daughters	
Father with American Name		6.203***		0.0333
		(0.766)		(0.391)
Immigrant Mother		-0.607		0.0629
		(0.611)		(0.459)
Siblings (aged 0-10)		-0.0234		-0.0229
		(0.198)		(0.134)
Child's Age		-0.106		0.884**
		(0.769)		(0.447)
Child's Age^2		0.104		-0.141
		(0.190)		(0.0995)
Child's Age^3		-0.00783		0.00749
		(0.0127)		(0.00631)
Years in the US (father)	1.091***	1.017***	0.266	0.230
	(0.371)	(0.361)	(0.229)	(0.227)
Years in the US^2 (father)	-0.276**	-0.255**	-0.130*	-0.116
	(0.127)	(0.124)	(0.074)	(0.073)
Years in the US^3 (father)	0.032**	0.030*	0.021**	0.020**
	(0.016)	(0.015)	(0.008)	(0.008)
Age (father)	1.972	1.809	0.504	0.262
	(1.469)	(1.468)	(1.097)	(1.128)
Age^2 (father)	-0.481	-0.448	-0.091	-0.045
	(0.352)	(0.350)	(0.261)	(0.267)
Age^3 (father)	0.037	0.035	0.005	0.002
	(0.028)	(0.027)	(0.020)	(0.021)
Log population	-0.429	-0.406	1.569	1.587
	(2.058)	(2.009)	(1.007)	(0.998)
Immigr. labor force participation rate	1.654	0.332	3.039	3.189
	(9.436)	(9.415)	(4.861)	(4.864)
Native labor force participation rate	-9.732	-10.18	-10.99	-11.02
	(10.19)	(10.21)	(6.990)	(6.968)
Immgr. log occ. score	-4.048	-3.387	1.670	1.582
	(4.954)	(4.944)	(2.693)	(2.672)
Native log occ. score	-0.629	-0.428	-4.911	-5.016
	(5.083)	(5.147)	(3.253)	(3.246)
Concentration Index	0.113	0.133	-0.049	-0.049
	(0.121)	(0.120)	(0.073)	(0.073)
Share of Immigrants	0.065	0.069	-0.009	-0.009
	(0.074)	(0.074)	(0.050)	(0.050)
Observations	33,827	33,827	32,170	32,170
R-squared	0.087	0.093	0.071	0.072
Fixed Effects				
Year	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES
County	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table C.10: American Name (LPM) - Sons and Daughters (Italian Father)

		Sons		Daughters
Father with American Name		7.534***		2.968***
		(0.776)		(0.601)
Immigrant Mother		-3.814***		-2.213***
		(0.943)		(0.735)
Siblings (aged 0-10)		0.450*		-0.0244
		(0.237)		(0.173)
Child's Age		-0.807		1.408**
		(0.792)		(0.689)
Child's Age^2		0.134		-0.348**
		(0.209)		(0.159)
Child's Age^3		-0.00568		0.0232**
		(0.0142)		(0.0104)
Years in the US (father)	0.736*	0.607	0.620	0.618
	(0.419)	(0.423)	(0.407)	(0.416)
Years in the US^2 (father)	-0.043	-0.047	-0.270	-0.290
	(0.167)	(0.169)	(0.182)	(0.186)
Years in the US^3 (father)	0.001	0.004	0.030	0.035
	(0.025)	(0.025)	(0.025)	(0.026)
Age (father)	-2.953	-3.428	-0.432	-0.616
	(2.438)	(2.314)	(1.581)	(1.592)
Age^2 (father)	0.608	0.744	0.068	0.133
	(0.595)	(0.569)	(0.388)	(0.394)
Age^3 (father)	-0.040	-0.051	-0.003	-0.009
	(0.047)	(0.045)	(0.031)	(0.032)
Log population	0.142	0.754	-0.850	-0.490
	(1.961)	(2.021)	(1.663)	(1.578)
Immigr. labor force participation rate	0.898	5.541	-1.062	-0.616
	(16.51)	(16.13)	(9.942)	(9.952)
Native labor force participation rate	-10.51	-14.87	-6.225	-7.039
	(19.62)	(18.73)	(15.39)	(15.53)
Immgr. log occ. score	-1.505	-2.015	-8.140	-8.364
	(8.027)	(7.932)	(5.426)	(5.354)
Native log occ. score	-18.23*	-17.11	4.727	4.596
	(10.64)	(10.61)	(9.133)	(9.024)
Concentration Index	0.123	0.125	0.109	0.135
	(0.307)	(0.305)	(0.207)	(0.207)
Share of Immigrants	-0.010	-0.014	-0.015	-0.011
	(0.129)	(0.127)	(0.115)	(0.111)
Observations	13,326	13,326	12,747	12,747
R-squared	0.055	0.065	0.053	0.057
Fixed Effects				
Year	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES
County	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table C.11: American Name (LPM) - Sons and Daughters (Former USSR Father)

