Labour Economics 62 (2020) 101778

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Please call me John: Name choice and the assimilation of immigrants in the United States, 1900–1930☆

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

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

ARTICLE INFO

Keywords:
Americanization
Culture
First name
Identity
Immigration JEL classification Code
J15
N32

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.

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, 20–30% 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

☆ 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).

* Corresponding author.
E-mail address: hfreis@bportugal.pt (H. Reis).

[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.

https://doi.org/10.1016/j.labeco.2019.101778
Received 7 August 2018; Received in revised form 15 November 2019; Accepted 15 November 2019
Available online 16 November 2019

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 examining 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 (1994) and Lazear (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 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 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 oth-

[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.

[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.

[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.

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

2.1. 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.

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]

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 Por-

[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/).

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

[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 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%).

tuguese, 12.4% are called John, 12.0% are Joseph and 7.1% are Frank. Interestingly, there are only two American names (Joe and John) among the top 10 names for Mexicans, suggesting that they had little need for a name change. Names such as Joseph, John and Frank are completely non-existent in Italian or Portuguese cultures. These are very distinctively American names.

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

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

Table 3 presents for both, males and females, the percentage of American names according to our definition by year and by immigrants' birthplace, respectively. Starting with males, in the pooled sample, the percentage of immigrants with an American name is about 77%, being relatively constant between 1900 and 1930. Nevertheless, there is significant heterogeneity across countries, with this percentage ranging from

Table 2
Top 10popular 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%).

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]

Fig. 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

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

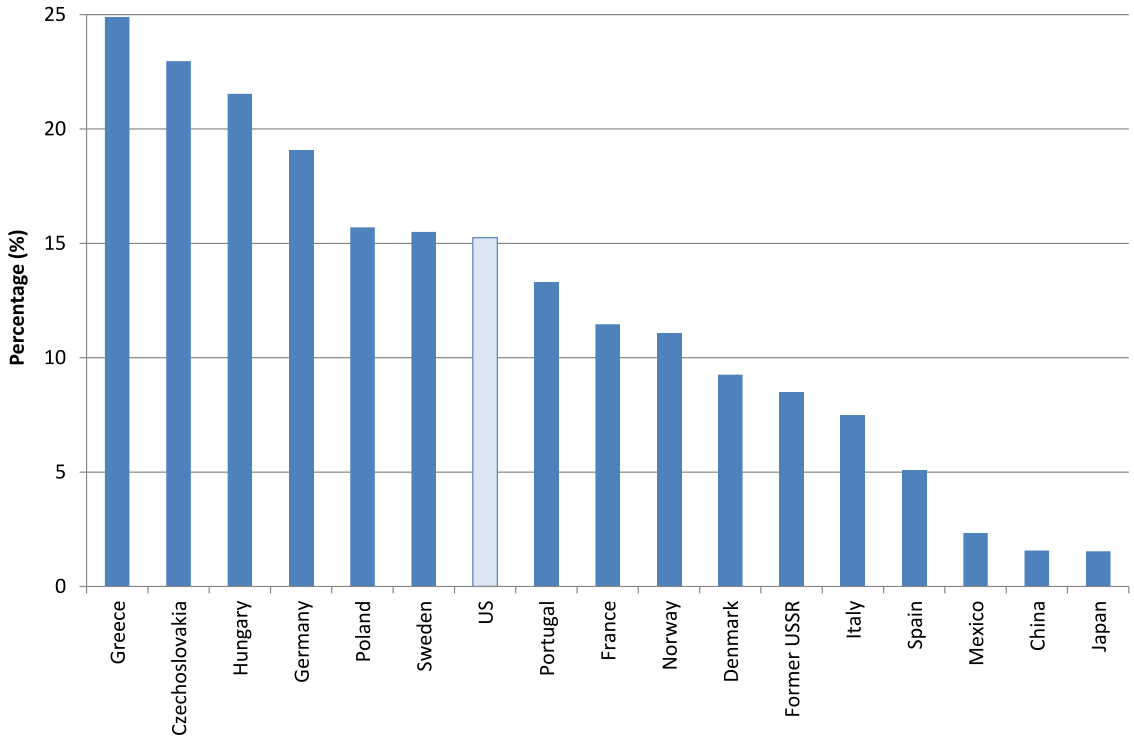

Fig. 1. Male immigrants - Top three American names. Notes: IPUMS pooled sample - weighted by census sample size: 1900 (5%) and 1910, 1920, 1930 (1%). Top three American names, John, William and George, are reported in Table A.2 - Column (2) from IPUMS.

Table 3
Percentage offemale 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.

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 Fig. 2, which shows the percentage of male Italian immigrants with American first names by number of years in the US.

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

Fig. 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 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.

One caveat behind Fig. 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.

