

NBER WORKING PAPER SERIES

CULTURAL ASSIMILATION DURING THE AGE OF MASS MIGRATION

Ran Abramitzky  
Leah Platt Boustan  
Katherine Eriksson

Working Paper 22381  
<http://www.nber.org/papers/w22381>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
July 2016, Revised September 2017

We are grateful for the access to Census manuscripts provided by Ancestry.com, FamilySearch.org and the Minnesota Population Center. We benefited from the helpful comments we received at the DAE group of the NBER Summer Institute, the Munich “Long Shadow of History” conference, the Irvine conference on the Economics of Religion and Culture, the Cambridge conference on Networks, Institutions and Economic History, the AFD-World Bank Migration and Development Conference, and the Economic History Association. We also thank participants of seminars at Arizona State, Berkeley, Columbia, Michigan, Ohio State, Stanford, UCLA, UCSD, Warwick, Wharton, Wisconsin and Yale. We benefited from conversations with Cihan Artunc, Sascha Becker, Hoyt Bleakley, Davide Cantoni, Raj Chetty, Dora Costa, Dave Donaldson, Joe Ferrie, Price Fishback, Avner Greif, Eric Hilt, Naomi Lamoreaux, Victor Lavy, Joel Mokyr, Kaivan Munshi, Martha Olney, Luigi Pascali, Santiago Perez, Hillel Rapoport, Christina Romer, David Romer, Jared Rubin, Fabian Waldinger, Ludger Woessmann, Gavin Wright, and Noam Yuchtman. David Yang provided able research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Cultural Assimilation during the Age of Mass Migration  
Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson  
NBER Working Paper No. 22381  
July 2016, Revised September 2017  
JEL No. J15,N32

**ABSTRACT**

Using two million census records, we document cultural assimilation during the Age of Mass Migration, a formative period in US history. Immigrants chose less foreign names for children as they spent more time in the US, eventually closing half of the gap with natives. Many immigrants also intermarried and learned English. Name-based assimilation was similar by literacy status, and faster for immigrants who were more culturally distant from natives. Cultural assimilation affected the next generation. Within households, brothers with more foreign names completed fewer years of schooling, faced higher unemployment, earned less and were more likely to marry foreign-born spouses.

Ran Abramitzky  
Department of Economics  
Stanford University  
579 Serra Mall  
Stanford, CA 94305  
and NBER  
ranabr@stanford.edu

Katherine Eriksson  
Department of Economics  
University of California, Davis  
One Shields Avenue  
Davis, CA 95616  
and NBER  
kaeriksson@ucdavis.edu

Leah Platt Boustan  
Princeton University  
Industrial Relations Section  
Louis A. Simpson International Bldg.  
Princeton, NJ 08544  
and NBER  
lboustan@princeton.edu

## I. Introduction

Immigration has emerged as a sharply divisive issue in both the United States and the European Union. Skepticism about whether new arrivals can assimilate into American society was a key concern in the 2016 presidential election and remains an ongoing theme in the public debate on immigration policy.<sup>1</sup> This controversy is not new. American voters have expressed repeated waves of hostility toward immigrants and today's concerns echo alarms sounded often in the past. Consider the following statement: Immigration "is bringing to the country people whom it is very difficult to assimilate and who do not promise well for the standard of civilization in the United States." The speaker was not Donald Trump on the campaign trail but Massachusetts Sen. Henry Cabot Lodge in 1891. Congress eventually passed strict immigration quotas in the early 1920s, putting an end to the first Age of Mass Migration.<sup>2</sup>

One of the central issues in the political debate over immigration is cultural assimilation: Are immigrants able to successfully integrate into American society by adopting the economic, social, and cultural norms of native-born Americans? Or are they likely to remain an alien presence long after they settle? Many commentators express opinions on the subject, but relatively little empirical evidence is available on how fully and quickly immigrants assimilate into US culture.

Measuring cultural assimilation is a challenge because data on many cultural practices – things like food, dress, and accent – are not systematically collected. But information on one aspect of culture, the names that parents choose for their children, has been recorded in both the past and the present, offering a revealing window into the cultural assimilation process.

This paper studies the cultural assimilation of immigrants during two waves of mass migration to the United States, the first from Europe (1850-1913) and the second (1990-today) from Asia and Latin America. Using five million census records from 1920 and 1940, and nearly

---

<sup>1</sup> Donald Trump first gained traction as the 2016 Republican presidential nominee by declaring that immigration has a detrimental effect on American culture. Trump proposed building a wall on the US-Mexico border, claiming that Mexican immigrants were prone to crime (Lind, 2015; Posner, 2015). After a series of attacks by Islamic extremists, first in Paris and then in Orlando, Trump called for a ban on immigration from Muslim countries. Likewise, the British vote to withdraw from the European Union and the rise of far-right candidate Marine Le Pen in France were driven, in part, by concerns about immigrant assimilation (Salam, 2016; Nossiter, 2017).

<sup>2</sup> Congress passed a literacy test for entry to the US in 1917 and a set of country-specific quotas that favored northern and western European countries in 1921 (modified in 1924). Goldin (1994) reviews the political economy of this legislation.

ten million California birth certificate records between 1989 and 2015, we construct a foreignness index indicating the probability that a given name is held by a foreigner or a native at the time the name was given.<sup>3</sup> In the past, for example, people with names like Hyman or Vito were almost certain to be children of immigrants, while children with names like Clay or Lowell were likely to have native parents.

Names are signals of cultural identity. The premise of this paper is thus that we can trace the assimilation process by examining changes in the names immigrants gave their offspring as they spend more time in the US. Unlike inter-marriage, which requires finding a willing spouse in the native population, naming is a pure choice for immigrant parents, unconstrained by discrimination on the part of natives.<sup>4</sup> Furthermore, giving a child an American-sounding name is a way of identifying with US culture at little financial cost, which provides a tool for assimilation that may be equally available for the rich and poor. At the same time, immigrants may be inclined to give their children ethnic names in order to retain their original cultural identity. Our measure captures how the tradeoff between assimilating into American culture versus retaining original identity changes with time spent in the US; these changes may reflect learning about US culture, identifying more with US culture, or a decision to stay in the US rather than return home.

We find that, in the early twentieth century, immigrants were less likely to give their children foreign-sounding names as they spent more time in the US.<sup>5</sup> After 20 years in the US, immigrants erased half of the gap in name choice with the native born. The shift in name choices happened at a roughly equal pace for sons and daughters, and somewhat faster for families with an illiterate household head. The pace of cultural assimilation varied substantially by country of

---

<sup>3</sup> Lieberman (2000) is the classic reference in the sociology of naming and includes some discussion of the name choices of immigrant parents in the US. Naming patterns have been used as a measure of social distance in a number of contexts. Zelinsky (1970) and Lieberman and Bell (1992) study differences in name choices by region and by parental education, respectively. On African-American naming practices, see Lieberman and Mikelson (1995) and Fryer and Levitt (2004); on Hispanics, see Sue and Telles (2007); and on immigrants to Europe, see Algan, et al. (2013) and Gerhards and Hans (2009).

<sup>4</sup> Moreover, unlike marriage, which typically happens only once, parents often had the opportunity to select names for many children born earlier or later in their immigration process.

<sup>5</sup> Similar to today, the 1900s and 1910s was a period of rapid in-migration, during which around 15 percent of the population was foreign born. Rates of assimilation may have been faster (or slower) when rates of immigration dropped after the border closed.

origin. Italians and other Eastern and Southern Europeans, who were particularly criticized by politicians for not trying to assimilate, were actually among the fastest to adopt American-sounding names in the 1900s and 1910s. Immigrants from Scandinavian countries had slower rates of name-based assimilation. The fact that parents did not fully adopt native naming patterns within the first generation highlights the value that immigrants place on maintaining their cultural identity, a pattern consistent with Akerlof and Kranton's (2000) and Benabou and Tirole's (2011) theories of the economics of identity and Bisin and Verdier's (2000) model of cultural transmission within families.<sup>6</sup>

The observed shifts in naming choices with parental time in the US are not simply capturing cultural naming practices that vary by rank in the birth order. The relationship between names and parental years in the US is robust to controlling for birth order directly and to controlling for sharing a parent's name, which is a more common trait of the eldest child. Furthermore, we find no association between birth order and name foreignness in contexts where birth order is not correlated with parental years in the US, such as for children who were born abroad or for children of third-generation or higher parents.

Having an American-sounding name was a marker of assimilation that may have conferred economic and social benefits. We examined the census records of over 800,000 children of immigrants, observed both in 1920, when they lived with their childhood families, and in 1940 as adults. Indeed, children with less-foreign-sounding names completed more years of schooling, earned more, and were less likely to be unemployed than their counterparts with more foreign-sounding names. In addition, children with less-foreign-sounding names were less likely to marry someone born abroad or with a foreign-sounding name.

However, the association between name foreignness and adult outcomes largely disappears when we compare brothers raised in the same household. This pattern suggests that the association between name foreignness and outcomes in the full population is driven by the selection of households that choose to use foreign names. Families that were quicker to assimilate, as proxied by choosing native-sounding names, raised sons with a different life trajectory. This finding is consistent with Fryer and Levitt (2004), who show that, after controlling for family background, having a highly black name is not associated with lower

---

<sup>6</sup> Jia and Persson (2016) extend and apply these theories to the choice of child's ethnicity in mixed marriages in China.

levels of education or earlier childbearing in a sample of African-American women born in California.

Other measures reinforce the picture of early twentieth century immigrants gradually taking on American cultural markers. By 1930, more than two-thirds of immigrants had applied for citizenship and almost all reported they could speak some English. A third of first-generation immigrants who arrived unmarried and more than half of second-generation immigrants wed spouses from outside their own (or parental) country of origin.

We compare the rates of cultural assimilation in the past and the present by examining name-based assimilation in California during the past twenty five years. California is a large immigrant receiving state – in 1990, California housed one third of the foreign-born population living in the US – and provides researchers with individual birth records with information on the names of both mothers and children. Over the past three decades, the US has experienced a second wave of mass migration, drawing migrants primarily from Asia and Latin America. Unlike in the early twentieth century, the number and characteristics of legal immigrants is highly regulated today and is accompanied by a large inflow of undocumented immigrants. And yet, contrary to the conventional wisdom that immigrants are slower to converge to US cultural norms today, we find that immigrants in California in the 1990s and 2000s shift away from foreign names at a similar rate as European immigrants in the early twentieth century, erasing one-third of the naming gap with natives after twenty years in the US. As in the past, there are substantial differences in the pace of assimilation by country of origin. Mexican immigrants have the fastest rate of assimilation, followed by the Vietnamese. Among Mexican families, assimilation is somewhat faster for less educated families.

We emphasize that our paper has no normative implications. That is, we do not wish to imply that immigrants *should* assimilate culturally into the US society. In fact, arguably part of what makes a society flourish is its openness to cultural diversity. Our paper simply documents what immigrants do in practice, observing the choices that immigrants make about their integration into American culture based on the tradeoff between maintaining their cultural identity and assimilating into the dominant society. We also acknowledge that our estimates only capture assimilation into the general “average” (and often white) society and they do not capture the multi-dimensional nature of culture in the US.

## II. Literature on immigrant assimilation and names as signals of identity

Our paper contributes to growing literatures on immigrant assimilation, ethnic and racial discrimination in the labor market, and the inter-generational transmission of cultural traits.

Studies of immigrant assimilation in economics have mainly focused on labor market outcomes – in particular, whether immigrants’ occupations and earnings converge to those of natives with time spent in the destination.<sup>7</sup> In a recent study, we show that, during the Age of Mass Migration, immigrants moved up the occupational ladder at the same rate as natives, preserving the initial gaps in economic status over time (Abramitzky, Boustan and Eriksson, 2014).<sup>8</sup> However, a lack of labor market assimilation need not foreclose cultural integration and so understanding the speed of cultural assimilation is an independent topic of interest.<sup>9</sup>

There is a long tradition of studying cultural assimilation in sociology, primarily by analyzing the rate of inter-marriage between immigrants and US natives (Gordon, 1964; Lieberson and Waters, 1988; see also Angrist, 2002 and Meng and Gregory, 2005 in economics). Pagnini and Morgan (1990) document high levels of in-group marriage for first-generation immigrants from southern and eastern Europe in 1910 that declined for second and higher generations (Alba and Golden, 1986). Existing work on inter-marriage in this period is hard to interpret because early Censuses do not allow researchers to screen out marriages that occurred in the home country. We improve these measures by using the age at first marriage question in the 1930 Census to focus on immigrants who arrived in the US before marriage.

European immigrants converged with natives in various social behaviors, including completed family size, political participation, and criminality, but this transition often took more than one generation (Watkins, 1994; Guinnane, Moehling and O’Grada, 2006; Foley and

---

<sup>7</sup> See Chiswick (1978), Borjas (1985) and Lubotsky (2007) for discussions of labor market assimilation in the contemporary period and the associated methodological issues. Abramitzky and Boustan (forthcoming) survey the literature on immigration assimilation in US history.

<sup>8</sup> Earlier work on labor market assimilation in the Age of Mass Migration found that immigrants held substantially lower-paid occupations than natives upon first arrival, but that they converged with natives after spending some time in the US (Hatton, 1997; Hatton and Williamson, 1998, chapter 7; Minns, 2000). Differences are primarily due to the use of cross-sectional versus panel datasets.

<sup>9</sup> Indeed, there is only a weak correlation across countries of origin between measured cultural assimilation and the extent of economic assimilation documented in Abramitzky, Boustan and Eriksson (2014, Figure 3). In particular, the rank correlation between changes in occupation-based earnings and changes in name-based assimilation (in Figure 5) is 0.28.

Guinnane, 1999; Shertzer, 2013; Moehling and Piehl, 2009, 2014). A related contemporary literature finds that immigrants draw closer to natives in their political preferences and gender norms, but that some gap remains even in the second generation (Fernandez and Fogli, 2009; Alesina et al., 2011; Luttmer and Singhal, 2011; Blau, et al., 2013). Using parental name choice allows us to trace out the convergence of immigrants to a native norm by time spent in the US, rather than simply across generations.<sup>10</sup>

Our paper also contributes to a growing literature using names to document discrimination against ethnic and racial groups in the labor market. In some contexts, names appear to be used as signals of ethnic or racial identity. One related paper, Goldstein and Stecklov (2015), shows that, during the Age of Mass Migration, men with foreign names faced lower occupation-based earnings in 1930, even after controlling for a proxy for family background. We find that this relationship disappears when comparing brothers within the same family. Other work documents a positive economic return for immigrants who change their own first or last name (Arai and Thoursie, 2009; Biavaschi, Giulietti and Siddique, 2013; Carneiro, Lee and Reis, 2015); negative effects of having a distinctively African-American name in the labor market or the classroom (Bertrand and Mullainathan, 2004; Figlio, 2005); and lower earnings for individuals with identifiably ethnic surnames (Oreopoulos, 2011; Rubinstein and Brenner, 2014). Fryer and Levitt (2004) instead show that black names are not associated with adult outcomes after controlling for family background.

Much of the research on the consequences of distinctive names uses observational data on the names bestowed on children by parents or the names that individuals select for themselves; in this case, names could be correlated with other aspects of family background or with individual motivation. The existing experimental studies look at short-run outcomes, like interview call-back rates. Because race is highly observable, black workers with racially ambiguous names might be rejected by employers later in the hiring process. By comparing siblings born into the same families, we improve the source of identification without sacrificing the use of actual labor market outcomes.

---

<sup>10</sup> Watkins and London (1994) document changes in naming practices for Italians and Jews during the Age of Mass Migration. In related historical work, Hacker (1999) uses name choice as a measure of secularization (primarily among the native born), documenting a decline in biblical names over the nineteenth century and a positive association between biblical naming and family size.



### III. Data and definitions

#### A. Measuring the foreignness of given names

Naming practices provide a useful measure of cultural assimilation. As immigrants spend more time in the US, they learn about US culture, including which names are currently popular among the native born. Immigrants who hope to assimilate into US society might then select more American or “native-sounding” names. Name choices are free from the financial constraint or the discriminatory barriers imposed by natives that might hamper other dimensions of assimilation (e.g., marriage or neighborhood location). High rates of name changing for first generation immigrants – Biavaschi, et al. (2013) show that 30 percent of immigrants in New York City who applied for naturalization by 1930 had Americanized their name – suggest that immigrants were aware of the potential value of native-sounding names.