	Sons		Daughters	
Father with American Name		2.566		2.057*
		(1.647)		(1.168)
Immigrant Mother		1.868		1.224
		(1.725)		(0.832)
Siblings (aged 0-10)		-0.263		-0.192
		(0.354)		(0.214)
Child's Age		0.641		0.423
		(1.008)		(0.689)
Child's Age^2		-0.195		-0.00778
		(0.258)		(0.165)
Child's Age^3		0.0132		-0.000345
		(0.0175)		(0.0107)
Years in the US (father)	-0.263	-0.193	0.674*	0.545
	(0.476)	(0.476)	(0.364)	(0.349)
Years in the US^2 (father)	-0.001	-0.003	-0.289*	-0.245*
	(0.208)	(0.206)	(0.153)	(0.146)
Years in the US^3 (father)	-0.002	-0.002	0.045**	0.043**
	(0.032)	(0.032)	(0.023)	(0.022)
Age (father)	-4.553	-4.191	0.115	0.112
	(2.800)	(2.789)	(1.196)	(1.213)
Age^2 (father)	0.118	0.110	-0.044	-0.061
	(0.729)	(0.724)	(0.305)	(0.303)
Age^3 (father)	-0.010	-0.094	0.006	0.008
	(0.062)	(0.061)	(0.025)	(0.025)
Log population	2.102	2.072	0.647	0.717
	(3.494)	(3.559)	(2.551)	(2.554)
Immigr. labor force participation rate	-33.34**	-34.25**	-6.218	-5.816
	(13.64)	(13.71)	(12.28)	(12.53)
Native labor force participation rate	2.955	3.255	-3.351	-2.964
	(11.66)	(11.73)	(10.39)	(10.52)
Immgr. log occ. score	0.760	1.003	9.458	9.684
	(8.943)	(9.023)	(6.501)	(6.418)
Native log occ. score	-8.523	-8.141	-15.51***	-14.86***
	(5.843)	(5.802)	(5.470)	(5.428)
Concentration Index	0.0269	0.0211	-0.180	-0.185
	(0.170)	(0.169)	(0.116)	(0.117)
Share of Immigrants	-0.247*	-0.255**	-0.030	-0.037
	(0.128)	(0.127)	(0.097)	(0.095)
Observations	11,040	11,040	10,738	10,738
R-squared	0.077	0.078	0.078	0.081
Fixed Effects				
Year	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES
County	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Appendix D. Using the Name Index Approach

For robustness we use an alternative measure of the American first name, based on the name index approach proposed by Fryer and Levitt (2004). Specifically, we define the "American Name Index" (ANI) to be

$$\text{ANI}_{name,t} = \frac{\text{Prob}(name|\text{US born}, t)}{\text{Prob}(name|\text{US born}, t) + \text{Prob}(name|\text{Immigrant}, t)} \times 100,$$

where *name* is a particular first name and t reflects the census year. The index ranges from 0 to 100, takes the value 100 if all individuals who have a specific name are US born, and is equal to 0, if only immigrants have a particular name.