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

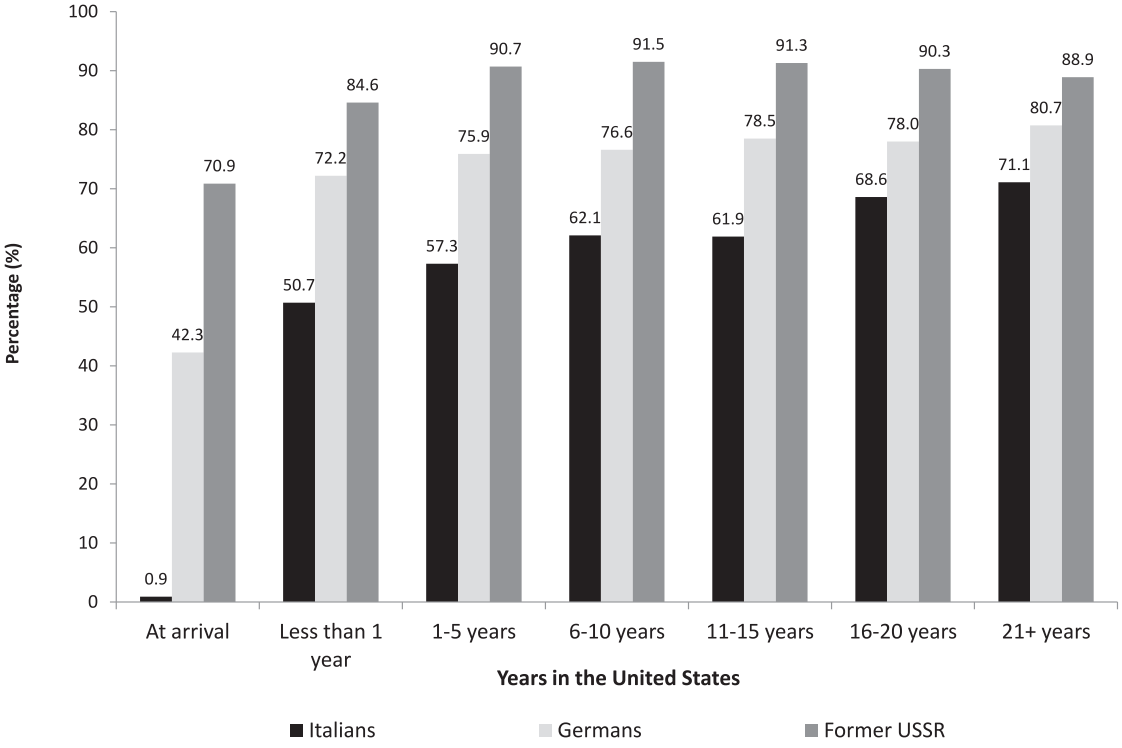

Fig. 2. Male immigrants with American first names. Notes: At arrival: US National Archive data - Italians to America Passenger Data File, 1855–1900; Germans to America Passenger Data File, 1850–1897; Former USSR immigrants to America Passenger Data File, 1834–1897 (Archives, 1977–2002). For each country 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.

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 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

[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 (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.

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%.

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.

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 in-

[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.

Table 5
Fathers' and sons'/daughters' naming patterns.

		Whole Sample					
		Father			Father		
		Non-American name Percentage	American name			Non-American name Percentage	American name
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%).

tegration 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), Biavaschi 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).

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

[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).

[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.

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 to 1835 up to 1926–1930).

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). 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 some-

[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.

[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.

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 to 1835 up to 1926–1930).

one 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 Eq. (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 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 Eq. (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

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

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 unobserv-

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

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

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 to 1835 up to 1926–1930).

able 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 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 Eq. (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 238,560 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

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

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

[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.

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 to 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 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 to 1835 up to 1926–1930).

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 to 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.

[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.

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 to 1835 up to 1926–1930).

4.1. Main results - Male immigrants

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

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 to 1835 up to 1926–1930).

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 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),

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 to 1835 up to 1926–1930).

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.

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.

[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.

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).

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 Eq. (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 Eq. (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 Eq. (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 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.

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.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.labeco.2019.101778 .

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. Am. Econ. Rev. 102 (5), 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. J. Dev. Econ. 102, 2–14.

Abramitzky, R., Boustan, L., Eriksson, K., 2014. A nation of immigrants: assimilation and economic outcomes in the age of mass migration. J. Political Econ. 122 (3), 467–506.

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

Akerlof, G., Kranton, R., 2000. Economics and identity. Q. J. Econ. 115 (3), 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. Centre for Economic Policy Research.

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

National, 1977–2002. Archives and Records Administration 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.

Bandiera, O., Rasul, I., Viarengo, M., 2013. The making of modern america: migratory flows in the age of mass migration. J Dev Econ 102, 23–47.

Benabou, R., Tirole, J., 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96 (3), 1652–1678.

Bertrand, M., Mullainathan, S., 2004. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. Am. Econ. Rev. 94 (4), 991–1013.

Biavaschi, C., Giulietti, C., Siddique, Z., 2017. The economic payoff of name americanization. J. Labor Econ. 35 (4), 1089–1116.

Card, D., 2011. Origins of the unemployment rate: the lasting legacy of measurement without theory. Am. Econ. Rev. 101 (3), 552–557.

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

Charness, G., Gneezy, U., 2008. What'S in a name? anonymity and social distance in dictator and ultimatum games. J. Econ. Behav. Org. 68 (1), 29–35.

Cook, L.D., Logan, T.D., Parman, J.M., 2014. Distinctively black names in the american past. Explor. Econ. Hist. 53, 64–82.

Fryer, R., Levitt, S., 2004. The causes and consequences of distinctively black names. Q. J. Econ. 119 (3), 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. Am. Sociol. Rev. 81 (1), 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 (2), 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. J. Political Econ. 107 (6), S95–S126.

Manning, A., Roy, S., 2010. Culture clash or culture club? national identity in britain. Econ. J. 120 (542), F72–F100.

Olivetti, C., Paserman, M.D., 2015. In the name of the son (and the daughter): intergenerational mobility in the united states, 1850-1940. Am. Econ. Rev. 105 (8), 2695–2724.

Rubinstein, Y., Brenner, D., 2014. Pride and prejudice: using ethnic-sounding names and inter-ethnic marriages to identify labour market discrimination. Rev. Econ. Stud. 81 (1), 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. Soc. Sci. Hist. 18 (2), 169–209.