Historical Census data contain individual records with details on first and last name and country of birth for the full population. The Census Bureau releases these complete manuscripts after 72 years. To develop a systematic measure of name foreignness, we use the newly-digitized complete-count 1920 and 1940 US census to calculate the relative probability ( $R$ ) that a given name was held by a foreigner versus a native by birth cohort. This measure has a natural interpretation; a relative probability of two means that a name is twice as likely to be used in the immigrant population as in the native population, and a relative probability of 0.5 means the name is twice as likely to be found among natives as among immigrants.<sup>11</sup> The foreignness of a name can change over time with shifts in the naming practices of either natives or immigrants. Therefore, to capture the foreignness of a name for a child born in year  $t$ , we calculate the relative probability of the name among individuals in the previous twenty birth cohorts.

The relative probability that a name is given to a child of foreign born parents is sensitive to outliers, especially to names that are unpopular among natives, which results in small values of the denominator. Thus, we also calculate a normalized index used by Fryer and Levitt (2004) in the context of distinctly black names. In particular, the Foreignness Index is defined as:

---

<sup>11</sup> The formula for  $R$  is given by: 
$$R = \frac{\frac{\# \text{foreigners}_{name}}{\text{total} \# \text{foreigners}}}{\frac{\# \text{natives}_{name}}{\text{total} \# \text{natives}}}$$

$$Foreignness\ Index_{name} = 100 \cdot \frac{\frac{\#\ foreigners_{name}}{total\ \# foreigners}}{\frac{\#\ foreigners_{name}}{total\ \# foreigners} + \frac{\# natives_{name}}{total\ \# natives}}$$

and ranges from zero to 100, with a value of zero reflecting the fact that no men in the US with a given first name were foreign born (i.e., a distinctively native name) and a value of 100 assigned to a child whose first name is distinctively foreign. Note that the F-index is a simple function of  $R$ , equivalent to  $R/(1-R)$ .

Table 1 lists the most foreign, neutral and native names for boys and girls in the birth cohorts of 1900-20. The neutral names of this period like Murray and Herman were equally common among the children of foreign-born and native parents. The most foreign names include Italian names like Vito and Mario and Jewish names like Hyman and Isidor. Some of the very native names are surnames used as first names, like Clay and Lowell, which was a particularly American tradition.

The first panel of Figure 1 graphs the Foreignness Index for all sons of immigrant fathers and native fathers born between 1850 and 1920.<sup>12</sup> In the earliest birth cohorts, the sons of immigrants received names with an average F-index of 50, while the sons of natives received names with an average F-index of 40. Both series trend downward slightly from 1850 to 1900. Starting with the birth cohort of 1900 and coinciding with a shift in sending countries toward Southern and Eastern Europe, immigrant parents chose increasingly foreign-sounding names for their sons, leading the F-index for the sons of immigrants to increase from 46 in 1900 to 53 in 1915. The swing in immigrant naming practices after 1900 was large, around the same order of magnitude as the growth of distinctively black names among the African-American community between 1965 and 1980.<sup>13</sup> The gap in the F-index between the children of immigrant and native fathers reached 20 points by 1910.

---

<sup>12</sup> We graph the F-index calculated from the 1920 (rather than the 1940) complete-count Census here so that we can extend the series back to the birth cohort of 1850. The F-index for the birth cohort of 1850 is calculated from individuals born between 1830 and 1849, who were already 71-90 years old in 1920. For younger birth cohorts, the series calculated from the 1920 and 1940 Censuses are nearly identical; any differences would be due to mortality, name changes, or in- or out-migration between 1920 and 1940.

<sup>13</sup> Fryer and Levitt (2004, Figure 3) show that the name of the average black child increased by 12 points on the Distinctively Black Index from 1965 to 1980.

The second panel of Figure 1 demonstrates that naming practices varied by cohort and country of origin. The earliest cohorts of sons born to fathers from Northern and Western Europe were given quite distinct names, but, by later cohorts, naming choices had converged to those of the native born (see the German case in Figure 1). In contrast, the sons of immigrants from Southern and Eastern Europe retained distinctive names in both early and later birth cohorts (see the Italian case in Figure 1). Sons of fathers from the United Kingdom were given names similar to the sons of the native born throughout this period (see the English case in Figure 1). We tested but did not find any evidence for breaks in the use of foreign names following World War I and other key political events in sending countries during the volatile decade of the 1910s.<sup>14</sup>

Figure 2 offers the first evidence of assimilation in naming patterns with time spent in the US. The graph shows the distribution of name foreignness for children living at home in the 1920 Census, either with native-born parents or with foreign-born parents who had been in the US less than (or more than) ten years. The distribution of names bestowed by native parents is shifted to the left, with a mean Foreignness score of 34, dropping off substantially after an index value of 60. For the children of foreign-born parents, the distribution of names given by recent immigrant arrivals (mean = 58) can be easily distinguished from the names given to more long-standing immigrants (mean = 50). Recent immigrants are far more apt to give names with an index value above 60.

## **B. Creating a linked Census sample: 1920-40**

The second part of our paper compares the adult outcomes of children who received more/less foreign names. For this part of the analysis, we create a matched dataset that follows the native-born sons of immigrant fathers from their childhood household in 1920 to the 1940 Census. By creating this matched sample we are able to compare siblings in 1940 who shared the same household in 1920. We link men over time by first and last name, age, and state of birth; details on the main linking procedure are provided in the Data Appendix and we consider

---

<sup>14</sup> For example, German-Americans faced increasing discrimination during World War I, and so German parents might have responded by giving their children less identifiably German names (Moser, 2012). Indeed, there is a trend break in the *Germaness* of names among German families following World War I but there is no trend break in overall foreignness index, suggesting that Germans shifted into names that were heavily used by other foreigners during WWI, rather than names that were commonly used by natives. Fouka (2015, 2017) studies the effect of the War on German naming practices in detail.

alternative matching algorithms in an appendix table below. We restrict our attention to men between the ages of 3 and 15 in 1920, who would have been young enough to be living at home with their parents in 1920 and were of prime labor market age in 1940. Our primary linked sample contains more than 800,000 men, around 300,000 of whom are in sets of matched brothers. We achieve a match rate of 35 percent, which is slightly higher than the standard for historical matched samples (e.g., Ferrie, 1996; Abramitzky, Boustan and Eriksson, 2012).<sup>15</sup>

Men who received a foreign name at birth may have been more likely to change their name in adulthood or to have their name mis-transcribed on a Census form or in the digitization process, any of which would prevent the linking of their records over time. Table 2 compares the men in our matched sample to the full population on a number of baseline characteristics. Indeed, men in the matched sample have less foreign names than in the population, scoring 2.5 points less on the Foreignness Index, or 5 percent below the population mean. Men in the matched sample otherwise resemble the full population; the differences across samples, although sometimes statistically significant in our large sample, represent 2 percent (or less) of the population mean in age, number of siblings, rank in the birth order, and length of first name.

#### IV. Parental name choice in the 1920 Census

This section explores the naming choices of immigrant parents in the complete-count 1920 Census. We estimate assimilation profiles relating parental name choice to time spent in the US, running the following two regression specifications:

$$FIndex_{ij} = \alpha_j + \beta_1 YearsUS_{ij} + \beta_2 X_{ij} + \gamma_{ij} + \varepsilon_{ij} \quad (1)$$

$$FIndex_{ij} = \alpha_j + \beta_1 BirthOrder_{ij} + \beta_2 X_{ij} + \gamma_{ij} + \varepsilon_{ij} \quad (2)$$

where the F-index is the Foreignness Index of the name of child  $i$  in household  $j$  measured at birth. Equation (1) is estimated for children (age 3-15) living in households with a foreign-born household head. The main right-hand side variable is parental years in the US at the date of child  $i$ 's birth ( $YearsUS_{ij}$ ). The regression also includes family fixed effects ( $\alpha_j$ ), which ensures that the

---

<sup>15</sup> Factors that contribute to higher match rates in the 1940 Census include better transcription, a more literate and numerate population able to report their name and age more accurately over time, and improvements in life expectancy. Furthermore, we match a younger sample that would have lower mortality rates than adult samples.

effect of parental time in the US is identified by differences between siblings born after their parents spent more/less time in the US.<sup>16</sup> We control for a set of dummies for child  $i$ 's birth year ( $\gamma_{ij}$ ) in five-year age bands to absorb secular trends in naming, and, in some specifications, we also include characteristics in a vector of controls ( $X_{ij}$ ), including child's rank in the birth order, an indicator for whether he has the same name as his father, a measure of name frequency/commonness, and indicators for whether the name is a saint or biblical name.<sup>17</sup>

Equation (2) replaces parental years in the US with a child's place in the birth order rank among sons (or daughters) observed in the household. Birth order is correlated with parental time in the US in the immigrant sample but is also defined for native households and for immigrant households observed before their move to the US (via children born abroad), allowing us to compare naming patterns across household types. For comparability, we limit our attention to non-black children who were born outside the South because few immigrants lived in the South in 1920. To minimize inaccurate measures of birth order due to the departures of older sibling from the childhood home, we further restrict the sample to children whose mother is less than 43 years old.<sup>18</sup>

Figure 3 illustrates our research design with the example of the Breitenbach family observed in the 1920 Census manuscript. The household head, August, was born in Germany and came to the US in 1904 at the age of 21. In 1920, August and his wife Emma had three sons, Emil (15) and Richard (14) and Edwin (9). Emil (F-index = 62) and Richard (F-index = 42) were born in 1905 and 1906, one and two years, respectively, after their parents arrived in the US, while Edwin (F-index = 19) was born six years later. For the Breitenbach family, six additional years in the US was associated with a 23 point drop in the F-index.

---

<sup>16</sup> There is no variation in parental years in the US across children born at different times in households with a native-born household head. Thus, we cannot include native households in specifications with 'parental years in the US' as our variable of interest. We instead estimate changes in name choice for households with foreign-born heads and compare the magnitude of this change to the size of the immigrant-native name gap in Figure 1.

<sup>17</sup> For an example of ethnic naming traditions that vary by birth order, see <http://homepages.rootsweb.ancestry.com/~cregan/patterns.htm>.

<sup>18</sup> More than 85 percent of 18 year old sons of immigrant are observed living at home with their parents in 1920. Of married foreign-born women who have at least one child, 88 percent have their first child by age 24. With a mother's age restriction of less than 43, we observe the oldest children living at home (and, thus, an accurate birth order) in at least 75 percent of the cases.

Figure 4 reproduces the relationship between a set of dummy variables for parental years in the US and name foreignness, as observed in the Breitenbach family, for all households in the 1920 Census with a foreign-born household head. Consistent with a process of cultural assimilation, we observe that immigrant parents gave both their sons and daughters less foreign names as they spent more time in US, in a linear fashion for the first twenty years and slowing down thereafter. Children born after their parents had spent over 20 years in the US scored 8-10 points lower on the Foreignness Index relative to their siblings born upon their parents' first arrival. The mean gap in the F-Index for the children of immigrants and natives in the 1920 Census was around 20 points, implying that immigrants closed half of this "cultural gap" with natives after spending some time in the US.

It is notable that parents shift their naming behavior with time spent in the US at a roughly equal pace for sons and for daughters. Sociologists have documented that parents are more open to new or creative names for girls, while boys tend to receive a more traditional set of names (Rossi, 1965; Sue and Telles, 2007). In this case, we may have expected a larger shift in daughter's names for a given underlying change in cultural assimilation. However, the cost of having a foreign name may have been larger for sons at the time, due to the potential for labor market discrimination by ethnicity and the historical gender gap in labor force participation.

Table 3 documents that the relationship between parental years in the US and name foreignness is robust to controlling for son's birth order and other features of names; Appendix Table 1 presents similar results for daughters. The first column reproduces the pattern in Figure 4 using a linear specification, wherein each year that a parent spent in the US reduces a child's name foreignness by 0.4 points. Column 2 controls for a son's rank in the birth order. Elder sons were, indeed, given more foreign names and thus including a birth order control reduces the coefficient on parental years in the US by 30 percent but the relationship remains highly significant. Column 3 adds controls for whether the son shared his father's name, as well as the popularity and religious content of the name.<sup>19</sup> Religious names and father's names were more foreign; commonly used names were less foreign. Yet, controlling for these features of names preserves the relationship between parental time in the US and name choice. The last column in

---

<sup>19</sup> We added these controls because we were concerned that some aspects of names are mechanically correlated with the foreignness index in a way that is unrelated to assimilation. However, name controls are themselves outcomes, and so we show specifications with and without these controls.

Table 3 uses the relative probability of the name (the ratio of the probability that a name is given to a foreigner versus a native) as the dependent variable, rather than the F-index. . Each year spent in the US lowers the relative probability of a name by 0.09 points. In 1920, the average child of immigrants had a relative probability measure of 2.3, while the average child of native parents had a relative probability measure of 0.8 (a gap of 1.5 points). By this measure, the naming gap between immigrants and natives would have completely closed after 20 years in the US.

The pace of name-based assimilation was somewhat faster for the poor than the rich. Table 4 shows that illiterate household heads shifted away from foreign names at a faster rate than literate heads, at 0.6 points rather than 0.4 points on the Foreignness index for each year in the US. Immigrants who rented their housing unit engaged in name-based assimilation at the same rate as immigrant homeowners.

Immigrants from sending countries that were culturally distant from the US or that faced high levels of discrimination may have had the largest benefit from name-based assimilation, but they also may have experienced the highest costs of assimilation, in terms of foregoing aspects of their cultural identity. Figure 5 documents that the speed of name-based assimilation differed by country of origin. Immigrants from Southern and Eastern Europe (Portugal, Italy and Russia) exhibited the fastest rates of cultural assimilation. Reassuringly, immigrants from English-speaking countries (England, Scotland, Wales) exhibit no changes in name choice with time spent in the US. There is a strong correlation between average F-index and changes in F-index with years spent in the US; immigrant groups that start out with names that differ the most from natives also undergo the greatest amount of name-based assimilation.<sup>20</sup>

Table 5 considers the speed of assimilation by two proxies for cultural distance from US natives. In each case, we find that immigrants that differ more from the native norm are fastest to engage in cultural assimilation. Our first proxy is an F-index of surnames (analogous to the F-index for given names described in Section III), which can vary both within and between sending countries.<sup>21</sup> Thus, the surname index captures the greater cultural distance of US natives from

---

<sup>20</sup> Wales and Italy, the countries with the lowest and highest initial F-indices, are separated by 25 index points. A 25 point difference in the initial F-index is associated with a change of 15 additional points on the F-index for 20 years spent in the US.

<sup>21</sup> We calculate the F-index of surnames using the 1940 complete-count Census. The sample in Table 5 is smaller than in other tables because only a subset of surnames in the 1920 Census

Slavic countries, say, than from countries with Romance languages, but may also pick up differences in religion or ethnicity within some sending countries (e.g., Protestants and Catholics in Germany). We interact parental years in the US with the surname index and find that households with more foreign last names shift toward native-sounding names more rapidly. Immigrants with very native surnames lose 0.32 points on the F-index for given names per year in the US, whereas immigrants with very foreign surnames instead shift at a rate of 0.48 points per year ( $=.32 + .16$ ).

Our second proxy for cultural distance is residence in an immigrant enclave, as measured by the share of residents in the enumeration district that were born in the family's own country of origin. Again, we find that immigrants living in neighborhoods with their own fellow countrymen (rather than with natives) are faster to assimilate. Immigrants living in areas with no others from their country of origin drop 0.33 points on the F-index per year in the US, whereas immigrants living in an immigrant-dense district instead shift at a rate of 0.40 points per year ( $=.32 + .69*.12$ , where 0.12 is one standard deviation in own-country share). However, we note that the association between living in an enclave and speed of assimilation primarily arises because immigrants from countries that assimilated faster were more likely to live in immigrant-dense areas. Of the 16 sending countries in the data, only three show a significant within-country association between residence in an enclave and assimilation.

Appendix Table 2 demonstrates that the effect of parental years in the US on name foreignness is invariant to alternative measures of the F-index. Results are not sensitive to adjusting the F-index for phonetically-equivalent names, rather than using raw names (for example, treating Roberto and Robert as the same name); fixing the F-Index in 1900, rather than assigning birth-cohort specific indices; calculating the F-index using the 1920 (rather than the 1940) complete-count Census; or using state-specific F-indices to allow for differential name trends by region.<sup>22</sup> Including southern residents in the analysis also leaves results unchanged.

---

match to a surname in 1940. The sample in Table 5 will under-represent families with rare surnames.

<sup>22</sup> Fixing the F-index at a point in time ensures that the results are not driven by a mechanical relationship between birth order and trends in the F-index over time. Any differences in the F-index calculated in 1920 and 1940 would be due to mortality, name changes, or net-migration by birth cohort.