Table D.1 shows the mean and standard deviation of the American Name Index (ANI) by year of census and country of birth, for female and male immigrants. In the remaining part of the appendix, Tables D.2, D.3, D.4, D.5, D.6, D.7, D.8 and D.9 report estimation results using the name index approach by repeating the exercises reported in Tables B.1, B.2, 8, 9, 10, 11, 12 and 13, respectively. We find that the main qualitative results remain intact. See each table for details.

Table D.1: American Name Index (ANI) for Female and Male Immigrants - By Year of Census and Country of Birth

	Mean(standard deviation)	
	Female	Male
Pooled Sample	43.8	40.7
	(4.6)	(18.5)
By year		
1900	43.8	40.8
	(3.6)	(18.6)
1910	43.7	40.3
	(4.6)	(19.3)
1920	44.0	40.8
	(4.7)	(18.3)
1930	43.8	40.9
	(5.0)	(18.0)
By Country of Birth		
Germany	43.5	45.7
	(3.2)	(16.8)
Italy	44.3	38.6
	(3.3)	(18.3)
Former USSR	43.7	40.7
	(3.4)	(16.0)
Poland	45.6	44.3
	(3.5)	(15.1)
Sweden	44.1	40.6
	(5.3)	(19.4)
Mexico	30.1	28.6
	(5.3)	(18.4)
Norway	42.7	34.9
	(6.1)	(22.2)
Hungary	46.1	42.8
	(3.8)	(14.4)
Czechoslovakia	44.4	45.6
	(3.9)	(13.2)
Denmark	42.6	35.8
	(8.2)	(22.6)
Greece	46.7	43.5
	(10.8)	(20.6)
France	45.6	46.5
	(7.6)	(18.6)
Japan	18.6	7.1
	(10.3)	(18.1)
China	26.9	17.1
	(18.0)	(21.1)
Portugal	44.7	37.6
	(4.1)	(15.6)
Spain	41.3	33.5
	(10.7)	(18.6)

Notes: Pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%). Immigrants between 16 and 65 years old.

Table D.2: Assimilation Outcomes - ANI - Male Immigrants

	Log Occupational Score	Labor Force Participation	Full-year Employed Last Year	American Wife (exc. Second generation)	Speaks English	US Citizenship
Total						
American Name Index (ANI)	0.018***	0.001	0.009	0.030***	0.047***	0.016***
	(0.006)	(0.004)	(0.007)	(0.004)	(0.007)	(0.005)
Observations	227,866	238,560	154,555	146,247	236,972	227,469
R-squared	0.244	0.065	0.069	0.137	0.315	0.399
Germany						
American Name Index (ANI)	0.020*	0.021***	-0.006	0.019**	0.011	0.002
	(0.011)	(0.008)	(0.009)	(0.008)	(0.008)	(0.011)
Observations	75,049	79,035	64,542	55,407	78,524	75,840
R-squared	0.271	0.085	0.069	0.191	0.147	0.355
Italy						
American Name Index (ANI)	0.033***	-0.006	0.014	0.044***	0.068***	0.040***
	(0.009)	(0.008)	(0.019)	(0.008)	(0.014)	(0.014)
Observations	31,940	33,329	16,691	18,669	33,074	31,605
R-squared	0.158	0.106	0.113	0.176	0.322	0.291
Former USSR						
American Name Index (ANI)	-0.009	0.012	0.045**	0.010	-0.012	0.012
	(0.012)	(0.009)	(0.018)	(0.007)	(0.016)	(0.018)
Observations	26,159	27,485	15,163	17,348	27,368	25,482
R-squared	0.293	0.119	0.110	0.192	0.338	0.381
Fixed Effects						
Year	YES	YES	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES	YES	YES
County	YES	YES	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES	YES	YES

Notes:
Besides the reported fixed effects, all regressions include as controls: Years in the US in cubic form, Age in cubic form, and Log Population.
Robust standard errors in parentheses clustered by county. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table D.3: Assimilation Outcomes - ANI - Female Immigrants

