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

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.

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

Leah Platt Boustan
Department of Economics
8283 Bunche Hall
UCLA
Los Angeles, CA 90095-1477
and NBER
lboustan@econ.ucla.edu

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

I. Introduction

Donald Trump first gained traction as the 2016 Republican presidential nominee by declaring that immigration has a detrimental effect on American culture.[1] Likewise, the British vote to withdraw from the European Union was driven, in part, by concerns about immigrant assimilation (Salam, 2016). In the past, much like today, politicians accused immigrants of maintaining distinct cultural norms, continuing to speak foreign languages and living in enclave communities.[2] Senator Henry Cabot Lodge, a leading advocate for border closure, argued in 1891 that 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—a matter as serious as the effect on the labor market." Congress finally passed strict immigration quotas in the early 1920s.[3]

This paper studies the cultural assimilation of immigrants during the Age of Mass Migration (1850-1913), a formative period in US history. We trace out immigrants' "cultural assimilation profile" with time spent in the US, using changes in the foreignness of names that immigrant parents selected for their children as a measure of cultural adaptation. Children's names offers an attractive measure of the assimilation process, because names carry cultural

[1] 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.
[2] Opponents of open immigration in the early twentieth century relied on cultural arguments to defend border restriction (Higham, 1955; Jones, 1992, chapter 9; Fetzer, 2000; King, 2000). Recent survey evidence suggests that cultural concerns are equally or more important than economic factors in determining attitudes towards immigration policy today (Citrin, Green, Muste and Wong, 1997; Hainmueller and Hiscox, 2007).
[3] 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.

content and because naming is a pure choice for immigrant parents, unconstrained by financial limitations or by discrimination on the part of natives.[4,5]

We find that immigrant parents in the 1900s and 1910s chose less foreign names for their children as they spent more time in the US, erasing half of the gap in name choice with natives after 20 years in the US.[6] Name-based assimilation occurred at the same rate for sons and daughters and regardless of household literacy or wealth (as proxied by homeownership). By this measure, the speed of assimilation varied substantially by country of origin and was stronger among immigrants with more cultural distance from US natives, generating cultural convergence over time. We interpret the shift away from foreign names with time spent in the US as driven by some combination of learning about US culture, acquiring a greater desire to integrate into US society, and deciding to stay in the US (rather than return to the home country). Some immigrants may arrive with a strong desire to assimilate, but with little knowledge about how to effectively do so (in this context: little understanding of which names are popular among US natives), while others may start out with concerns about assimilation which lessen over time.

[4] Lieberson (2000) is the classic reference in the sociology of naming. Consistent with our claim that names can be used as a measure of cultural assimilation, Lieberson argues that "given the strong thrust toward assimilation among most groups that migrate to the United States, the names immigrants give to their American-born children often differ from those preferred in the homeland" (p. 175). Lieberson offers some suggestive evidence on this point based on the names of the children of white ethnic families in Illinois in the 1980s. Naming patterns have been used as a measure of social distance in a number of other social contexts in both economics and sociology. Zelinsky (1970) and Lieberson and Bell (1992) study differences in name choices by region and by parental education, respectively. On African-American naming practices, see Lieberson 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).

[5] In contrast, inter-marriage requires finding a willing spouse. 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.

[6] 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.

The observed shifts in naming choices with parental time in the US are not simply capturing naming practices that vary by rank in the birth order, which is correlated with parental time in the US. The relationship between names and parental years in the US is robust to controlling for birth order directly or for an indicator for sharing a name with a parent, which is more common for 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.

The cultural assimilation of immigrant households had social and economic consequences for children in school and in the labor and marriage markets. We link over a million children of immigrants born between 1900 and 1920 across historical Censuses. Following men from their childhood families in 1920 into adulthood in 1940, we find that children with less foreign names completed more years of schooling, earned more and were less likely to be unemployed. Children with less foreign names were also less likely to marry a spouse who was born abroad or who had a very foreign name herself.