The standard F-index compares the names of all foreigners and all natives, a group that includes the native-born children of immigrant parents. We refine the F-index in two ways, first by creating a country-specific index and then by classifying the children of immigrant parents as “foreign” rather than “native.” Estimated name-based assimilation is 50 percent larger when using a country-specific F-index, which compares the names of immigrants from a particular sending country to the names used by all other US residents. Yet, name-based assimilation with time spent in the US is cut in half if we instead classify names given to second-generation immigrants as “foreign,” leaving only the names of third-generation immigrants (or higher) in the native category. It appears that, with time spent in the US, immigrants shifted away from names commonly used in their own culture but were not particularly drawn to names given by families many generations removed from Europe.

If naming patterns are evidence of cultural assimilation, we should find shifts in name choices for immigrant parents but not for native-born parents or for children who are born before their parents moved to the US. While the parental years in the US measure is only defined for the native-born children of immigrants, birth order rank is a correlated measure that allows us to compare across household types. Figure 6 reports the implied difference in name foreignness between the first and fourth child in the birth order for the children of immigrants and the native born. Relative to their oldest son, immigrant parents gave their fourth-born sons names that were around 4 points lower on the F-index. The estimated change is smaller than in Figure 4, presumably because birth order is correlated with but not a perfect proxy for time spent in the US.

We find a much smaller effect of birth order on name foreignness for the sons and daughters of immigrants who were themselves born abroad, suggesting that the observed pattern is not due to naming traditions imported from the home country. Similarly, the relationship between birth order and name foreignness is an order of magnitude smaller and positive for third-or-higher generation Americans, whose first and fourth-born sons are separated by only 0.4 points on the F-index.

We do find a sizeable effect of birth order on name foreignness for the children of second-generation immigrants. It seems unlikely that second-generation immigrants, who themselves were born in the US, would become more aware of US culture between births. However, second-generation immigrants may continue to adhere to cultural naming traditions by

giving classic “ethnic” names to their first born children, or may be particularly influenced by their own parents in selecting the names of their oldest kids. In this case, we would expect second-generation immigrants to give foreign names to their first-born son but to exhibit no relationship between birth order and name foreignness for higher-order births. In contrast, we expect a more linear effect of birth order on name foreignness for first-generation immigrant arrivals, consistent with the pace of learning about or adapting to US culture.

Appendix Table 3 estimates separate dummy variables for each step in the birth order separately for households with two, three, or four or more sons. For households with two foreign-born parents, each step along the birth order is associated with at least a one point decline in the F-index. In contrast, for households with second- (and third-plus) generation parents, the linear birth order effect is driven by oldest sons. First-born sons score up to 0.7 points higher on the F-index than their younger brothers, but there is no systematic differences between second, third or fourth sons.<sup>23</sup> A similar pattern is observed for children born abroad. We suspect that oldest sons are particularly likely to be named for grandfathers or other family members, even in families that have lived in the US for many generations, and receive more culturally-specific names. Progressive declines in name foreignness down the birth order – as we see for the children of immigrant parents – is instead consistent with a gradual process of cultural assimilation.

Appendix Table 4 subdivides households with foreign-born heads and/or spouses into categories based on age at arrival in the US, inter- vs. intra-marriage, and gender of foreign-born parent (mother vs. father). Name-based assimilation in immigrant households in which at least one parent arrived in the US as a child was not noticeably different than in households in which both parents arrived as adults. Households with immigrant parents from two different countries of origin shift name choices at the same rate as immigrant parents in endogamous marriages. Among households with only one foreign-born parent, having a foreign-born father is associated with faster name-based assimilation than having only a foreign-born mother.

---

<sup>23</sup> In percentage terms, a 0.7 point shift in the F-index is larger for the children of second- and third-generation immigrants, whose mean F-index value is around 35 (compare to 55 for the children of first-generation immigrants). However, it is not clear that a 1 point shift in the F-index is equally meaningful at all spots along the unit interval. Above and below certain thresholds, it may have already been obvious to employers that workers were native- or foreign-born. Thus, we prefer to interpret level shifts, rather than percentage changes in the F-index.

## V. Foreign names and outcomes in the labor and marriage markets: 1940 Census

Immigrant households took steps to assimilate into US culture as they spent more time in the US. In this section, we study whether immigrant children who received a more foreign sounding name had different outcomes in school and/or in labor and marriage markets. We then explore how much of this association is driven by differences within a household and how much by the selection of families that choose to give foreign names to their children.

Foreign names could have negative effects on both schooling and labor market outcomes, even for brothers within the same family, for a number of reasons. First, it could be that children with foreign names were reminded of their ethnic identity and thus perceived lower returns to their education and exerted less effort in school. Second, the negative effect of having a foreign name could be the result of discrimination by teachers or employers who used names as a signal of ethnicity. A similar dynamic may have been at play in the marriage market: men who identified more strongly with their ethnic group may have preferred to find a spouse within their own ethnic community. At the same time, men with more ethnic names may have been overlooked or rejected by native-born spouses from other backgrounds.

In our main specification, we regress the adult outcome of son  $i$  from family  $j$  ( $y_{ij}$ ) on the foreignness of the son's name at birth ( $FIndex_{ij}$ ), controlling for a vector of dummies for the son's birth year ( $\gamma_{ij}$ ):

$$y_{ij} = \alpha_j + \beta_1 FIndex_{ij} + \beta_2 BirthOrder_{ij} + \beta_3 YearsUS_{ij} + \beta_4 \gamma_{ij} + \beta_5 X_{ij} + \varepsilon_{ij} \quad (3)$$

In some specifications, we control for parental years in the US and child's place in the birth order and/or include family fixed effects ( $\alpha_j$ ) in order to compare brothers who were given names with a different foreignness index. The vector of controls  $X$  can also include the F-index of an individual's name in adulthood, which is separately identified by changes in the relative popularity of names over time. Our sample contains sons of foreign-born parents who were between the ages of 3-15 in the 1920 Census, were not born in the South, and can be successfully matched to the 1940 Census. The overall matched sample contains more than 800,000 men, around 300,000 of which are matched brothers.

The Breitenbach family in Figure 3 illustrates the within-household component of this research design. We were able to follow Emil and Richard Breitenbach forward to 1940, at

which point Emil was a machinist with eight years of education earning \$1,600 and Richard was a photo-engraver with nine years of education earning \$2,500.

Figure 7 presents estimates of the relationship between name foreignness and a series of economic and marriage market outcomes. The coefficients underlying the figure are presented in Appendix Tables 5 and 6. Name foreignness is negatively related to educational attainment and employment. A 20 point shift in the F-index (the typical gap between the children of immigrants and natives) is associated with around two months fewer years of schooling and a 0.5 point increase in the probability of unemployment. Yet, the estimated effects of name foreignness on adult outcomes nearly disappear after including family fixed effects.<sup>24</sup> This specification compares brothers who were given names with different ethnic content. Foreign names had a very small effect on educational attainment within brother pairs (20 points of F-index associated with two weeks less schooling) and no effect on unemployment. The population-based estimates in the black bars are picking up other differences between immigrant families that chose foreign or native name for their children, rather than the effect of names themselves.<sup>25</sup>

The ethnic signal of names that parents select for their children at birth can be attenuated (or augmented) as the name becomes more/less popular among certain groups. For example, Nick, one of the most foreign names in the data in 1920, is commonly given by native parents today. More relevant to an employer's perception of a worker's ethnic identity might be the Foreignness Index of his name at the time of labor market entry. In Appendix Table 5, we consider a specification that includes two F-indices on the right-hand side – one calculated at birth and the other at labor market entry (see column 3). By controlling for the F-index at birth,

---

<sup>24</sup> In an earlier version of this paper, we found a persistent association between name foreignness and adult outcomes, even within pairs of brothers (Abramitzky, Boustan and Eriksson, 2016). These results were driven by a few households with co-resident children who were likely not related and could not be easily identified with Census data provided by Family Search (e.g., two or more families living within a single household, or borders or lodgers living with a family). The updated results in this draft are instead based on Census data from the Minnesota Population Center housed at the NBER which provides detailed information on families, even within households, as coded by the variable “serial number.” With more accurate measures, we find little association between name and adult outcomes within brother pairs.

<sup>25</sup> Results are robust to clustering by father's country of origin interacted with state of residence, the level at which one might expect information about names is transmitted. Estimates are little changed by weighting the linked sample so that the distribution of father's country-of-origin matches the full population in 1940. Weighting adjusts for potential variation in match rates by ethnic group.

we can identify the effect of name foreignness based on trends in name popularity over time, which are hard to predict and therefore likely exogenous to family background. We find that F-index at labor market entry is more quantitatively important than the F-index at birth in predicting unemployment, which is suggestive evidence that employers hired or promoted workers differently based on the ethnic content of their name.<sup>26</sup>

Beyond the labor market, having a foreign name may have influenced men's success in the marriage market. Men with foreign names may more closely identify with their own ethnic group and therefore seek out a spouse within their own ethnic community. Alternatively, native-born spouses may discriminate against men who they perceive to be "too foreign." Figure 7 considers two measures of the foreignness of a man's spouse: whether the spouse, herself, was born abroad and the Foreignness Index of spouse's first name, an indicator of either being born abroad or being raised in a less culturally assimilated family in the US.<sup>27</sup> In both cases, we find that men with foreign names are more likely to marry women with a stronger ethnic identity. A 20 point difference in a man's F-index is associated with a 0.2 percentage point increase in the probability of having a foreign-born spouse (on a base of 5.4 percent) and a 1 point increase the F-index of his spouse's name.<sup>28</sup> Yet, as with the labor market outcomes, adding family fixed effects reduces the effect of foreign names considerably and neither relationship remains statistically significant.

Appendix Table 7 considers a series of additional labor market outcomes, including annual earnings and its subcomponents (hourly wages, weeks worked during the year, and hours

---

<sup>26</sup> A conceptually superior specification would also include household fixed effects, in which case identification would come from households with two or more matched brothers whose names follow different trends over a twenty year period. We do not find significant effects of F-index at 20 in this case, likely reflecting the high demands that this specification requires of the data.

<sup>27</sup> Until 1930, the Census asked all respondents about parents' birthplace, which would allow us to classify whether spouses were also second-generation immigrants. However, in the 1940 Census, the question about parental birthplace became a 'sample-line characteristic' asked of only five percent of the population.

<sup>28</sup> Results on spouse characteristics are restricted to the subsample of men who were 25 years or older in 1940 and who were married in that year. Men with a more foreign name are less likely to be married by 1940, but this effect is economically small. 20 points on the F-index is associated with a 1 percentage point decline in the probability of being married (on a base of 68 percent).

worked during the week) and various forms of employment.<sup>29</sup> Consistent with the association between name foreignness and unemployment at the time of the Census, men with more foreign names work less time during the year in both hours and weeks but, conditional on being employed, they do not receive a lower wage. The correlation with employment is equal and opposite to the relationship with unemployment, implying no effect of names on the probability of being out of the labor force in this prime-age sample (the omitted category). Men with foreign names were also more likely to hold a public works job through the New Deal, an indication of under-employment. In all cases, the relationship between name foreignness and economic outcomes disappears when comparing brothers within households. These results are robust to six alternative linking algorithms that increasingly introduce more conservative requirements on what is considered a successful match.<sup>30</sup>

## **VI. Additional measures: Inter-marriage, citizenship and facility with English**

Our name-based measure reveals that immigrants achieved a substantial amount of cultural assimilation in the Age of Mass Migration, which contradicts the view of politicians and contemporary observers that new immigrants were unwilling or unable to assimilate into American culture. We confirm our findings by considering a broader set of measures, including rates of inter-marriage, application for US citizenship and facility with English.

Inter-marriage has been used extensively in sociology as a marker of cultural assimilation. Inter-marriage is a relatively stringent measure of cultural assimilation because it requires not only that immigrants want to integrate into their new society, but also that natives (or members of other immigrant groups) are willing to interact with them (Kalmijn, 1998). Rates of inter-marriage reflect direct preferences over potential spouses as well as the degree of cultural segregation between groups. That is, immigrants may be more likely to marry each other simply because they interact more regularly in segregated neighborhoods or schools.

---

<sup>29</sup> The 1940 Census only contains information on wage and salary income. As a result, results on annual earnings exclude the self-employed. Appendix Table 7 shows that shifts in the F-index have no effect on the probability of self-employment.

<sup>30</sup> See Appendix Table 8. For brevity, we focus on highest grade completed, the one outcome that appears to be significantly related to name foreignness even within households but the patterns presented here are similar for other outcomes. See Abramitzky, et al. (2017) for a detailed description of alternative matching approaches.

We construct the proportion of out of ethnic group marriages for first- and second-generation immigrants by country of origin in 1930, excluding immigrants who were likely married abroad (based on reported immigration year and age at first marriage). In particular, we calculated the share of married immigrants whose spouse was either a first- or second-generation immigrant from the same country of origin; here, we report patterns from the 1930 5 percent IPUMS sample.<sup>31</sup>

Figure 8 presents rates of marriage outside of one's ethnic group for immigrants from 16 sending countries. Panel A reports means from the raw data and Panel B controls for the size and sex ratio within each immigrant group by state (a proxy for the scope of the marriage market). These controls address mechanical differences in the likelihood of meeting a potential spouse from one's own group. The mean probability of out-group marriage rises from 39 percent for the first generation to 68 percent for the second generation (endogamy rates were slightly higher for women; see Appendix Figure 1). The out-group marriage of immigrants who arrived as children (which we term the 1.5 generation) fits between these two values. We also find sizeable variation in the out-group marriage rate across countries of origin. For example, only 11 percent of first-generation immigrants from Italy were married to a non-Italian, compared with 72 percent of first generation immigrants from Scotland. The Finns, Portuguese and Russians also exhibit a strong tendency toward endogamy, while the French and the English are unlikely to marry fellow countrymen. There is considerable persistence in the out-group marriage rate across generations; the within-country correlation between the out-group marriage rates of first and second-generation immigrants is 0.90.

Appendix Figure 2 considers two additional aspects of cultural assimilation: the decision to apply for US citizenship (Panel A) and facility with English (Panel B). Each panel plots the average likelihood of the given activity for first-generation immigrants by gender and country of

---

<sup>31</sup> We focus on 1930 because it was the first Census to add the "age at first marriage" variable. Previous work on inter-marriage in this period (e.g., Pagnini and Morgan, 1990; Angrist, 2002) analyze earlier Censuses and so they cannot separate marriages that occurred in the US from those that occurred prior to migration. If the endogamy rate for the first generation is biased upward by marriages that occurred abroad, the rate of marital assimilation by the second generation will also be overstated. Including marriages that occurred abroad will differentially bias the endogamy rate for country-of-origin groups that tended to migrate in families, rather than as single individuals.

origin, controlling for group size and sex ratio at the state level.<sup>32</sup> Over 75 percent of immigrants from most sending countries had either received citizenship or started the application process by 1930 (at which point, the average immigrant had been in the US for 24.5 years). As with the rates of inter-marriage, applications for US citizenship were lower among Russians and among Portuguese, Italians and Finns, three groups with high rates of return migration to Europe (Gould, 1980). Across all sending countries, women exhibited lower rate of citizenship application than men. This pattern is consistent with the fact that foreign-born women were more likely to marry spouses from their country of origin and were thus ineligible for citizenship via marriage to a US citizen. Immigrants from most sending countries reported near-universal facility with English, although the ability to speak English was somewhat lower for the Portuguese, Italians, and Finns, and for women relative to men. Note that the historical Census measure simply indicates the ability to speak *some* English; unlike today, this question did not ask about language spoken at home or distinguish between levels of English-speaking ability.

Appendix Table 9 presents the pairwise correlation between the various measures of cultural assimilation at the country-of-origin and individual level. The average foreignness of names given to sons born in the US is coded to be negatively correlated with cultural assimilation, while the other measures (out-group marriage, applications for US citizenship and the ability to speak English) are positively correlated with cultural assimilation. These measures are all related with each other in the expected way (e.g., immigrants in an out-group marriage are more likely to be able to speak English and to have applied for US citizenship, and are less likely to select a foreign name for their children). The correlations are all highly significant at the individual level, providing strong validation for using name choice as a measure of cultural assimilation.

## **VII. Parental name choice in the CA birth certificate records: 1989-2015**

A name-based measure of cultural assimilation allows us to compare the speed of assimilation in the past and the present across two waves of mass migration. We use California birth records to observe the names given to children born to foreign and native mothers in the

---

<sup>32</sup> Results look qualitatively similar without these controls. We add controls for the size of the marriage market because citizenship is often achieved through marriage. Furthermore, the return to learning English may be a function of the size of one's ethnic community.



1990s and 2000s. By our metric, we find that the speed of assimilation is nearly the same today as in the past. Immigrants from Mexico and Vietnam exhibit the fastest rates of assimilation, followed by immigrants from China and the Philippines. Within the subset of Mexican migrants, who make up two-thirds of the foreign sample, assimilation is somewhat faster among mothers with less than a college degree.