	Log Occupational Score	Labor Force Participation	Full-year Employed Last Year	American Husband (exc. Second generation)	Speaks English	US Citizenship
Total						
American Name Index (ANI)	-0.226	0.030	-0.003	0.072**	0.168***	-0.021
	(0.139)	(0.028)	(0.022)	(0.029)	(0.033)	(0.040)
Observations	31,260	175,802	38,846	128,070	173,598	59,772
R-squared	0.298	0.146	0.101	0.191	0.301	0.332
Germany						
American Name Index (ANI)	0.190	0.138*	-0.125	0.029	0.084	-0.206
	(0.531)	(0.072)	(0.136)	(0.103)	(0.076)	(0.151)
Observations	10,194	65,402	15,354	46,375	64,536	10,273
R-squared	0.205	0.156	0.118	0.226	0.178	0.443
Italy						
American Name Index (ANI)	-0.980	-0.118	0.215	-0.062	0.246*	-0.149
	(0.788)	(0.088)	(0.321)	(0.055)	(0.140)	(0.277)
Observations	2,651	18,721	3,558	15,167	18,460	10,533
R-squared	0.255	0.151	0.134	0.104	0.278	0.183
Former USSR						
American Name Index (ANI)	-1.046	-0.005	0.258	0.073	-0.090	0.010
	(0.758)	(0.105)	(0.298)	(0.086)	(0.152)	(0.255)
Observations	4,359	22,175	6,308	16,151	21,949	10,016
R-squared	0.291	0.263	0.154	0.186	0.329	0.255
Fixed Effects						
Year	YES	YES	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES	YES	YES
County	YES	YES	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES	YES	YES

Notes:
Besides the reported fixed effects, all regressions include as controls: Years in the US in cubic form, Age in cubic form, and Log Population.
Robust standard errors in parentheses clustered by county. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table D.4: American Name Index (ANI) - Male Immigrants - Years in the US

	Years in the US				
	[0-5]	[6-10]	[11-15]	[16-20]	+21
Years in the US	-0.264	-55.68*	138.5	93.14	0.590**
	(1.073)	(31.82)	(95.36)	(157.9)	(0.291)
Years in the US^2 / 10	5.39	76.36*	-110.2	-52.42	-0.102
	(6.86)	(42.95)	(76.64)	(90.00)	(0.078)
Years in the US^3 / 100	-10.4	-34.3*	29.1	9.81	0.007
	(11.3)	(19.1)	(20.4)	(17.1)	(0.007)
Age	-0.326	-0.612**	-0.0543	0.0724	0.163
	(0.297)	(0.272)	(0.265)	(0.267)	(0.371)
Age^2 / 10	0.067	0.105	-0.072	-0.010	-0.010
	(0.009)	(0.076)	(0.073)	(0.069)	(0.083)
Age^3 / 100	-0.006	-0.006	0.009	0.011**	0.009
	(0.008)	(0.007)	(0.006)	(0.005)	(0.006)
Log population	-1.610*	-1.036	0.526	0.492	0.311
	(0.966)	(0.646)	(0.724)	(0.495)	(0.328)
Immigr. labor force participation rate	3.526	-2.691	-7.186	-2.896	1.531
	(7.145)	(4.821)	(5.390)	(3.810)	(1.968)
Native labor force participation rate	5.046	-9.710*	-5.258	-4.223	-3.010
	(7.372)	(5.172)	(6.081)	(4.475)	(2.572)
Immgr. log occ. score	3.692	3.316	0.855	5.062**	-0.205
	(3.225)	(2.484)	(2.605)	(2.202)	(1.018)
Native log occ. score	0.175	4.291	-3.980	0.626	-0.602
	(3.359)	(2.676)	(3.181)	(2.180)	(1.251)
Concentration Index	-0.292***	-0.023**	-0.028***	-0.025***	-0.023***
	(0.058)	(0.009)	(0.008)	(0.007)	(0.005)
Share of Immigrants	-0.029	0.025	0.018	-0.081***	-0.023
	(0.036)	(0.030)	(0.036)	(0.026)	(0.018)
Observations	29,512	42,519	33,973	48,760	83,748
R-squared	0.208	0.151	0.196	0.158	0.153
Fixed Effects					
Year	YES	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES	YES
County	YES	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES	YES