The association between name foreignness and adult outcomes decline somewhat but are still present when we compare brothers raised in the same household, even brothers who were born one year apart and thus grew up with similar levels of cultural and financial resources. This pattern suggests that a portion of this association is not driven by the selection of households that choose to use foreign names. Having a foreign name could influence an individual's behavior through his own perceived identity or through the experience of discrimination in school or on the job. Using changes in naming trends over time, we find that name foreignness at labor market entry matters more than name foreignness at birth for earnings and unemployment, which is consistent with employer discrimination playing some role.

The lack of complete convergence in name choices within a generation highlights the value that immigrants place on maintaining their cultural identity. Children with less foreign names earn higher incomes, and choosing a native-sounding name does not carry a financial cost, yet some immigrant parents still choose to endow their children with culturally-specific names. The observed value on maintaining one's ethnic identity is 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.[7]

We end by exploring other aspects of cultural assimilation, which confirm that immigrants experienced substantial convergence with natives in the early twentieth century. By 1930, more than two-thirds of immigrants had applied for US citizenship and almost all reported some ability to speak English. A third of first-generation immigrants who arrived before marrying and more than half of second-generation immigrants married spouses from different origins.

II. Literature on immigrant assimilation and on 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

[7] Jia and Persson (2016) is a nice application of these theories to the choice of child's ethnicity in mixed marriages in China.[8] 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.

natives with time spent in the destination.[8] 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).[9] However, this lack of labor market assimilation need not foreclose cultural integration and so understanding the speed of cultural assimilation is an independent topic of interest.[10]

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, which 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 age at first marriage, completed family size, political participation and criminality, but this transition often took more than one generation (Watkins, 1994; Guinnane, Moehling and O'Grada, 2006;

[8] 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.

[9] 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.

[10] 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 6) is 0.14.

Foley and 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.[11]

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 received lower occupation-based earnings in 1930, even after controlling for a proxy for family background. 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; Fryer and Levitt, 2004; Figlio, 2005); and lower earnings for individuals with identifiably ethnic surnames (Rubinstein and Brenner, 2014).[12]

[11] 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.

[12] Fryer and Levitt (2004) show that, after controlling for family background, having a blacker name is not associated with poorer adult outcomes. Race is highly observable; therefore, even if black workers with racially ambiguous names garner more interviews, they might be rejected by employers later in the hiring process.

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 (Bertrand and Mullainathan, 2004). By comparing siblings born into the same families and leveraging changes in name trends over time, we improve the source of identification without sacrificing the use of actual labor market outcomes.

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.[13] 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 the foreign born 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:

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

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*). We discuss robustness to a number of alternative specifications of the Foreignness Index in Section IV.

Table 1 lists the most foreign, neutral and native names for boys and girls in the birth cohorts of 1900-20. Neutral names like Murray and Herman were equally common among the

[13] The formula for *R* is given by: $R = \frac{\frac{\#\ foreigners_{name}}{total\ \#\ foreigners}}{\frac{\#\ natives_{name}}{total\ \#\ natives}}$

children of foreign-born and native parents. The most foreign names in this period 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.[14] 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 shifted 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 adoption of distinctively black names among the African-American community between 1965 and 1980.[15] The gap in the F-index between the children of immigrant and native fathers reached 20 points by 1910.

The second panel of Figure 1 demonstrates that naming practices varied by immigrants' 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 the figure). In contrast, the sons of immigrants from

[14] 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 1830-49, 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.

[15] 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.

Southern and Eastern Europe retained distinctive names in both early and later birth cohorts (see the sons of Italian fathers 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). We tested but did not find any evidence for breaks in naming trends during the volatile decade of the 1910s, when we might have expected changes in naming practices during World War I or following key political events in sending countries.[16]

Figure 2 offers the first evidence of assimilation in naming patterns with time spent in the US, graphing 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

[16] 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). However, we find no trend break in naming practices in German households during World War I at the national level. Fouka (2015) shows that German immigrants in states that introduced anti-German language policies during the war responded by choosing visibly German names, perhaps as a show of community support.