Politicians and cultural commentators often claim that assimilation is slower today than in the past, asserting that immigrants today— especially low-skilled arrivals from Mexico – are more likely to live in enclaves, avoid learning English and to engage in circular migration between their home country and the US.<sup>33</sup> Our approach is the first (that we know of) to allow for a rigorous comparison of past and present, free from the nostalgia that might distort memories of how quickly ancestors were able to assimilate in the past. Indeed, we find that name assimilation today is remarkably similar to that in the past, with immigrants halving the gap with natives after spending twenty years in the US.

We use the universe of births occurring in the state of California between 1989 and 2015 (excluding 2011, which has incomplete information), a total of more than 10 million records. Nearly five million of these births are to mothers who are born abroad and around 2 million of these are to mothers born in Mexico. Like the historical census records, modern birth certificates record the mother’s birth place. Unlike the historical census records, however, the birth certificates record do not record the mother’s year of arrival to the US. By linking mothers across their births, we make use of the fact that within a mother (i.e. after including mother fixed effect), mother’s age at the time of birth is perfectly correlated with years in the US for foreign born mothers. The birth certificates also do not record the birth place of a mother’s own parents and so we cannot classify native-born mothers into second or third generation immigrants, as in the historical data. Instead, we create a foreignness index of last names and classify native born mothers into those with the most and least foreign last names. We report results for native born mothers with 1<sup>st</sup> quartile and 4<sup>th</sup> quartile last names, as proxies for native born mothers with native born parents or native born mothers with immigrant parents, respectively.

Our historical data is organized at the household level, allowing us to observe the names of all children residing in the household in 1920. We create ‘households’ in the birth certificate

---

<sup>33</sup> Saletan (2016) summarizes the views of Trump and other politicians on immigrant assimilation. Douhat (2013) is an example of cultural criticism.

data by linking mothers over multiple births; we link mothers using mother's first name, maiden name, place of birth, and exact birth date. Although we are only able to observe births that take place in California between 1989 and 2015, we create an accurate measure of birth order by using mother's report of "total children ever born." That is, the first birth observed in our dataset is recorded as a first birth if the mother reports having only one child and is instead recorded as a second birth if the mother reports having two children ever born by that date.

For mothers born abroad, we use two proxies of years in the US at the time of a child's birth: (1) mother's age at time of birth, and (2) child's place in the birth order. By including mother fixed effects in our analysis, we absorb a mother's calendar year of arrival and age at arrival in the US. Within mother, then, years since arrival in the US at time of birth is perfectly correlated with mother's age at time of birth. We assess the validity of this method in the historical data, for which we have both years in the US at time of birth and mother's age at time of birth. Patterns for mother's age at time of birth look nearly identical to Figure 4 (results available upon request).

Figure 9 presents our main analysis of changes in the foreignness index of child's name with mother's age at birth (our proxy for mother's time in the US) in the contemporary data.<sup>34</sup> The figure presents coefficients from a version of equation 1, in which we regress name foreignness on single-year dummy variables for mother's age at birth, indicators for child's year of birth in five year bands, and mother fixed effects. We find that immigrant mothers give children born at older ages (i.e., born after spending more time in US) less foreign-sounding names. After 20 years, mothers eliminate 12 points on the name foreignness index. Likely second generation immigrants (that is, native white mothers in the fourth quartile of the last name index) also eliminate 8 points of the F-index over 20 years. In contrast, we see no such relationship between mother age and the foreignness index for native white mothers who likely have native born parents (first quartile on last name index) or for native black mothers. The average F-index for foreign mothers in the sample is 67 and the average F-index for native-born mothers is 33. Immigrant mothers erase one-third of this 34 point gap after 20 years in the US.

---

<sup>34</sup> As in the historical data, we find similar adjustment speed for the names of sons and daughters with our proxies for time in the US in the modern data and thus report here combined results for all children.

Table 6 instead reports estimated changes in the F-index by child's place in the birth order. Following equation 2, we regress F-index on birth order either linearly or as a series of indicators for place in the birth order. Regressions also include indicators for child's year of birth in five year bands and mother fixed effects. For sons, each place in the birth order reduces the F-index by nearly 0.5 points in immigrant households, resulting in a change of 1.4 points moving from first to fourth in birth order. In contrast, moving one place in the birth order is associated with a small *increase* (less than 0.1 points) for children in households with native-born mothers of likely native parentage. The shift in names with place in the birth order is somewhat smaller for the daughters of foreign-born mothers (0.35 points) and, as above, is not present for the daughters of native-born mothers of likely native parentage.

Table 7 uses the birth order specification to explore heterogeneity in the speed of cultural assimilation by country of origin and education level. We start by considering name assimilation by country of origin for child's first names (Panel A) and middle names (Panel B); anecdotally, immigrants from Asian countries may give their children one foreign-sounding name and a one native-sounding name to aid with the assimilation process.<sup>35</sup> We find that Mexican immigrants are fastest to assimilate, reducing the F-index by more than 0.5 points for each spot in the birth order. This finding is consistent with the cross-country evidence from the past that immigrants from newer sending countries (such as Italy, Russia, Portugal) who tended to be poorer and to have less cultural connection to US natives were fastest to undergo cultural assimilation. We note that our measure of shifts in F-index with time spent in the US may underestimate the assimilation of Asian migrants, who appear to give all of their children native-sounding names (names with low F-index values) regardless of years spent in the US. Asian immigrants may select native-sounding names on the birth certificate for the purpose of social interaction in the US while maintaining an Asian name for household use.<sup>36</sup>

Our results do not support the frequent claims that Mexican migrants avoid assimilation because they live in immigrant enclaves or plan to return to Mexico. We tried a specification that

---

<sup>35</sup> As expected, Chinese and Filipino immigrants are more likely to shift toward native-sounding middle names – rather than first names – with time spent in the US, although this shift is not statistically significant.

<sup>36</sup> Our approach could also underestimate the assimilation of immigrants in the opposite case in which immigrants choose a foreign name on the birth certificate but end up using a native-sounding name in daily life.

interacts birth order and residence in an immigrant enclave, as measured by share of residents in mother's zip code who are foreign born. As in the historical data (reported in Table 5), we find that immigrants living in an immigrant enclave assimilate somewhat faster than immigrants living in more integrated neighborhoods. However, in the California data, this pattern is driven entirely by the country-of-origin mix living in immigrant enclaves. Mexican migrants exhibit the fastest speeds of assimilation and are most likely to live in an enclave. We find no association between speed of assimilation and neighborhood residence within country-of-origin groups.

Panel B of Table 7 subdivides Mexican mothers by highest degree obtained in the data. We find no differences in speed of assimilation by education level, with identical changes in F-index for high school dropouts, high school graduates and mothers with some college, and slightly lower changes for college graduates.

## **VIII. Conclusion**

We study the cultural assimilation of immigrants during two Ages of Mass Migration drawing on historical census data for the early twentieth century and California birth certificate records for today. Both then and now, immigrants chose less foreign names for their sons and daughters as they spent more time in the US, reducing the difference in name choice with natives by one-half in the past and by one-third today after 20 years in the US. There are substantial differences in the pace of name-based assimilation by country of origin, with the immigrant groups most often accused of a lack of assimilation actually assimilating most rapidly. In the past, Italians and Portuguese were among the fastest to assimilate, and today Mexican immigrants have the fastest rate of name assimilation. In both periods, cultural assimilation is somewhat faster for immigrants from lower socio-economic status. The rapid pace of cultural assimilation observed in our names-based measure is consistent with other indicators, including learning to speak English, applying for US citizenship, and marrying spouses from different origins. In the early twentieth century, receiving a "native sounding" name was associated with a series of positive outcomes for the children of immigrants; yet, we find that brothers with more and less foreign names enjoyed similar levels of schooling, earnings, and employment.

Our findings suggest that immigrants' identification with US culture grows stronger with time spent in the country. The gradual adoption of American-sounding names appears to have been part of a process of assimilation through which newcomers learned US culture, made a

commitment to build roots in their adoptive country, and came to identify as Americans. Some immigrants may have arrived with a strong desire to assimilate, but little knowledge of American customs, including popular names preferred by natives. Other immigrants may not have cared about integrating into American society at first, but eventually felt the urge to blend in. Perhaps for both reasons, immigrants appear to navigate the dominant culture with greater ease as time goes by. The naming patterns also highlight the tradeoff that immigrant families face between maintaining their cultural identity and assimilation into society at large. Giving an ethnic-sounding name can enhance self-identification with an ethnic group but, at the same time, this signal of ethnic identity might generate discrimination from teachers and employers. This tradeoff is still salient for immigrants in the US and ethnic minorities around the world today.<sup>37</sup>

Overall though, lessons from the Age of Mass Migration suggest that fears that immigrants cannot or will not fit into American society are misplaced. It would be a mistake to determine immigration policy based on the belief that immigrants will remain foreigners, preserving their old ways of life and keeping themselves at arm's length from the dominant culture. The evidence suggests that over time immigrant populations come to resemble natives, and that new generations form distinct identities as Americans.

---

<sup>37</sup> See, for example, recent articles about the complicated decision of naming children in Asian American and Muslim American communities today (Ramakrishnan, 2015; Ali, 2015).

## References

- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102, no. 5 (2012): 1832-1856.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122, no. 3 (2014).
- Abramitzky, Ran and Leah Platt Boustan. "Immigration in American Economic History." *Journal of Economic Literature*, forthcoming.
- Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "Cultural Assimilation during the Age of Mass Migration." NBER Working Paper 22381, July 2016.
- Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James Feigenbaum, Santiago Perez and Tom Zohar. "Best Practices for Automated Linking Methods using Historical Data." Manuscript 2017.
- Akerlof, George A. and Rachel E. Kranton. "Economics and Identity." *Quarterly Journal of Economics* (2000): 715-53.
- Alba, Richard D. and Reid M. Golden. "Patterns of Ethnic Marriage in the United States." *Social Forces* 65, no. 1 (1986).
- Alesina, Alberto and Paola Giuliano. 2011. "Preferences for Redistribution." In *Handbook of Social Economics*, edited by Alberto Bisin and Jess Benhabib, p. 93-131. Amsterdam: Elsevier.
- Algan, Yann, Thierry Mayer, and Mathias Thoenig. "The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France." Manuscript, 2013.
- Ali, Wajahat. "For Muslim Americans: Baby Aiden or Baby Mohammed?" *New York Times* October 3, 2015.
- Arai, Mahmood, and Peter Skogman Thoursie. "Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings." *Journal of Labor Economics* 27, no. 1 (2009): 127-147.
- Benabou, Roland and Jean Tirole. "Identity, Morals and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2011): 805-855.
- Bertrand, Marianne, and Sendhil Mullainathan. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94, no. 4 (2004): 991-1013.

Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique. "The Economic Payoff of Name Americanization." IZA Discussion Paper No. 7725, 2013.

Bisin, Alberto and Thierry Verdier. "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits." *Quarterly Journal of Economics* (2000): 955-988.

Blau, Francine D., Lawrence M. Kahn, Albert Yung-Hsu Liu, and Kerry L. Papps. 2013. "The Transmission of Women's Fertility, Human Capital, and Work Orientation Across Immigrant Generations." *Journal of Population Economics* 26 (2): 405-35.

Borjas, George J. "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics* (1985): 463-489.

Carneiro, Pedro, Sokbae Lee and Hugo Reis. 2015. "Please Call me John: Name Choice and the Assimilation of Immigrants in the United States, 1900-30." Institute for Fiscal Studies Working Paper.

Chiswick, Barry R. "The Effect of Americanization on the Earnings of Foreign-born Men." *Journal of Political Economy* (1978): 897-921.

Citrin, Jack, Donald P. Green, Christopher Muste, and Cara Wong. "Public Opinion Toward Immigration Reform: The Role of Economic Motivations." *Journal of Politics* 59, no. 03 (1997): 858-881.

Douthat, Ross. "When Assimilation Stalls." *New York Times*. April 27, 2013.

Fernández, Raquel, and Alessandra Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics* 1(1): 146-77.

Fetzer, Joel S. *Public Attitudes toward Immigration in the United States, France, and Germany*. New York: Cambridge University Press, 2000.

Foley, Mark C., and Timothy W. Guinnane. "Did Irish Marriage Patterns Survive the Emigrant Voyage? Irish-American Nuptiality, 1880-1920." *Irish Economic and Social History* 26 (1999): 15-34.

Fouka, Vasiliki. 2015. "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I." Manuscript.

Fouka, Vasiliki. 2017. "How do Immigrants Respond to Discrimination? The Case of Germans in the US during World War I." Manuscript.

Figlio, David N. "Names, Expectations and the Black-white Test Score Gap." National Bureau of Economic Research Working Paper No. w11195, 2005.

Fryer, Roland G., and Steven Levitt. "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics* 119, no. 3 (2004): 767-805.

Gerhards, Jürgen, and Silke Hans. "From Hasan to Herbert: Name-Giving Patterns of Immigrant Parents between Acculturation and Ethnic Maintenance." *American Journal of Sociology* 114, no. 4 (2009): 1102-1128.

Goldin, Claudia. "The Political Economy of Immigration Restriction in the United States, 1890 to 1921." In *The Regulated Economy: A Historical Approach to Political Economy*, pp. 223-258. University of Chicago Press, 1994.

Goldstein, Joshua R. and Guy Stecklov. 2015. "From Patrick to John F.: Ethnic Names and Occupational Success in the Last Era of Mass Migration." *American Sociological Review*.

Gordon, Milton M. *Assimilation in American Life: The Role of Race, Religion, and National Origins*. New York: Oxford University Press, 1964.

Gould, John D. 1980. "European Inter-Continental Emigration. The Road Home: Return Migration from the USA." *Journal of European Economic History* 9 (1): 41-112.

Guinnane, Timothy W., Carolyn M. Moehling, and Cormac Ó Gráda. "The Fertility of the Irish in the United States in 1910." *Explorations in Economic History* 43, no. 3 (2006): 465-485.

Hacker, J. David. "Child Naming, Religion, and the Decline of Marital Fertility in Nineteenth-century America." *History of the Family* 4, no. 3 (1999): 339-365.

Hainmueller, Jens, and Michael J. Hiscox. "Attitudes toward Highly Skilled and Low-skilled Immigration: Evidence from a Survey Experiment." *American Political Science Review* 104, no. 01 (2010): 61-84.

Hatton, Timothy J. "The Immigrant Assimilation Puzzle in Late Nineteenth-Century America." *Journal of Economic History* 57, no. 01 (1997): 34-62.

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

Higham, John. *Strangers in the Land: Patterns of American Nativism, 1860-1925*. Rutgers University Press, 1955.

Jia, Ruixue and Torsten Persson. "Individual vs. Social Motives in Identity Choice: Theory and Evidence from China. Manuscript 2016.

Jones, Maldwyn Allen. *American Immigration: Second Edition*. Chicago, University of Chicago Press, 1992.



Kalmijn, Matthijs. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology* (1998): 395-421.

King, Desmond. *Making Americans: Immigration, Race, and the Origins of the Diverse Democracy*. Harvard University Press, 2009.

Liebersohn, Stanley. *A Matter of Taste: How Names, Fashion, and Culture Change*. New Haven: Yale University Press, 2000.

Liebersohn, Stanley, and Mary C. Waters. *From Many Strands: Ethnic and Racial Groups in Contemporary America*. Russell Sage Foundation, 1988.

Liebersohn, Stanley, and Eleanor O. Bell. "Children's First Names: An Empirical Study of Social Taste." *American Journal of Sociology* (1992): 511-554.

Liebersohn, Stanley, and Kelly S. Mikelson. "Distinctive African American Names: An Experimental, Historical, and Linguistic Analysis of Innovation." *American Sociological Review* (1995): 928-946.

Lind, Dara. "Donald Trump's Anti-immigrant Demagoguery Works Because It's Not About Jobs. It's About Fear." Vox.com. July 29, 2015.

Lubotsky, Darren. "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings." *Journal of Political Economy* 115, no. 5 (2007): 820-867.

Luttmer, Erzo F. P., and Monica Singhal. 2011. "Culture, Context, and the Taste for Redistribution." *American Economic Journal: Economic Policy*, 3 (1): 157-79.

Meng, Xin, and Robert G. Gregory. "Intermarriage and the Economic Assimilation of Immigrants." *Journal of Labor Economics* 23, no. 1 (2005): 135-174.

Minns, Chris. "Income, Cohort Effects, and Occupational Mobility: A New Look at Immigration to the United States at the Turn of the 20th Century." *Explorations in Economic History* 37, no. 4 (2000): 326-350.

Moehling, Carolyn, and Anne Morrison Piehl. "Immigration, Crime, and Incarceration in Early Twentieth-Century America." *Demography* 46, no. 4 (2009): 739-763.