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table D.5: American Name Index (ANI) - Male Immigrants - Birthplace

	(1)	(2)
Years in the US	0.392***	0.387***
	(0.044)	(0.045)
(Years in the US)^2 / 10	-0.095***	-0.094***
	(0.014)	(0.014)
(Years in the US)^3 / 100	0.008***	0.008***
	(0.002)	(0.002)
Age	-0.063	-0.067
	(0.100)	(0.100)
Age^2 / 10	-0.034	-0.033
	(0.026)	(0.026)
Age^3 / 100	0.004*	0.004*
	(0.002)	(0.002)
Log population	0.030	0.019
	(0.279)	(0.287)
Immigr. labor force participation rate	0.384	0.580
	(1.412)	(1.409)
Native labor force participation rate	-3.866**	-3.775**
	(1.762)	(1.770)
Immgr. log occ. score	1.149	1.180
	(0.758)	(0.760)
Native log occ. score	0.186	0.204
	(0.882)	(0.890)
Concentration index	-0.118***	
	(0.014)	
Share of Immigrants	0.011	0.013
	(0.013)	(0.014)
Concentration index$^+$		-0.048**
		(0.021)
Concentration index x Southern		-0.077
		(0.068)
Concentration index x Nordic		-0.126***
		(0.034)
Concentration index x Eastern		-0.036
		(0.038)
Concentration index x Asian		-0.279***
		(0.078)
Concentration index x Mexico		-0.111***
		(0.043)
Observations	238,560	238,560
R-squared	0.137	0.137
Year	YES	YES
Birthplace	YES	YES
County	YES	YES
Cohort of Entry group	YES	YES

Notes: The table reports the effects in percentage points. + Ommited group: Germany. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930). Group of countries: Nordic (Norway, Sweden, and Denmark), Southern (Italy, Greece, France, Portugal, and Spain), Eastern (Czechoslovakia, Hungary, Poland, and former USSR), and Asian (Japan and China).

Table D.6: American Name Index (ANI) - Robustness Check (Unemployment Rate) - Male Immigrants

	(1)	(2)
Years in the US	0.236***	0.295***
	(0.060)	(0.060)
(Years in the US)^2 / 10	-0.069***	-0.095***
	(0.026)	(0.024)
(Years in the US)^3 / 100	0.008**	0.011***
	(0.003)	(0.003)
Age	-0.074	-0.062
	(0.136)	(0.135)
Age^2 / 10	-0.022	-0.025
	(0.035)	(0.035)
Age^3 / 100	0.002	0.003
	(0.003)	(0.003)
Log population	-0.196	0.152
	(0.553)	(0.561)
Immigr. unemp. rate	-0.646	-0.908
	(1.071)	(1.058)
Native unemp. rate	-1.411	-0.712
	(1.759)	(1.731)
Immgr. log occ. score	1.321	0.911
	(1.322)	(1.314)
Native log occ. score	1.496	1.301
	(1.406)	(1.378)
Concentration index	-0.168***	-0.159***
	(0.021)	(0.022)
Share of Immigrants	0.024	-0.023
	(0.187)	(0.190)
Observations	159,305	159,305
R-squared	0.146	0.146
Fixed Effects		
Year	YES	
Birthplace	YES	
County	YES	YES
Cohort of Entry group	YES	YES
Birthplace x Year		YES
Joint Hypothesis Test		
Economic Variables	0.95	0.70
(p-value)	(0.436)	(0.591)
Network Variables	34.26	25.85
(p-value)	(0.000)	(0.000)

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table D.7: American Name Index (ANI) - Robustness Check (Entry in the US: Before and After) - Male Immigrants