Census. We link men over time by first and last name, age and state of birth; details on the linking procedure are provided in the Data Appendix. We restrict our attention to men between the ages of 3 and 18 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 linking procedure generates a final sample of just over one million men, 688,875 of whom also have a matched brother. 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).[17]

Men who received a foreign name at birth may have been more likely to change their name in adulthood or to have their name misspelled on a Census form or mis-transcribed 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 5 points less on the Foreignness Index, or 10 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.

It seems unlikely than many second-generation immigrants changed their names. In the full population, name changers would appear with a highly foreign name in 1920 (when they were children) and a more Americanized name in 1940 (when they were adults). Yet, the name foreignness of second-generation immigrants was similar in 1920 and 1940, both in the full

[17] 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.

population and in the linked sample.[18] Even if name changing was rare, we might overstate the penalty of having received a foreign name at birth if men who adopted new names avoided some of the costs associated with their birth name. We take 30 percent as a (far) upper bound on the share of second-generation immigrants who would change their name, which was the reported share of first-generation immigrants who were name-changers (Biavaschi, et al., 2013). In this case, the lower-bound population estimate might be recovered by multiplying our estimate by 0.7.[19]

IV. Household 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:

$$FIndex_{ij} = \alpha_j + \beta_1 YearsUS_{ij} + \beta_2 Age_{ij} + \beta_3 X_{ij} + \varepsilon_{ij} \quad (1)$$

$$FIndex_{ij} = \alpha_j + \beta_1 BirthOrder_{ij} + \beta_2 Age_{ij} + \beta_3 X_{ij} + \varepsilon_{ij} \quad (2)$$

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 0-18) living in households with a foreign-born

[18] A similar argument applies to selective mortality. It is possible that men with more foreign names had higher mortality rates, and were thus less likely to survive until 1940. However, in this case, we would expect the set of names held by men in the matched sample to more closely resemble the population in 1940 than in 1920. Yet, we find a similar difference in the Foreignness Index between the matched sample and the population in both years.

[19] Say that 30 percent of men changed their name and that, for these men, their birth name had no effect on adult outcomes. We estimate a penalty of X for the 70 percent of men who did not change their name. The lower bound would be achieved when there is no association between name foreignness and outcomes for the 30 percent of men who changed their name. Thus, the lower-bound population penalty would be $(0.7 \cdot X + 0.3 \cdot 0)$.

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 household fixed effects (α_j), which ensures that the effect of parental time in the US is identified by differences between siblings born after their parents spent more/less time in the US.[20] After including household fixed effects, note that our estimates are *not* based on a comparison of the more foreign names favored by recently-arrived southern and eastern European immigrants to the less foreign names of more established northern and western European groups. We control for a set of dummies for child *i*'s birth year (Age_{ij}) 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.[21]

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. In both specifications, our analysis sample is limited 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

[20] 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 household fixed effects. 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.

[21] For an example of ethnic naming traditions that vary by birth order, see http://homepages.rootsweb.ancestry.com/~cregan/patterns.htm.

sibling from the childhood home, we further restrict the sample to children whose mother is less than 43 years old.[22]

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.

Figure 4 reproduces the relationship observed in the Breitenbach family between parental years in the US and name foreignness for the full sample of households in the 1920 Census with a foreign-born household head. For graphical representation, we present the effect on a child's F-index of yearly dummies for parental years in the US at the time of the child's birth. 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, linearly 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.

[22] 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.

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. Given the similarity of results by gender, we only present patterns for sons in the body of the paper but show results for daughters in the Appendix.

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 given more foreign names. Thus, given the correlation between birth order and parental time in the US, the main effect declines by 30 percent but is still 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.[23] Names shared with a father were more foreign, as were religious names; 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 Table 3 replaces the F-index with the relative probability measure as a dependent variable. Each year

[23] 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.

spent in the US lowers the relative probability of a name by 0.1 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 is completely closed after 20 years in the US.