Moehling, Carolyn M., and Anne Morrison Piehl. "Immigrant Assimilation into US Prisons, 1900-1930." *Journal of Population Economics* 27, no. 1 (2014): 173-200.

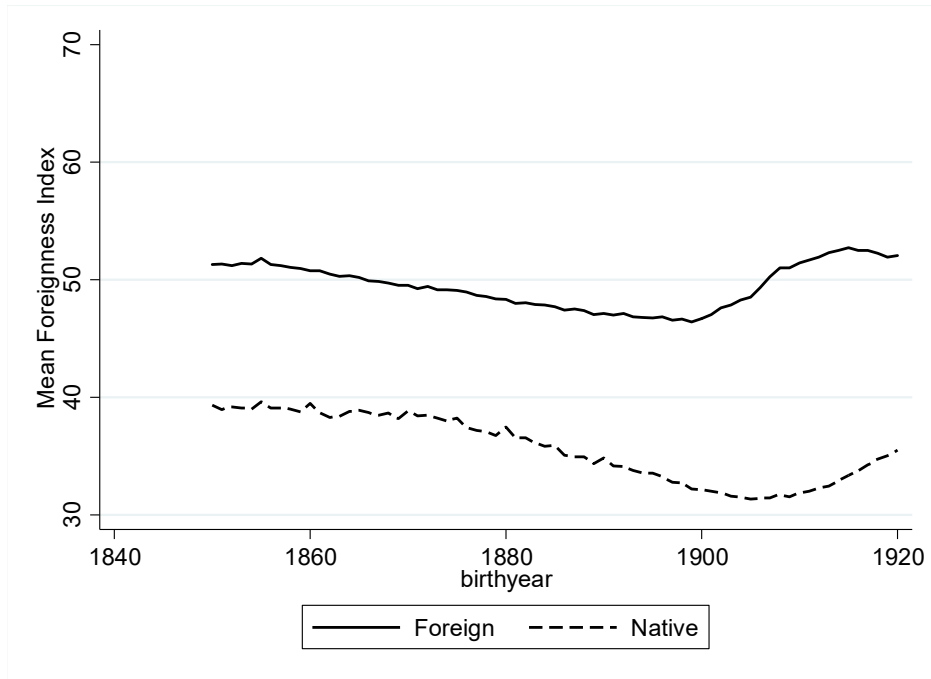
Moser, Petra. "Taste-Based Discrimination Evidence from a Shift in Ethnic Preferences after WWI." *Explorations in Economic History* 49, no. 2 (2012): 167-88.

Nossiter, Adam. "Marine Le Pen Leads Far-Right Fight to Make France 'More French.'" *New York Times*, April 20, 2017.

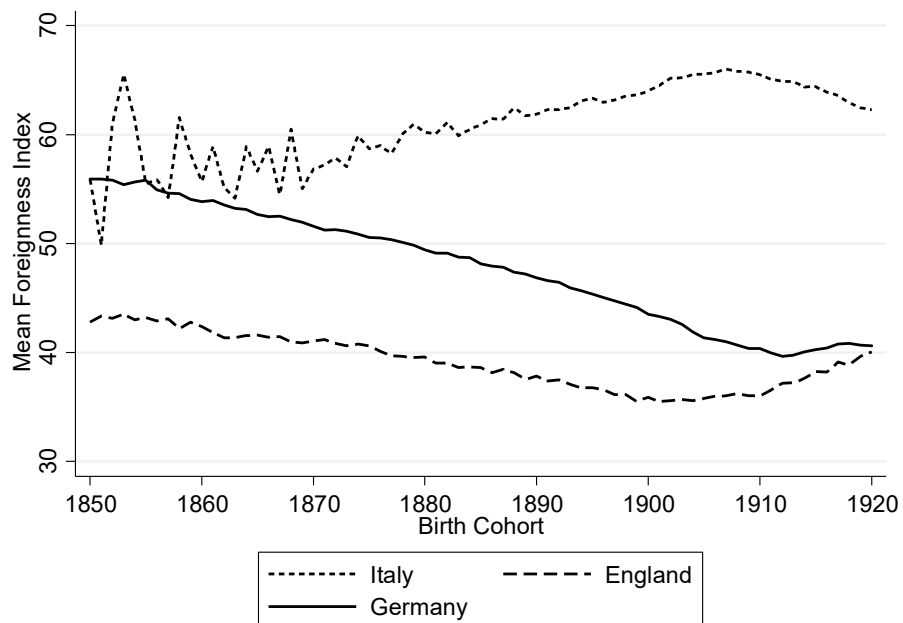
- Oreopoulos, Philip. "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes." *American Economic Journal: Economic Policy* 3, no. 4 (2011): 148-171.
- Pagnini, Deanna L., and S. Philip Morgan. "Intermarriage and Social Distance among US Immigrants at the Turn of the Century." *American Journal of Sociology* (1990): 405-432.
- Posner, Eric. "Trump Is the Only Candidate Talking About a Taboo Subject." *Slate.com*. August 25, 2015.
- Ramakrishnan, Karthick. "'Strange' vs. 'Simple Old American' Names." *Los Angeles Times*. May 28, 2015.
- Rossi, Alice S. "Naming Children in Middle-Class Families." *American Sociological Review* (1965): 499-513.
- Rubinstein, Yona, and Dror Brenner. "Pride and Prejudice: Using Ethnic-Sounding Names and Inter-Ethnic Marriages to Identify Labour Market Discrimination." *Review of Economic Studies* 81, no. 1 (2014): 389-425.
- Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2010.
- Salam, Reihan. "Why Immigration Pushed Britons to Brexit." *Slate.com*. June 24, 2016.
- Saletan, William. "Donald Trump's Next-Generation Bigotry." *Slate.com*. June 23, 2016.
- Shertzer, Allison. "Immigrant Group Size and Political Mobilization: Evidence from European Migration to the United States." NBER Working Paper 18827, February 2013.
- Sue, Christina A., and Edward E. Telles. "Assimilation and Gender in Naming." *American Journal of Sociology* 112, no. 5 (2007): 1383-1415.
- Watkins, Susan Cotts, ed. *After Ellis Island: Newcomers and Natives in the 1910 Census*. New York: Russell Sage Foundation, 1994.
- Watkins, Susan Cotts, and Andrew S. London. "Personal Names and Cultural Change: A Study of the Naming Patterns of Italians and Jews in the United States in 1910." *Social Science History* (1994): 169-209.
- Zelinsky, Wilbur. "Cultural Variation in Personal Name Patterns in the Eastern United States." *Annals of the Association of American Geographers* 60, no. 4 (1970): 743-769.

**Figure 1: Average name foreignness, Birth cohorts of 1850-1920**

**A. Sons of immigrant and native-born fathers**

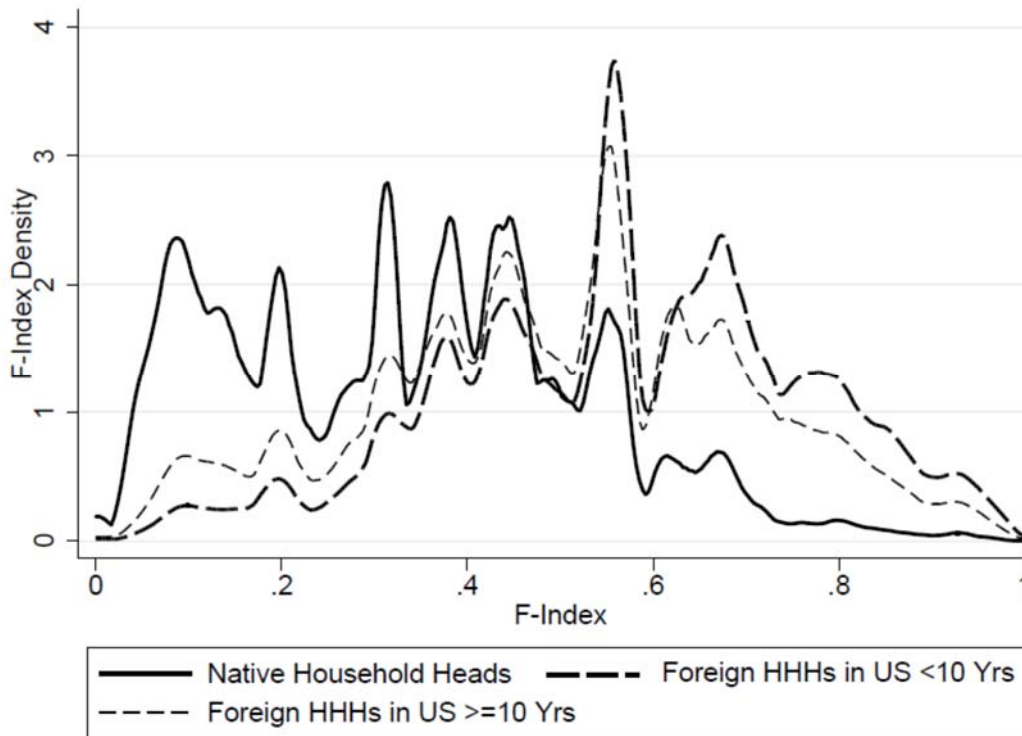


**B. Examples of countries with large reduction in name foreignness (Germany), high name foreignness (Italy) and low name foreignness (England)**



Notes: Sample is restricted to men born in the US. Father's nativity is determined using the father's place of birth variable. The Foreignness Index is calculated for each name and birth cohort using the complete-count (100 percent) sample of the 1920 census. The F-Index value for men in birth cohort  $t$  is based on men born in  $t-1$  through  $t-20$ .

**Figure 2: Kernel density estimates of name Foreignness Index in 1920, Children of native-born or foreign-born in the US more/less than 10 years**



Notes: Sample includes non-black children born in the US (outside of the South), living with their parents, and between the ages of 3-15 in the 1920 Census. Households are classified as native- or foreign-born based on the place of birth of the household head. Foreign-born households are further divided by time spent in the US (more/less than 10 years).

Figure 3: Census manuscripts for the Breitenbach family

A. Childhood household in 1920

Year	Household	Name	Relationship	Age	Sex	Marital	Occupation	Education	Place of Birth	Country
1914	Breitenbach	Thany G	Wife	29	F	M			England	Eng
		Josephine	Daughter	9	F	S			Massachusetts	USA
		August	Head	37	M	M			Massachusetts	USA
		Emma	Wife	41	F	M			Massachusetts	USA
		Emil	Son	15	M	S			Massachusetts	USA
		Richard	Son	14	M	S			Massachusetts	USA
1925	Leonard	Edwin	Son	9	M	S			Massachusetts	USA
		Jeanette	Daughter	7	F	S			Massachusetts	USA
		Charles	Head	32	M	M			Massachusetts	USA
		Minnie	Wife	32	F	M			England	Eng
		Leitreda	Daughter	9	F	S			Massachusetts	USA

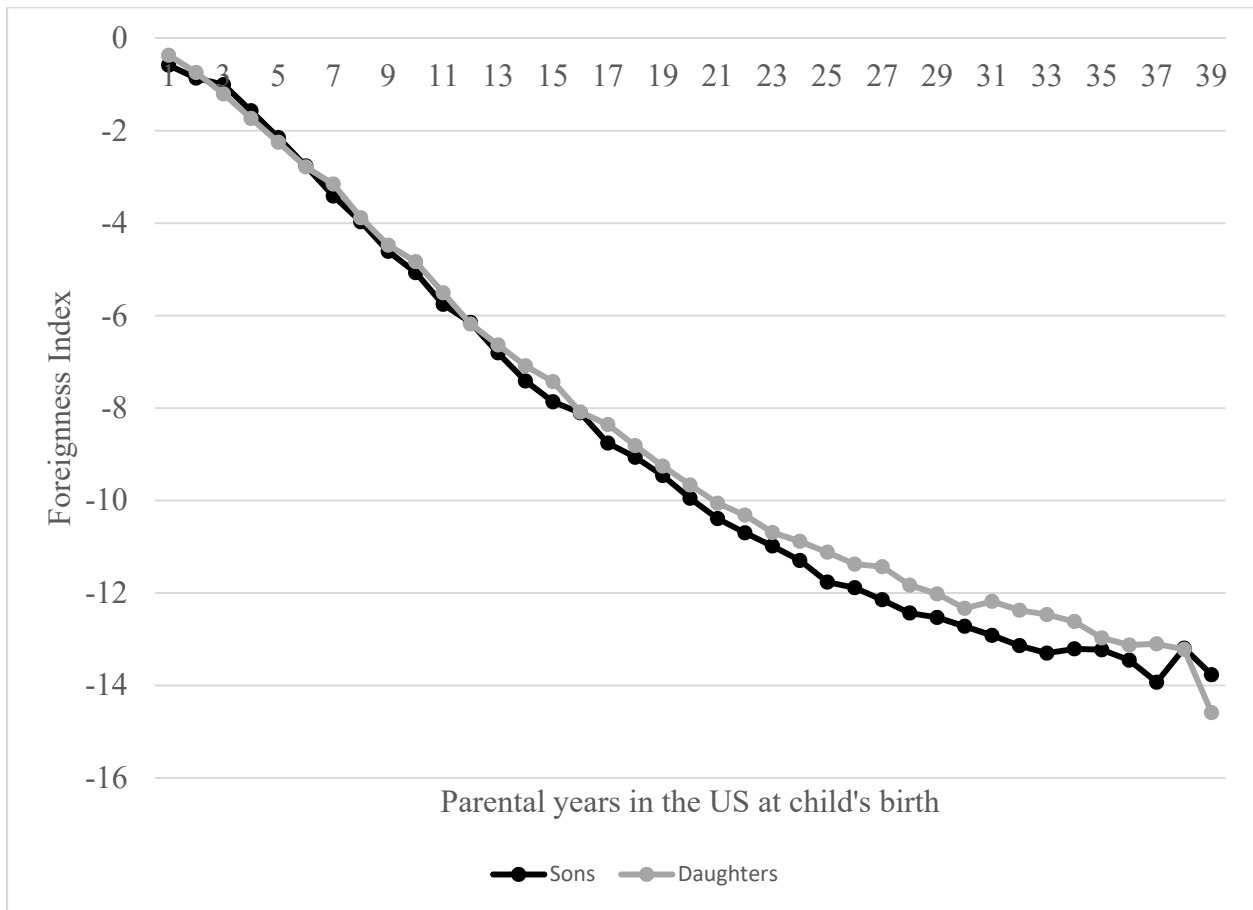
B. Emil Breitenbach in 1940

Year	Name	Relationship	Age	Sex	Marital	Occupation	Education	Place of Birth	Country
1940	Emil	Head	33	M	M			Massachusetts	USA
	Josephine	Wife	40	F	M			Massachusetts	USA
	Richard	Son	20	M	M			Massachusetts	USA
	Edwin	Son	17	M	M			Massachusetts	USA
	Leitreda	Daughter	14	F	S			Massachusetts	USA
	Charles	Daughter	12	F	S			Massachusetts	USA
	Minnie	Daughter	10	F	S			Massachusetts	USA
	Josephine	Daughter	8	F	S			Massachusetts	USA
	August	Daughter	7	F	S			Massachusetts	USA
	Emma	Daughter	6	F	S			Massachusetts	USA