	(1)	(2)	(3)	(4)
Years in the US	0.435***	0.308***	0.312***	0.311***
	(0.060)	(0.059)	(0.061)	(0.059)
(Years in the US)^2 / 10	-0.106***	-0.074***	-0.083***	-0.083***
	(0.019)	(0.019)	(0.019)	(0.019)
(Years in the US)^3 / 100	0.009***	0.008***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Age	-0.086	-0.099	-0.098	-0.134
	(0.104)	(0.104)	(0.105)	(0.107)
Age^2 / 10	-0.031	-0.026	-0.026	-0.018
	(0.027)	(0.027)	(0.027)	(0.028)
Age^3 / 100	0.003	0.003	0.003	0.003
	(0.002)	(0.002)	(0.002)	(0.002)
Log population	-0.0575	0.102		
	(0.289)	(0.247)		
Immigr. labor force participation rate (entry after)	-0.824	-0.461	-0.585	-0.969
	(0.548)	(0.532)	(0.866)	(0.901)
Immigr. labor force participation rate (entry before)	0.0960	1.299**	1.895*	1.650
	(0.638)	(0.622)	(1.056)	(1.132)
Native labor force participation rate	-3.520*	-2.543		
	(1.879)	(1.796)		
Immgr. log occ. score (entry after)	0.921**	1.026**	0.721	0.573
	(0.408)	(0.399)	(0.616)	(0.662)
Immgr. log occ. score (entry before)	-0.557	-0.430	-0.568	-0.302
	(0.450)	(0.445)	(0.646)	(0.667)
Native log occ. score	0.904	0.262		
	(1.021)	(0.959)		
Concentration index (entry after)	-0.166***	-0.151***	-0.165***	-0.110***
	(0.018)	(0.018)	(0.020)	(0.034)
Concentration index (entry before)	-0.069***	-0.111***	-0.124***	-0.079**
	(0.019)	(0.022)	(0.024)	(0.036)
Share of Immigrants (entry after)	0.0198	0.00267	0.286***	0.268***
	(0.015)	(0.016)	(0.051)	(0.055)
Share of Immigrants (entry before)	0.014	-0.063	0.289***	0.281***
	(0.017)	(0.017)	(0.053)	(0.057)
Observations	226,259	226,259	226,259	226,259
R-squared	0.133	0.138	0.159	0.204
Fixed Effects				
Year	YES			
Birthplace	YES			
County	YES	YES		
Cohort of Entry group	YES	YES	YES	YES
Birthplace x Year		YES	YES	YES
County x Year			YES	YES
County x Birthplace				YES
Joint Hypothesis Test				
Economic Variables	2.08	2.28	-	-
(p-value)	(0.053)	(0.034)	-	-
Network Variables	24.71	20.12	-	-
(p-value)	(0.000)	(0.000)	-	-

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table D.8: American Name Index (ANI) - Female Immigrants

	(1)	(2)	(3)	(4)
Years in the US	0.052***	0.002	-0.001	-0.001
	(0.014)	(0.013)	(0.012)	(0.007)
(Years in the US)^2 / 10	-0.015***	0.003	0.001	0.002
	(0.005)	(0.004)	(0.004)	(0.002)
(Years in the US)^3 / 100	0.001	-0.001	-0.0004	-0.0002
	(0.001)	(0.001)	(0.001)	(0.0003)
Age	0.008	0.007	0.016	-0.004
	(0.028)	(0.027)	(0.025)	(0.015)
Age^2 / 10	-0.005	-0.004	-0.007	0.004
	(0.007)	(0.007)	(0.006)	(0.004)
Age^3 / 100	0.0003	0.0004	0.001	-0.0005
	(0.001)	(0.001)	(0.001)	(0.0003)
Log population	0.264	0.392*		
	(0.237)	(0.202)		
Immigr. labor force participation rate	1.306	1.681*		
	(0.923)	(0.896)		
Native labor force participation rate	-0.703	-1.119		
	(1.260)	(1.166)		
Immgr. log occ. score	-0.497**	-0.336		
	(0.235)	(0.222)		
Native log occ. score	0.0433	-0.191		
	(0.416)	(0.400)		
Concentration index	-0.042**	-0.063***	-0.063***	-0.022
	(0.017)	(0.016)	(0.016)	(0.034)
Share of Immigrants	-0.010	-0.019		
	(0.013)	(0.013)		
Observations	168,631	168,631	168,631	168,631
R-squared	0.235	0.276	0.397	0.757
Fixed Effects				
Year	YES			
Birthplace	YES			
County	YES	YES		
Cohort of Entry group	YES	YES	YES	YES
Birthplace x Year		YES	YES	YES
County x Year			YES	YES
County x Birthplace				YES
Joint Hypothesis Test				
Economic Variables	1.73	1.74	-	-
(p-value)	(0.141)	(0.139)	-	-
Network Variables	3.65	9.75	-	-
(p-value)	(0.026)	(0.000)	-	-