The pace of name-based assimilation is no different for the poor than for the rich, and is faster for immigrants with greater cultural distance from US natives. Name choice is a financially inexpensive means for immigrants to express their interest in integrating into US culture. Table 4 shows that both literate and illiterate household heads experienced a similar shift away from foreign names for each year spent in the US, (around 0.4 points on the Foreignness index for each year in the US). Immigrants who rented their housing unit also 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 6 documents that the speed of name-based assimilation differed by country of origin. Immigrants from Scandinavian countries (Denmark, Norway, Sweden) had among the fastest rates of cultural assimilation, followed by immigrants from Southern Europe (Italy, Portugal). The slowest rates of name-based assimilation were experienced by Russian immigrants (many of whom were Russian Jews) and Finns. Reassuringly, immigrants from English-speaking countries (England, Scotland, Wales) exhibit no changes in name choice with time spent in the US.[24]

[24] In the full set of 16 sending countries, we find no correlation (or, if anything, a positive relationship) between initial F-index and changes in F-index with time spent in the US. However, this weak relationship is driven by the fact that immigrants from the three Scandinavian countries

Table 5 proxies for cultural distance with an F-index of surnames (analogous to the F-index for given names described in Section III), positing that households with more foreign-sounding last names hail from more dissimilar cultures. The surname F-index can vary both within and between sending countries. We interact parental years in the US with the surname index to examine whether immigrants that started out further from the native norm took more steps to assimilate or found themselves further behind. We find that households with more foreign last names shift toward native-sounding names more rapidly. In this sense, cultural assimilation generated convergence between immigrant families from different backgrounds.[25]

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.[26] Including southern residents in the analysis also leaves results unchanged.

(Denmark, Norway and Sweden) start out with low F-indices (mean = 37) but yet are quick to select less foreign names with time spent in the US. After excluding these three countries, we find that a 25 point difference on the initial F-index (or, the difference between Wales and Italy, the countries with the highest and lowest initial F-indices) is associated with a change of 3 additional points on the F-index for 20 years spent in the US.

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

[26] 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 5 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 2 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 no such effect of birth order on name foreignness for the sons of immigrants who were themselves born abroad (and a much smaller effect for daughters), 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 for third-or-higher generation Americans, whose first and fourth-born sons are separated by only 0.3 points on the F-index; this effect is statistically different from the estimates for immigrant parents.

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 a 0.5-0.6 point decline in the F-index. However, for households with second- (and third-plus) generation parents, the linear birth order effect is driven by first-born sons. First-born sons score up to 0.7 points higher on the

F-index than their younger brothers, but we see no systematic differences between second, third or fourth sons.[27]

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 appreciable different than in households in which both parents arrived as adults. Households with immigrant parents from two countries of origin shift name choices nearly twice as fast as immigrant parents in endogamous marriages, perhaps because they are less constrained by adherence to a single ethnic tradition. Among households with only one foreign-born parent, having a foreign-born father is associated with somewhat faster name-based assimilation than having only a foreign-born mother.

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 years later in school and/or in labor and marriage markets. We then explore how much of this association is driven by the selection of families that choose to give foreign names to their children. We start by broadly comparing all children of immigrant parents who received more/less foreign names, with name choice being used as a proxy for parental assimilation at the time of the child's birth. We find that men who received

[27] 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.

more foreign names attended fewer years of schooling, were more likely to be unemployed and earned lower pay. They were also more likely to marry a spouse who was foreign born and who had a more foreign-sounding name herself.

The estimated disadvantages are reduced by around one third when we add family fixed effects (although the marriage results are hardly changed), suggesting that a portion of the observed association between name foreignness and economic outcomes was driven by differences in household environment during childhood, including parental language ability and cultural awareness. We continue to find a strong negative association between name foreignness and outcomes even when we compare pairs of brothers born only one year apart. The Breitenbach family in Figure 3 illustrates this within-household design. We were able to follow Emil and Richard 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.