C. Richard Breitenbach in 1940

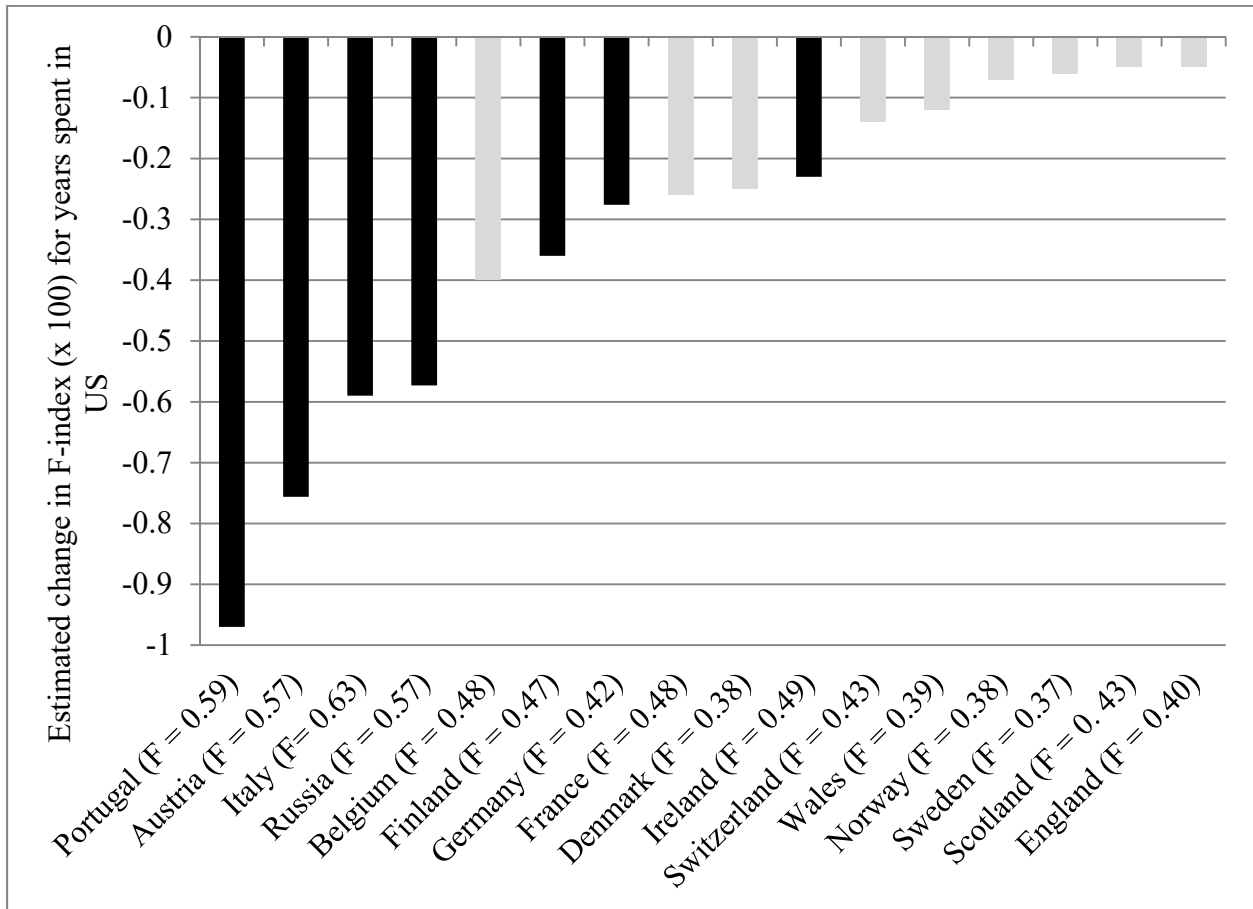
Year	Name	Relationship	Age	Sex	Marital	Occupation	Education	Place of Birth	Country
1940	Richard	Head	20	M	M			Massachusetts	USA
	Sarah	Wife	20	F	M			Massachusetts	USA
	Josephine	Daughter	10	F	S			Massachusetts	USA
	Edwin	Son	7	M	S			Massachusetts	USA
	Leitreda	Daughter	5	F	S			Massachusetts	USA
	Charles	Daughter	4	F	S			Massachusetts	USA

**Figure 4: Immigrants selected less foreign names for children after spending time in US, (Dependent variable = F-index)**



Notes: The graph reports coefficients from estimates of Equation 1, a regression of the F-index on a set of dummy variables for years that the household head had spent in the US by the time of the child's birth. Regressions also include dummy variables for child's age in five year bands and a set of family fixed effects. Data from 1920 complete-count Census. Sample includes children aged 3-15 who were born in a non-southern state and are living with their parents. Households must have a foreign-born head and the spouse (mother) must be less than 43 years old ( $N$  (sons) = 2,590,634;  $N$  (daughters) = 2,528,418).

**Figure 5: Effect of parental years in US on name foreignness, by sending country**

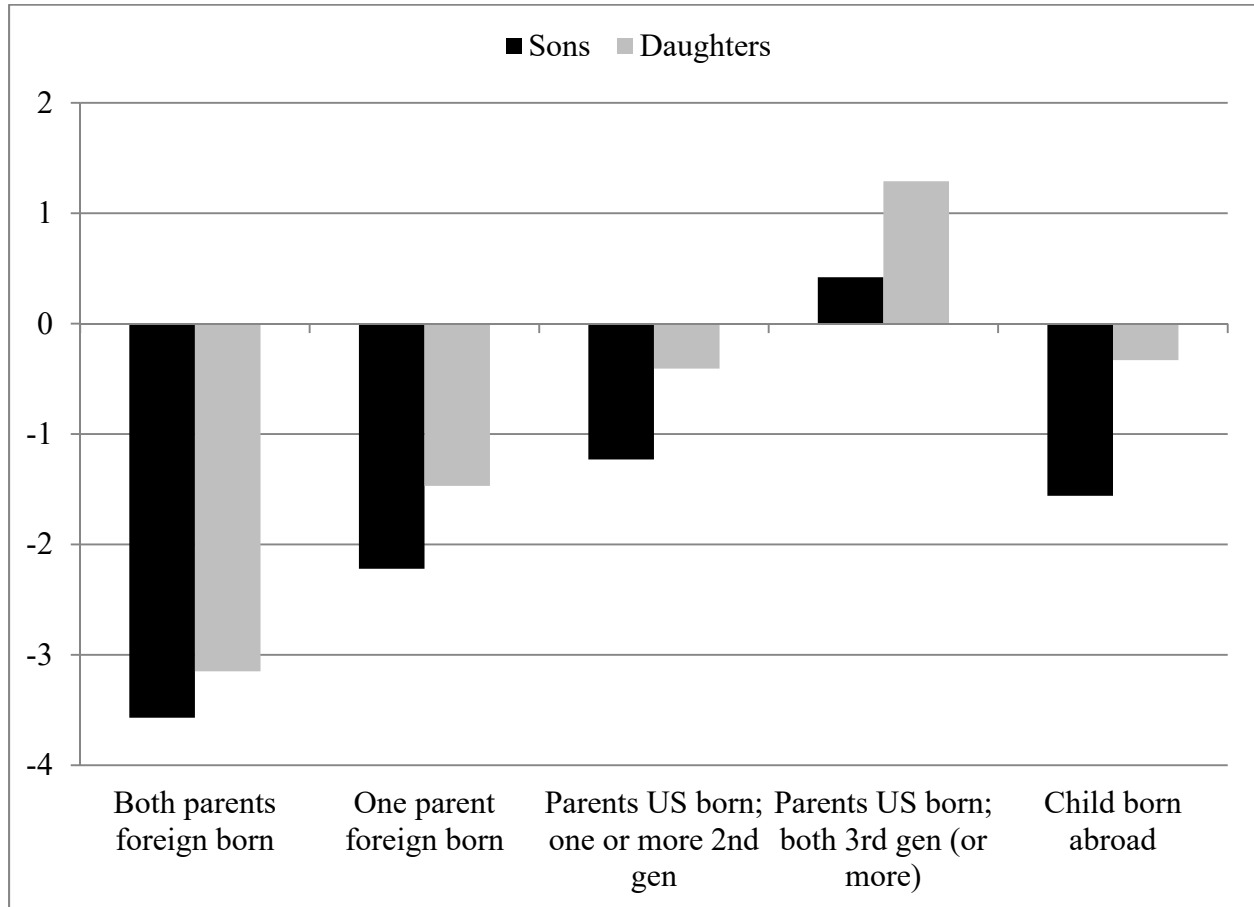


Notes: Reported coefficients from estimates of Equation 1, controlling for family fixed effects, child's birth year in five year bins, whether child has same name as father, and overall name frequency. Dark bars are statistically significant at the 5 percent level. Country labels also report average F-index for names given to children of immigrants from each country of origin. Data is from the complete-count 1920 census. Sample includes children aged 3-15 who were born outside the South, were living with their parents, and were living in a household with a foreign-born head. Spouse of household head (mother) must be less than 43 years old. Country of origin is defined by the place of birth of the household head.



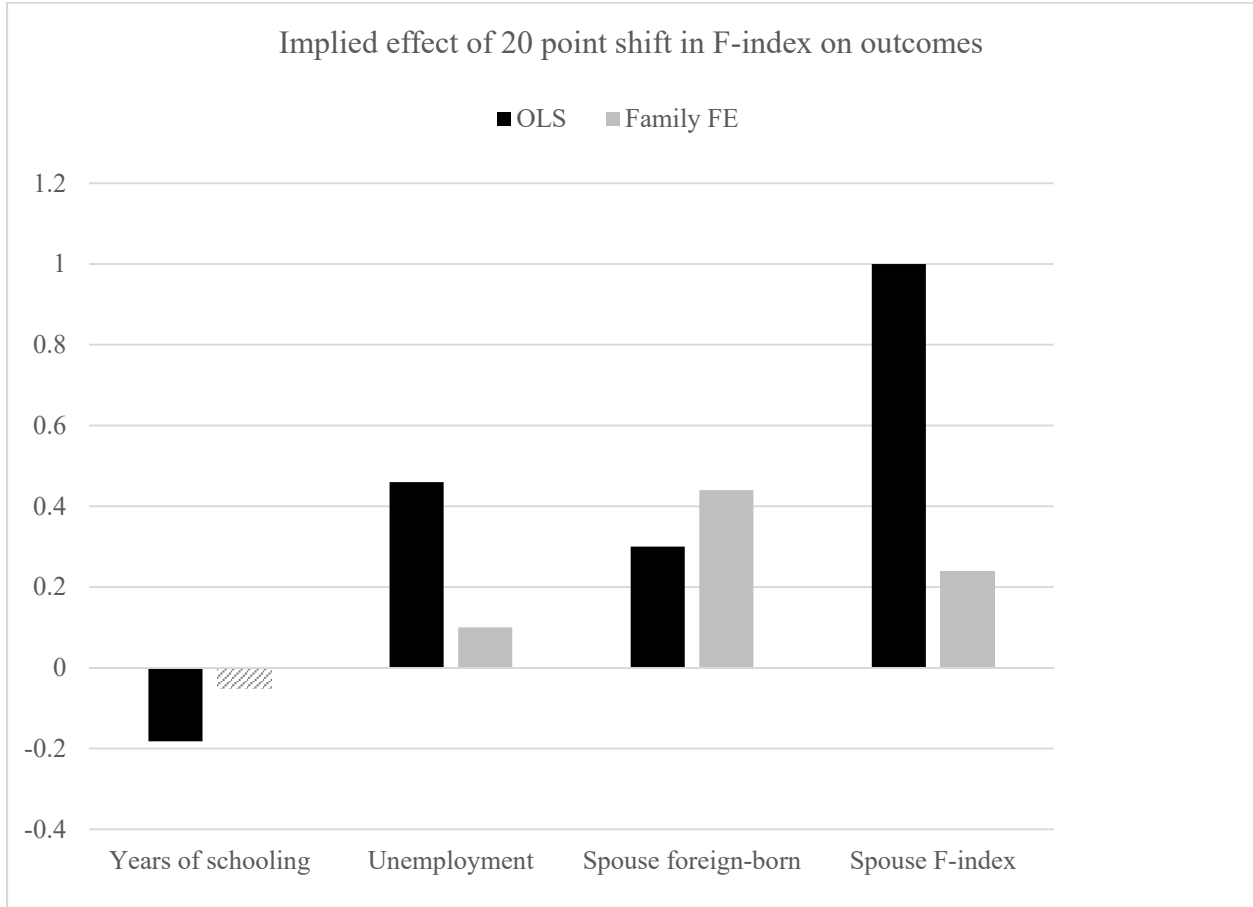
**Figure 6: Children of immigrants received less foreign names later in birth order, but children of other household types did not**

Implied effect on F-index of moving from first to fourth in birth order



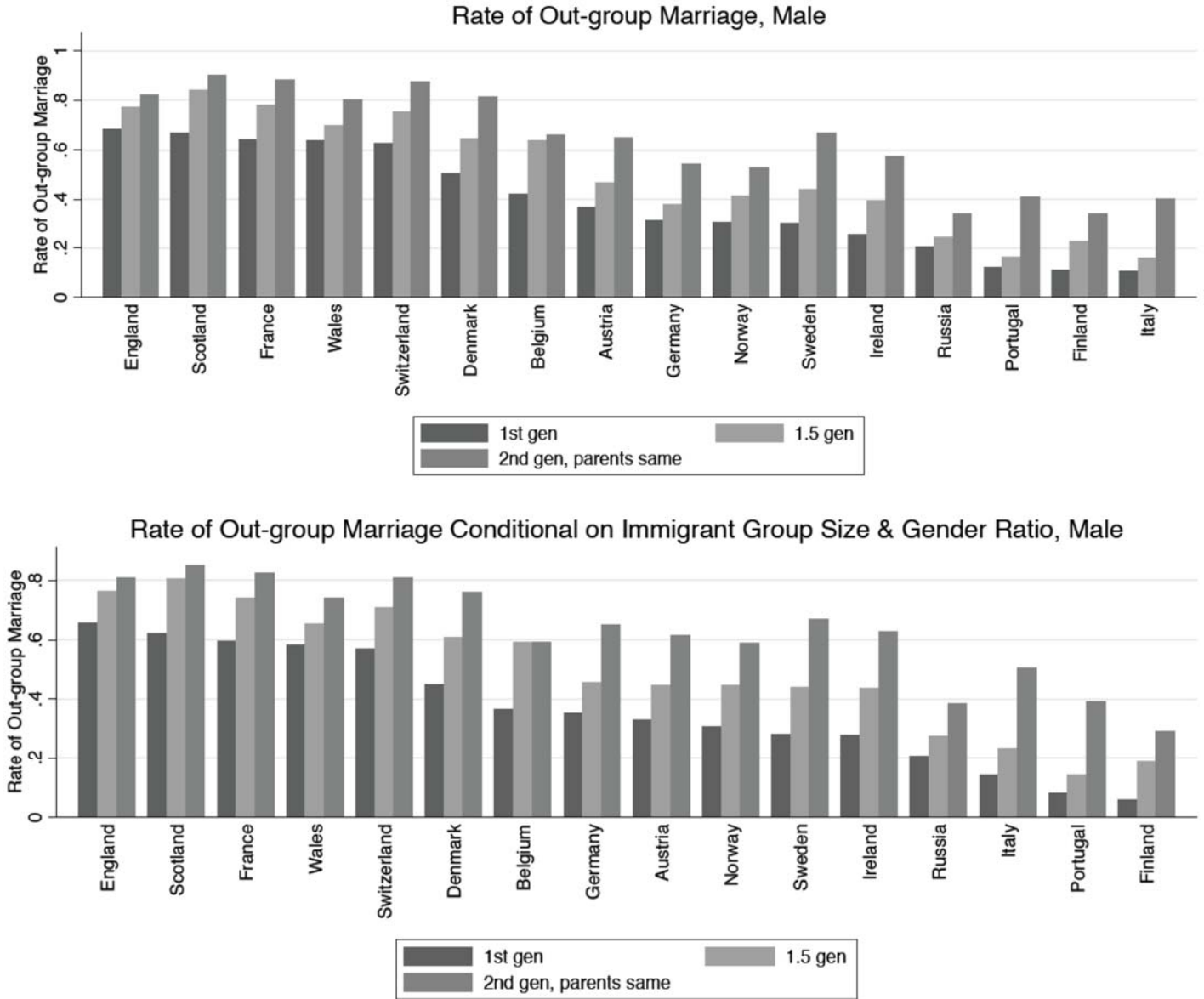
Notes: Data from the complete-count 1920 census. Bars report the implied effect on F-index of moving from first to fourth in birth order. Sample includes non-black children aged 3-15 living with their parents in a non-southern state. Spouse of household head (mother) must be less than 43 years old. Underlying regression also controls for family fixed effects, child's birth year, whether child has same name as father, and the overall frequency of the name.

**Figure 7: Effect of name foreignness on adult outcomes**



Note: Sample includes men matched between 1920 and 1940 complete-count Censuses. Men must be 3-15 in 1920, born outside the South and living at home with parents in 1920 in a household whose head was foreign-born. All regressions control for a vector of dummies for child's birth year, parental years in the US and child's rank in the birth order. Coefficients underlying the graph are reported in Appendix Tables 5 and 6. All OLS coefficients are statistically significant at the 1% level. Statistically significant coefficients in specifications that add family FE are marked with striped bars.