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).

Table D.9: American Name Index (ANI) - Sons and Daughters

	Sons		Daughters	
Father with American Name		9.258***		2.072***
		(0.619)		(0.505)
Immigrant Mother		-2.871***		-2.585***
		(0.274)		(0.227)
Siblings (aged 0-10)		-0.645***		-0.678***
		(0.0752)		(0.0663)
Child's Age		-0.264		0.247
		(0.261)		(0.222)
Child's Age^2		0.034		-0.075
		(0.063)		(0.053)
Child's Age^3		-0.004		0.003
		(0.004)		(0.003)
Years in the US (father)	0.488***	0.604***	0.211*	0.344***
	(0.120)	(0.121)	(0.112)	(0.117)
Years in the US^2 / 10 (father)	-0.045	-0.106**	0.014	-0.043
	(0.051)	(0.051)	(0.044)	(0.045)
Years in the US^3 / 100 (father)	-0.002	0.005	-0.007	-0.001
	(0.007)	(0.007)	(0.006)	(0.006)
Age (father)	-0.305	0.570	-0.255	0.592
	(0.508)	(0.512)	(0.453)	(0.445)
Age^2 / 10 (father)	-0.001	-0.158	0.002	-0.153
	(0.124)	(0.124)	(0.112)	(0.111)
Age^3 / 100 (father)	0.003	0.012	0.002	0.011
	(0.010)	(0.010)	(0.009)	(0.009)
Log population	0.230	0.251	0.752	0.796
	(0.664)	(0.644)	(0.556)	(0.530)
Immigr. labor force participation rate	-0.0625	-0.141	-0.603	-0.635
	(3.632)	(3.595)	(2.886)	(2.909)
Native labor force participation rate	-7.453	-7.807*	1.254	0.787
	(4.663)	(4.712)	(3.240)	(3.245)
Immgr. log occ. score	-0.465	-1.105	4.668***	4.289***
	(1.883)	(1.870)	(1.486)	(1.471)
Native log occ. score	-0.081	-0.232	-2.683	-2.669
	(1.843)	(1.840)	(1.648)	(1.644)
Concentration Index	-0.059*	-0.042	-0.063***	-0.052**
	(0.033)	(0.033)	(0.022)	(0.022)
Share Immigrants	0.014	0.016	-0.018	-0.014
	(0.030)	(0.030)	(0.020)	(0.020)
Observations	94,440	94,440	90,953	90,953
R-squared	0.127	0.135	0.131	0.137
Fixed Effects				
Year	YES	YES	YES	YES
Birthplace	YES	YES	YES	YES
County	YES	YES	YES	YES
Cohort of Entry group	YES	YES	YES	YES

Notes: The table reports the effects in percentage points. Robust standard errors in parentheses clustered by county. *** Significant at 1%; ** Significant at 5%; * Significant at 10%. Cohort of Entry group: 5 year group intervals fixed effects (groups from 1831-1835 up to 1926-1930).