Two mechanisms, which we cannot tell apart, could have driven the negative effect of foreign names on both schooling and labor market outcomes even for brothers within the same family. 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.

Our main regression equation estimates the relationship between an adult outcome *y* and the F-index value of a man's name at birth, controlling for the individual's age:

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

In some specifications, we control for parental years in the US and child's place in the birth order and/or include household fixed effects (α_j) in order to compare brothers who were given names 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 3-18 in the 1920 Census, were not born in the South and can be successfully matched to the 1940 Census. The overall matched sample contains over one million observations, 688,875 of which are matched brothers.

Table 6 presents estimates of the relationship between name foreignness and our three economic outcomes: highest grade completed, unemployment, and annual earnings, conditional on positive earning. The first three columns use variation across households, while the remaining columns add household fixed effects. We start by comparing the population estimates in column 1 to the within-household estimates in column 4. By multiplying each coefficient by 20, we find that a 20 point shift in the F-index (the typical gap between the children of immigrants and natives) is associated with nearly two months less of schooling, or one month between brothers; $200 lower annual earnings in 2010 dollars, or a $130 (0.6 percent) decline in earnings between

brothers; and a 0.5 point increase in the probability of unemployment.[28] Smaller effects of name foreignness within households imply that a portion of the population-wide estimate is driven by the correlation between names and family background.[29] If we assume that the earnings gap is entirely explained by differential schooling, the estimated effect of name foreignness on annual earnings is consistent with a 5.4 percent return to education; if, instead, a portion of the earnings gap was due to discrimination in the labor market, the implied return to schooling would be lower.[30]

Results are not sensitive to controlling for parental years in the US or child birth order (column 2), implying that names may have been used as signals of ethnicity even for sons raised in households with similar levels of cultural assimilation. Furthermore, the estimated effect of names is not merely picking up a relationship between birth order and economic success.[31]

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

[28] The 1940 Census only contains information on wage and salary income. As a result, results on annual earnings exclude the self-employed. Appendix Table 5 shows that shifts in the F-index have no effect on the probability of self-employment.

[29] 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.

[30] A 20 point increase in the F-index is associated with a one month decline in schooling and a 0.6 percent decline in earnings, implying that a one year (= nine month) decline in schooling is associated with a 5.4 percent decline in earnings.

[32] 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.

Foreignness Index of his name at the time of labor market entry. Column 3 includes two F-indices on the right-hand side – one calculated at birth and the other at labor market entry. By controlling for the F-index at birth, 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 earnings and equally important in predicting unemployment, suggesting that employers hired or promoted workers differently based on the ethnic content of their name.[32]

Even within households, there may be an association between name foreignness and aspects of the home environment, including parental income and degree of cultural assimilation. In column 5, we restrict our analysis to brothers who were born within two years of each other and thus likely grew up in a very similar setting. The results are quantitatively similar in this subsample, consistent with the idea of names as a signal of identity even for men with otherwise similar levels of cultural assimilation.

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." Table 7 considers two measures of the foreignness of a man's spouse. The first panel examines whether the spouse, herself, was born abroad, a relatively rare outcome for second-generation immigrant

[32] 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.

men, only 5.4 percent of whom were married to a foreign wife.[33] The second panel instead measures 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. 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 0.8 percentage point increase in the probability of having a foreign-born spouse (on a base of 5.4 percent) and a 1.5 point increase the F-index of his spouse's name.[34]

Appendix Table 5 considers a series of additional labor market outcomes, including the subcomponents of annual earnings (hourly wages, weeks worked during the year, and hours worked during the week) and various forms of employment. Consistent with the effect of name foreignness on unemployment at the time of the Census, brothers 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 effect of name foreignness on employment is equal and opposite to the effect on unemployment, implying no effect of names on the probability of being out of the labor force in this prime-age sample (the omitted category). Brothers with foreign names were no more likely to hold a public works job through the New Deal, despite facing higher unemployment, which could reflect some discrimination in access to public relief employment.