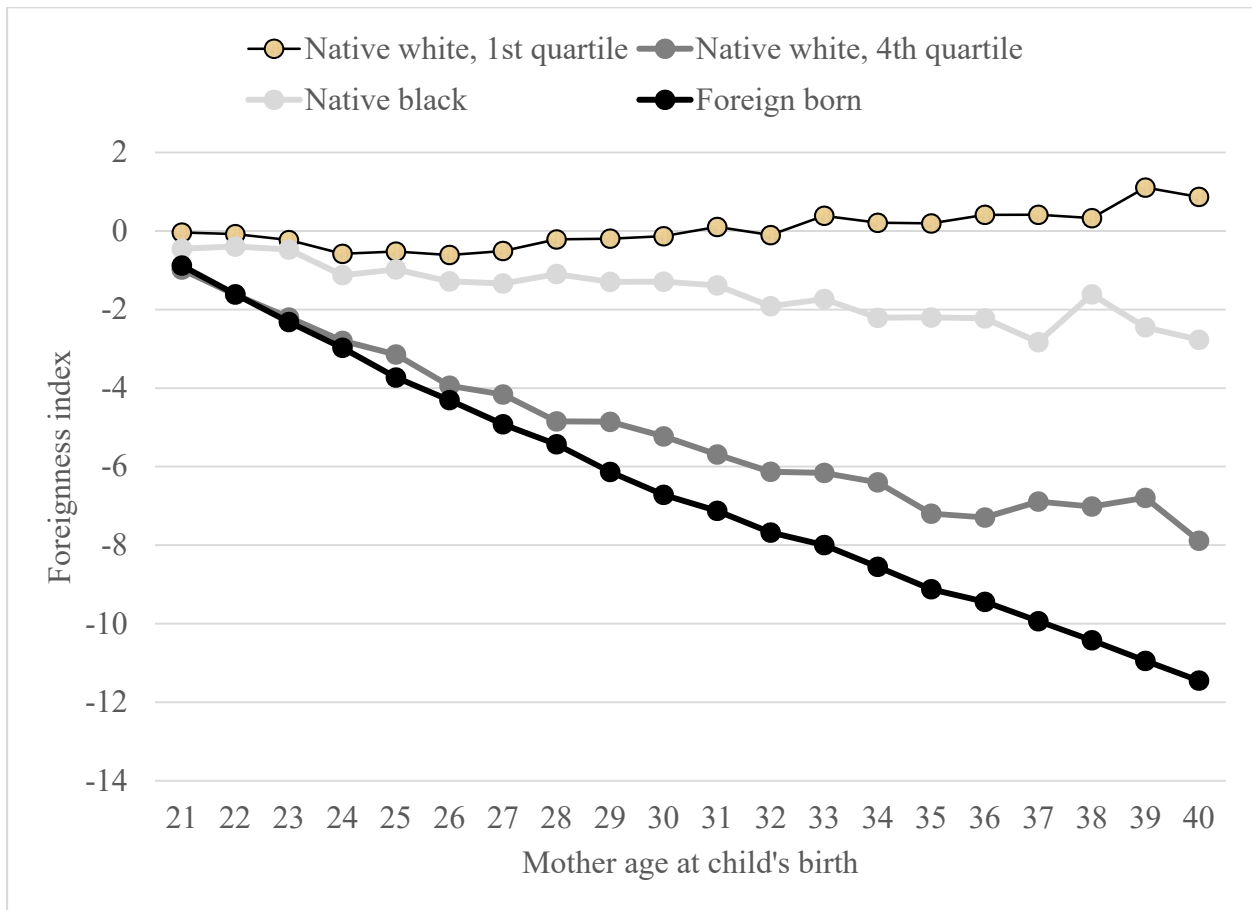
**Figure 8: Share of first and second generation immigrant men in out-group marriages, by country of origin, 1930**



Note: Figure based on men in IPUMS 5% sample of 1930 census who are currently married and whose age at first marriage occurred after arrival in the US. Men whose spouse (or spouse's parents) were born in the same country of origin as he (or his parents) are considered to be in an 'endogamous' marriage. We graph the complement here, namely men in out-group marriages. Panel B reports country of origin fixed effects from a regression whose dependent variable is a dummy equal to one for endogamous marriage with controls for the group size and gender ratio of the corresponding immigrant group at the state level.

Immigrant group size is defined as the number of immigrants (first or second generation) with a particular ancestry, relative to the total population. Gender ratio is defined as the ratio between total number of male to female immigrants (first or second generation) with a particular ancestry.

**Figure 9: Foreignness index by mother's age at child's birth and mother's nativity status and race, California birth records**



Notes: The graph reports coefficients from estimates of Equation 1, a regression of the F-index on a set of dummy variables for mother's age at time of child's birth. Regressions also include indicators for child's birth year in five year bands and a set of mother fixed effects. Data from California birth certificate registry. Sample includes children born to mothers aged 20-40 at the time of birth. Results presented separately for mother born abroad and native born mothers by race and quartile of foreignness of mother's last name.

**Table 1: Examples of foreign, neutral, and native names (1900-1920 birth cohorts)**

<b>Most foreign (F-Index &gt;0.90)</b>	<b>Most neutral (0.5 &lt; F-Index &lt; 0.52)</b>	<b>Most native (F-Index &lt;0.025)</b>
<b>A. Male names</b>		
Vito	Orlando	Gaylord
Mario	Benjiman	Doyle
Hyman	Murray	Clay
Pasquale	Otto	Lowell
Isidor	Theodor	Dale
Nick	Herman	Wayne
<b>B. Female names</b>		
Sonia	Margaret	Bethany
Antoinette	Deborah	Merlene
Concetta	Helene	Garnet
Johanna	Kathleen	Arlyce
Molly	Beatrice	Joellen
Carmela	Fay	Opal

Notes: Names with 100 or more observations selected for having high/lowest/most neutral F-index values in 1920 complete-count Census for the birth cohorts of 1900-20.

**Table 2: Comparing matched sample to population, 1920**

	Matched	Population	Difference (matched – pop.)
Foreignness Index, 1920	51.16 (21.29)	53.68 (21.27)	-2.521*** (0.156)
# characters, first name	5.724 (1.253)	5.698 (1.313)	0.026*** (0.002)
Age	8.712 (3.700)	8.721 (3.681)	-0.009* (0.004)
Number of Siblings	4.095 (2.061)	4.103 (2.143)	-0.008** (0.003)
Number of Brothers	2.571 (1.399)	2.579 (1.469)	-0.009*** (0.002)
Rank in brother order	1.803 (1.016)	1.772 (1.034)	0.029*** (0.001)

Notes: This table compares the matched sample to the full population of sons in households headed by a foreign-born household head in 1920 (N= 3,012,804 for the full population). For Row 2, we compare the matched sample to sample-line individuals in the 1940 Census for whom father's birthplace is available.

**Table 3: Did immigrants give less foreign name to sons after spending time in US?**

	F-index			Relative probability
	(1)	(2)	(3)	(4)
Years in US	-0.450*** (0.010)	-0.305*** (0.013)	-0.254*** (0.012)	-0.085*** (0.002)
Birth order controls	N	Y	Y	Y
Name control	N	N	Y	Y
<i>N</i>	2,520,347	2,520,347	2,520,347	2,520,347

Notes: Data is from the complete-count 1920 census. Sample includes sons aged 3-15 who were born outside the South, were living with their parents in 1920, and were living in a household with a foreign-born household head. All regressions control for dummy variables for child's age in five year bands, as well as a set of family fixed effects. To observe complete birth order, sample restricted to households in which mother is less than 43 years old.



**Table 4: Foreignness of sons' names with time spent in the US,  
By literacy and tenure status of household head**

Dependent variable = F-index

	Household head is...			
	Literate	Not literate	Homeowner	Renter
Years in US	-0.447*** (0.011)	-0.642*** (0.023)	-0.449*** (0.016)	-0.497*** (0.013)
<i>N</i>	2,150,897	369,450	992,097	1,528,250

Notes: Data from complete-count 1920 census. Sample includes sons age 3-15 living with their parents who were born in a non-southern state. Regressions control for dummy variables for child's age in five year bands, as well as a set of family fixed effects. To observe complete birth order, sample restricted to households in which mother is less than 43 years old.

**Table 5: Foreignness of sons' names with time spent in the US,  
By proxy for cultural distance (F-index of last name)**

Dependent variable = F-index		
	(1)	(2)
Years in US	-0.316*** (0.019)	-0.322*** (0.012)
Years in US x F-index of last name	-0.162*** (0.034)	
Years in US x % own country in district		-0.699*** (0.040)
<i>N</i>	1,254,001	1,688,760

Notes: Data from the complete-count 1920 census. Sample includes sons with a foreign-born head of household who were between the ages of 3-15 who were born outside the South and lived with their parents. Regressions control for dummy variables for child's age in five year bands, as well as a set of family fixed effects. To observe complete birth order, sample restricted to households in which mother is less than 43 years old. Sample restricted to households with surnames with at least 100 observations in the 1940 Census. Column (2) is further restricted to the 16 sending countries analyzed in Abramitzky, Boustan and Eriksson (2014).

**Table 6: Foreignness of Child’s Name by Rank in Birth Order and Mother’s Race and Nativity Status, California Birth Records**

	Foreign born	Native white 4 <sup>th</sup> quartile of last name F-index	Native white 1 <sup>st</sup> quartile of last name F-index	Native black
	<b>Panel A: Linear birth order</b>			
Birth order rank	-0.459*** (0.031)	-0.468*** (0.069)	0.048 (0.057)	-0.252*** (0.096)
	<b>Panel B: Birth order indicators</b>			
2 <sup>nd</sup> born	-0.224*** (0.049)	-0.514*** (0.113)	-0.037 (0.081)	-0.524*** (0.172)
3 <sup>rd</sup> born	-0.789*** (0.077)	-1.036*** (0.175)	0.278*** (0.138)	-0.689*** (0.255)
4 <sup>th</sup> born	-1.424*** (0.109)	-1.643*** (0.247)	0.077 (0.213)	-0.908*** (0.346)
<i>N</i>	4,801,260	739,978	732,032	474,855

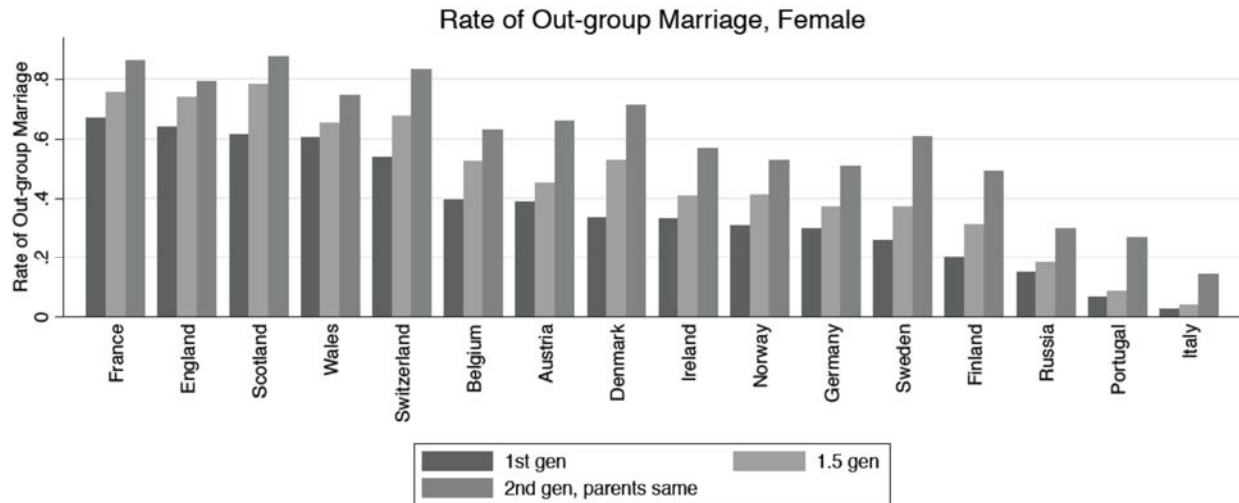
Notes: Data from California birth certificate registry. Sample includes children born to mothers aged 20-40 at the time of birth. Regressions control for mother fixed effects and indicators for child’s birth year in five-year bands. Panel B also contains indicators of rank in birth order up to rank = 20. For brevity, we report only second through fourth born. Column (1) includes births to mothers who were born abroad, while columns (2)-(4) contain births to native born mothers separated by race and by quartile of foreignness of mother’s last name.

**Table 7: Foreignness of Child's Name by Rank in Birth Order and Mother Characteristics, California Birth Records**

		<b>Panel A: By country of origin</b>				
		<b>First name</b>				
Birth rank		Mexico -0.542*** (0.036)	China -0.230 (0.231)	Philippines 0.025 (0.195)	Vietnam -0.437*** (0.160)	Rest world -0.351*** (0.071)
<i>N</i>		2,907,331	201,443	169,952	185,373	1,247,237
<i>Mean Findex</i>		76.2	47.1	38.7	50.7	57.2
		<b>Middle name</b>				
		-0.506*** (0.087)	-0.415 (0.538)	-0.346 (0.244)	-0.175 (0.216)	-0.392*** (0.111)
<i>N</i>		1,438,231	62,175	160,740	116,265	849,184
		<b>Panel B: By education level (Mexico only)</b>				
Birth rank		Less than HS -0.545*** (0.048)	HS grad -0.551*** (0.066)	Some college -0.558*** (0.103)	College + -0.396*** (0.167)	
<i>N</i>		1,641,792	754,906	333,763	176,870	

Notes: Data from California birth certificate registry. Sample includes children born to mothers aged 20-40 at the time of birth. Regressions control for mother fixed effects and indicators for child's birth year in five-year bands. Panel A considers foreignness of first name and middle name of child. Panel B focuses on foreignness of child's first name.

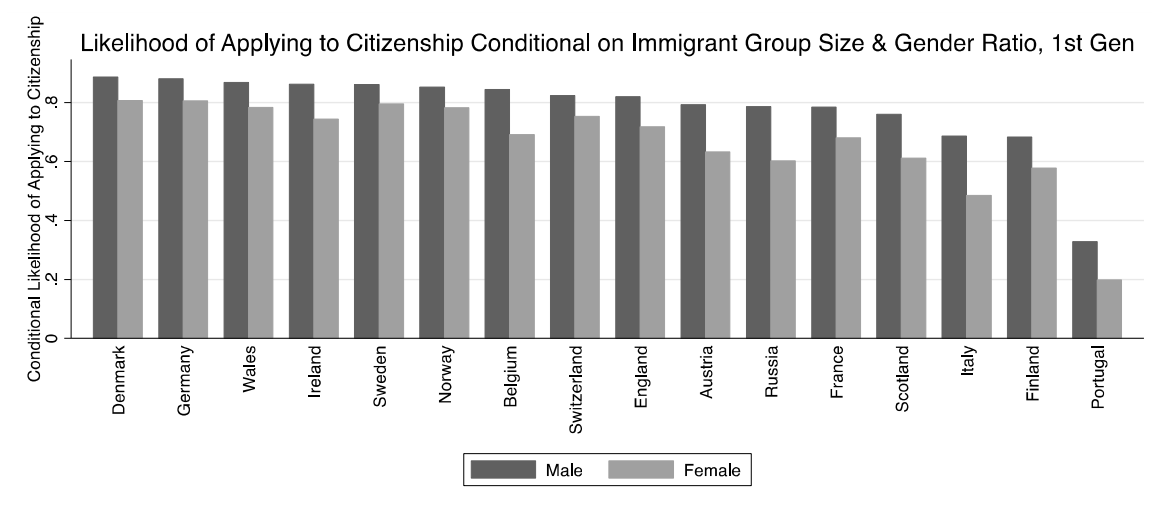
**Appendix Figure 1:  
Share of first and second generation immigrant women in out-group marriage, by country of origin, 1930**



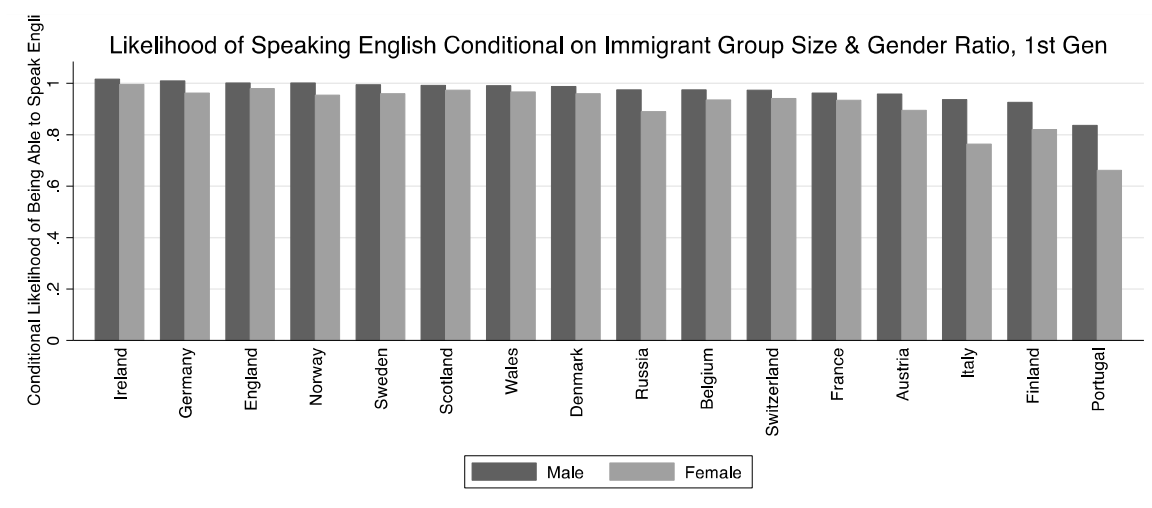
Note: Figure based on women in IPUMS 5% sample of 1930 census who are currently married and whose age at first marriage occurred after arrival in the US. Women whose spouse (or spouse’s parents) were born in the same country of origin as she (or her parents) are considered to be in an ‘endogamous’ marriage. We graph the complement here, namely women in out-group marriages. Panel B reports country of origin fixed effects from a regression whose dependent variable is a dummy equal to one for endogamous marriage with controls for the group size and gender ratio of the corresponding immigrant group at the state level. See notes to Figure 8 for definitions of control variables.

**Appendix Figure 2:  
Share of immigrants who engaged in other forms of cultural assimilation by country of origin, 1930**

**A. Applied for US citizenship**



**B. Reports ability to speak English**



Note: Figure is based on IPUMS 5% sample of 1930 census. The sample is restricted to individuals who were born outside of the US in one of the listed countries. Panel A reports country of origin fixed effects from a regression whose dependent variable is a dummy equal to one for immigrants who report being naturalized US citizens or having applied for first papers. The regression also controls for the group size and gender ratio of the corresponding immigrant group at the state level. Panel B follows the same format for a regression whose dependent variable is a dummy equal to one for immigrants reporting the ability to speak English. See the notes to Figure 8 for definitions of the other controls.

**Appendix Table 1:  
Did immigrants give less foreign name to daughters after spending time in US?**

	F-index			Relative probability
	(1)	(2)	(3)	(4)
Years in US	-0.446*** (0.010)	-0.339*** (0.013)	-0.438*** (0.010)	-0.055*** (0.001)
Birth order control	N	Y	Y	Y
Name controls	N	N	Y	Y
<i>N</i>	2,426,309	2,426,309	2,426,309	2,426,309

Notes: Data is from the complete-count 1920 census. Sample includes daughters aged 3-15 who were born outside the South, were living with their parents in 1920, and were living in a household with a foreign-born household head. All regressions control for child's age in five year bands and household FE. To observe complete birth order, sample restricted to households in which mother is less than 43 years old.

**Appendix Table 2:  
Foreignness of sons' names with time spent in the US,  
Robustness to measurement of F-Index**

Dependent variable = F-index; Coefficient on parental years in US

	(1)
Baseline <i>N</i> = 2,520,347	-0.476*** (0.010)
Adjust names with NYSIIS <i>N</i> = 2,233,135	-0.420*** (0.011)
Fix F-index in 1900 <i>N</i> = 2,520,347	-0.459*** (0.012)
F-index calculated in 1920 <i>N</i> = 2,512,422	-0.449*** (0.010)
F-index, by state of birth <i>N</i> = 2,443,062	-0.450*** (0.010)
F-index calculated by country <i>N</i> = 1,643,836	-0.736*** (0.019)
F-index, 2 <sup>nd</sup> gen foreign <i>N</i> = 2,513,315	-0.303*** (0.009)
Include South <i>N</i> = 2,647,144	-0.450*** (0.010)

Notes: Data from the complete-count 1920 census. Sample includes sons living in a household with a foreign-born head. All children in the sample were born outside the South, were between the ages of 3-15 in 1920 and lived with their parents in 1920. To observe complete birth order, sample restricted to households in which mother is less than 43 years old. Regression controls for child's age in five year bands, a dummy equal to one if child has same name as father, and name frequency.



**Appendix Table 3:  
Name foreignness and rank in birth order,  
Sons of immigrant and native parents**

	2 sons	3 sons	4 or more sons
<b>A. Both parents foreign born (Mean F-index = 56.1)</b>			
=1 if second born	-1.582*** (0.057)	-1.373*** (0.071)	-1.304*** (0.089)
=1 if third born		-2.652*** (0.097)	-2.328*** (0.112)
=1 if fourth or more			-3.645*** (0.138)
<i>N</i>	696,413	484,201	389,763
<i>F-test: 2<sup>nd</sup> son vs. 3<sup>rd</sup> son</i>		341.18	135.91
<i>F-test: 3<sup>rd</sup> son vs. 4<sup>th</sup> son</i>			246.49
<b>B. Parents born in US; At least one parent second generation immig.(Mean F-index = 36.7)</b>			
=1 if second born	-0.781*** (0.051)	-0.795*** (0.075)	-0.648*** (0.103)
=1 if third born		-0.704*** (0.103)	-0.926*** (0.132)
=1 if fourth or more			-0.798*** (0.164)
<i>N</i>	817,795	443,246	303,908
<i>F-test: 2<sup>nd</sup> son vs. 3<sup>rd</sup> son</i>		1.47	6.89
<i>F-test: 3<sup>rd</sup> son vs. 4<sup>th</sup> son</i>			1.58
<b>C. Parents born in US; both parents third generation immig. or more (Mean DV = 33.5)</b>			
=1 if second born	-0.016 (0.050)	0.070 (0.073)	-0.029 (0.100)
=1 if third born		0.456*** (0.100)	0.261** (0.129)
=1 if fourth or more			0.742*** (0.159)
<i>N</i>	923,360	491,979	322,367
<i>F-test: 2<sup>nd</sup> son vs. 3<sup>rd</sup> son</i>		28.99	7.91
<i>F-test: 3<sup>rd</sup> son vs. 4<sup>th</sup> son</i>			24.12

---

**D. Children born abroad (Mean F-index = 88.0)**

=1 if second born	-0.420 (0.267)	-1.096*** (0.290)	-0.977*** (0.330)
=1 if third born		-1.200** (0.501)	-1.223** (0.529)
=1 if fourth or more			-1.590** (0.739)
<i>N</i>	68,537	44,316	31,237
<i>F-test: 2<sup>nd</sup> son vs. 3<sup>rd</sup> son</i>		1.69	0.33
<i>F-test: 3<sup>rd</sup> son vs. 4<sup>th</sup> son</i>			0.43

---

Notes: The sample includes non-black sons aged 3-15 who were born outside the South and are living with their parents. To observe complete birth order, sample restricted to households in which mother is less than 43 years old. All specifications contain controls for child's birth year, a dummy for same name as father, name frequency and household FE.

**Appendix Table 4:  
Name foreignness and rank in birth order,  
Results for sub-groups of immigrant households**

Dependent variable = F-index; Coefficient on linear birth order

	(1)
<b><i>Baseline – 2 parents foreign</i></b> <i>N = 2,195,799</i>	<b><i>-1.199***</i></b> <b><i>(0.032)</i></b>
Both parents arrived as adults <i>N = 1,383,045</i>	-1.218*** (0.040)
One or more parents arrived as child <i>N = 812,754</i>	-1.169*** (0.054)
Parents from same sending country <i>N = 1,990,041</i>	-1.184*** (0.033)
Parents from different sending countries <i>N = 205,758</i>	-1.343*** (0.127)
<b><i>Baseline – 1 parent foreign</i></b> <i>N = 949,326</i>	<b><i>-0.744***</i></b> <b><i>(0.059)</i></b>
Only mother foreign <i>N = 348,842</i>	-0.461*** (0.097)
Only father foreign <i>N = 600,484</i>	-0.882*** (0.074)

Notes: Data from the complete-count 1920 census. Sample includes non-black children with two foreign-born parents (rows 1-5) or one foreign-born parent (rows 6-8) who were between the ages of 3-15, were born outside of the South and lived with their parents. To observe complete birth order, sample restricted to households in which mother is less than 43 years old. All regressions control for child's age in five year bands, a dummy equal to one if child has same name as father, and name frequency.

**Appendix Table 5: Effect of name foreignness on education, earnings and unemployment**

	No household FE			Household FE
	(1) Baseline	(2) Add controls	(3) Add F-index at 20	(4) Full Sample
<b>Panel A</b>				
<b>Dependent variable: Highest grade completed (Mean = 10.10)</b>				
F-index	-0.0094*** (0.0002)	-0.0091*** (0.0002)	---	-0.0025*** (0.0005)
<i>N</i>	669,149	602,242		602,242
<b>Panel B</b>				
<b>Dependent variable: =1 if unemployed <math>\times</math> 100 (Mean = 9.0)</b>				
F-index	0.022*** (0.002)	0.023*** (0.002)	0.009** (0.005)	0.005 (0.007)
F-index at 20			0.016*** (0.005)	
<i>N</i>	683,779	615,327	615,327	615,327

Note: Sample includes men matched between 1920 and 1940 complete-count Censuses. Men must be 3-15 in 1920, born outside the South and living at home with parents in 1920 in a household whose head was foreign-born. All regressions control for a vector of dummies for child's birth year. Columns 2-4 control for parental years in the US and child's rank in the birth order. Column 4 adds family fixed effects.

**Appendix Table 6: Effect of name foreignness on spouse characteristics**

	No household FE			Household FE
	(1) Baseline	(2) Add controls	(3) Add F-index at 20	(4) Full Sample
<b>Panel A</b>				
<b>Dependent variable: =1 if spouse foreign born x 100 (Mean = 5.4)</b>				
F-index	0.009*** (0.003)	0.015*** (0.003)	0.001 (0.009)	0.022 (0.015)
F-index at 20			0.017* (0.005)	
<i>N</i>	348,842	308,653	308,653	308,653
<b>Panel B</b>				
<b>Dependent Variable: F-index of spouse (Mean = 43.5)</b>				
F-index	0.051*** (0.002)	0.050*** (0.002)	0.034*** (0.005)	0.012 (0.011)
F-index at 20			0.018*** (0.005)	
<i>N</i>	333,895	295,401	295,401	295,401

Note: Sample includes men matched between 1920 and 1940 complete-count Censuses. Men must be born outside the South and living at home with parents in 1920 in a household whose head was foreign-born. The sample is further restricted to men who were at least 25 years old in 1940 and who report being married. 67.7 of the sample was married in this year. All regressions control for a vector of dummy variables for child's birth year. Columns 2-4 control for parental years in the US and child's rank in the birth order. Column 4 add family fixed effects.

**Appendix Table 7: Name foreignness and other labor market outcomes**

Dependent variable	Coefficient on F-index		
	Mean	Without HH FE	With HH FE
Annual earnings ( <i>N</i> = 430,011)	\$20,484	-17.95*** (1.449)	-1.41 (5.23)
Hourly wage ( <i>N</i> = 382,600)	11.38	-0.001 (0.001)	0.001 (0.0007)
Weeks worked per year ( <i>N</i> = 435,252)	44.63	-0.012*** (0.001)	-0.002 (0.005)
Hours worked per week ( <i>N</i> = 408,933)	42.99	-0.010*** (0.001)	-0.002 (0.005)
Employed ( <i>N</i> = 615,327)	85.8	-0.026*** (0.002)	-0.005 (0.0009)
Public emergency work ( <i>N</i> = 527,995)	3.5	0.008*** (0.001)	0.003 (0.006)
Self employed ( <i>N</i> = 527,995)	14.5	-0.006 (0.003)	-0.003 (0.010)

Note: Sample includes men matched between 1920 and 1940 complete-count Censuses. Men must be between the ages of 3-15 and living at home with their parents in 1920. Sample restricted to men whose fathers were foreign-born. The first two rows are further restricted to men with non-zero earnings who were not self-employed in 1940. All regressions control for child's birth year and household head's years spent in the US. Public emergency workers, many of whom worked for New Deal programs, are identified by the 'class of worker' variable.

**Appendix Table 8: Robustness to alternative matching approaches**

Coefficient on F-index			
Dependent variable: Highest grade attained			
Panel A: Basic approach			
	5 year band	+ exact name	+ exact age
No household FE	-0.0094*** (0.0002)	-0.0095*** (0.0003)	-0.0094*** (0.0003)
With household FE	-0.0015* (0.0008)	-0.0008 (0.0011)	-0.0014 (0.0016)
<i>N</i>	384,958	219,473	214,207
Panel B: Using Jaro-Winkler			
	Unique +/- 1 Match +/- 2	Unique +/- 2 Match +/- 2	Unique +/- 2 Match +/- 1
No household FE	-0.0085*** (0.0001)	-0.0087*** (0.0002)	-0.0088*** (0.0002)
With household FE	-0.0017*** (0.0005)	-0.0010* (0.0006)	-0.0009 (0.0006)
<i>N</i>	780,744	562,262	519,153

Notes: Sample includes men matched between 1920 and 1940 complete-count Censuses. Sample restrictions follow notes to Appendix Table 5. All regressions control for a vector of dummies for child’s birth year, parental years in the US and child’s rank in the birth order. The first column of Panel A restricts the sample to observations that are unique by first name, last name and age within a five year age band. The second column of Panel A maintains uniqueness within five years and replaces standardized names from the NYSIIS algorithm with exact names. The third column of Panel A maintains these two adjustments and requires that successful links have exact age matches. Panel B instead reports results that match based on Jaro-Winkler score of first and last name, rather than exact or standardized name. As above, all potential links must first be unique within their dataset (either 1920 or 1940 Census); here “unique” means that the individual is the only observation within a Jaro-Winkler string distance of 0.1 for both first and last name within a given age band. For all successful matches, (1) the matched pair must have first and last name within a Jaro-Winkler string distance of 0.1, and (2) there must be no other observations with a Jaro-Winkler string distance of 0.1 or less within one year band (column 1) or two years band (column 2) around the reported age. Column 3 additionally requires that a person is considered a “match” only when the reported age in both censuses results in a year of birth that is within one year of each other (column 1 and 2 allow for up to two years apart).

**Appendix Table 9:  
Pairwise correlations of various measures of cultural assimilation**

Variables	Out-group marriage	Speaks English	Applied for citizenship
<b>A. 16 sending countries</b>			
1			
2	0.642***		
3	0.476*	0.944***	
4	-0.489*	-0.767***	-0.666***
<b>B. Individual (N = 23,043)</b>			
1			
2	0.065***		
3	0.075***	0.258***	
4	-0.247***	-0.106***	-0.156***

Note: IPUMS 5% sample of 1930 census. (+) and (-) indicate positive and negative indicators of cultural assimilation. All samples restricted to white men who were born abroad and were 10 years or older. For *out-group marriage rate*, sample is further restricted to men who were currently married, and whose marriage took place after arrival in the US. For *average F-index of sons*, sample is restricted to men whose spouse is younger than 43 years old and whose oldest child is below 18 years old or less. The list of 16 sending countries underlying the correlations in Panel A can be found in Figure 2. The sample underlying the individual correlations in Panel B imposes all of the above restrictions (age, marital status, spouse's age and so on).



### **Data Appendix—Matching**

This appendix describes the procedure by which we match men from the 1920 Census to the 1940 Census. We use the Restricted Full Count data, shared by Ancestry.com, available on the NBER server. We begin by identifying children living at home in the full 1920 Census data who are born outside of the South. Because we are interested in the second generation, we match all men born in these states and then restrict the final sample to those whose father was born in one of 16 large European sending countries.<sup>38</sup> We restrict to boys between ages 3 and 15 in 1920. We select age 15 as the upper limit because more than 85 percent of the 15 year old sons of immigrants are observed living with their parents in 1920.

We search for viable matches for these men in 1940 using the iterative matching strategy developed by Ferrie (1996) and employed more recently by Abramitzky, Boustan and Eriksson (2012, 2014) and Ferrie and Long (2013). More formally, our matching procedure proceeds as follows:

- (1) We begin by standardizing the first and last names of boys in our 1920 samples to address orthographic differences between phonetically equivalent names using the NYSIIS algorithm (see Atack and Bateman, 1992). We restrict our attention to boys in 1920 who are unique by first and last name, birth year, and place of birth (either state or country) in the 1920 Census. We do so because, for non-unique cases, it is impossible to determine which of the records should be linked to potential matches in 1940.

---

<sup>38</sup> These countries are the same as those used in Abramitzky, Boustan, and Eriksson (2014): Austria, Belgium, Denmark, England, Finland, France, Germany, Ireland, Italy, Norway, Portugal, Russia, Scotland, Sweden, Switzerland, and Wales. These countries represented 91 percent of European immigrants living in the US in 1900.

- (2) We match observations forward from 1920 to the full count Census index in 1940 using an iterative procedure. We start by looking for a match by first name, last name, state of birth and exact birth year. There are three possibilities: (a) if we find a *unique* match, we stop and consider the observation “matched”; (b) if we find multiple matches for the same birth year, the observation is thrown out; (c) if we do not find a match at this first step, we try matching within a one-year band (older and younger) and then with a two-year band around the reported birth year; we only accept unique matches. If none of these attempts produces a match, the observation is discarded as unmatched.
- (3) Finally, after finding all unique matches in 1940, we restrict to those with foreign-born fathers in the 1920 census.