[33] 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.

[34] 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).

VI. Additional measures of cultural assimilation: 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 contemporary view 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. 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. 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).

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.[35]

[35] 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

Figure 7 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 marriage). 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 first and second-generation 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 origin, controlling for group size and sex ratio at the state level.[36] 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

---

bias the endogamy rate for country-of-origin groups that tended to migrate in families, rather than as single individuals.

[36] 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.

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 higher female endogamy, because one route to citizenship is 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 6 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. Conclusion

We use complete count historical census data to study the cultural assimilation of immigrants during the Age of Mass Migration. We document that immigrants chose less foreign names for their sons and daughters as they spent more time in the US, halving the difference in name choice with natives after 20 years in the US. Within a single generation, many immigrants also learned to speak English, applied for US citizenship, and married spouses from different origins. Receiving a "native sounding" name had positive consequences for the children of immigrants. We follow over a million children of immigrants from their childhood to adulthood and find that brothers with less foreign names completed more years of schooling, earned more and were less likely to be unemployed.

These findings suggest that, in this formative period of US history, immigrants assimilated quite considerably into US society. 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 can generate discrimination from teachers and employers. This tradeoff is still salient for immigrants in the US and ethnic minorities around the world today.[37]

[37] See, for example, recent articles about the complicated decision of how to name one's children in Asian American and Muslim American communities (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.

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.

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 Immigation 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.

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.

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

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

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

Lieberson, 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. 2012. “Taste-Based Discrimination Evidence from a Shift in Ethnic Preferences after WWI.” *Explorations in Economic History* 49 (2): 167-88.

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.

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.

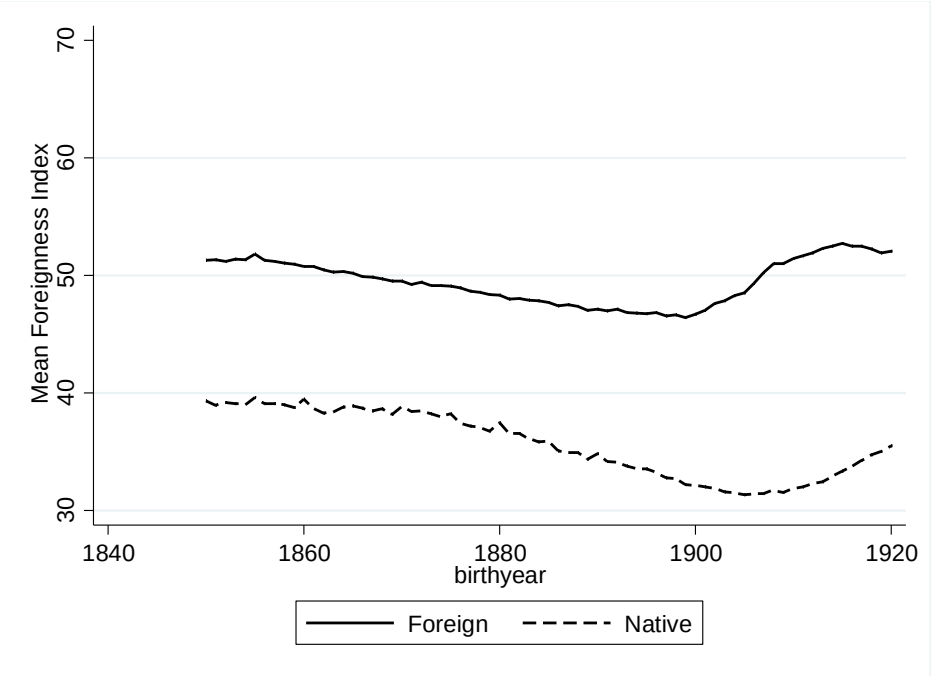
Watkins, Susan Cotts, ed. *After Ellis Island: Newcomers and Natives in the 1910 Census*. New York: Russell Sage Foundation, 1994.

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.

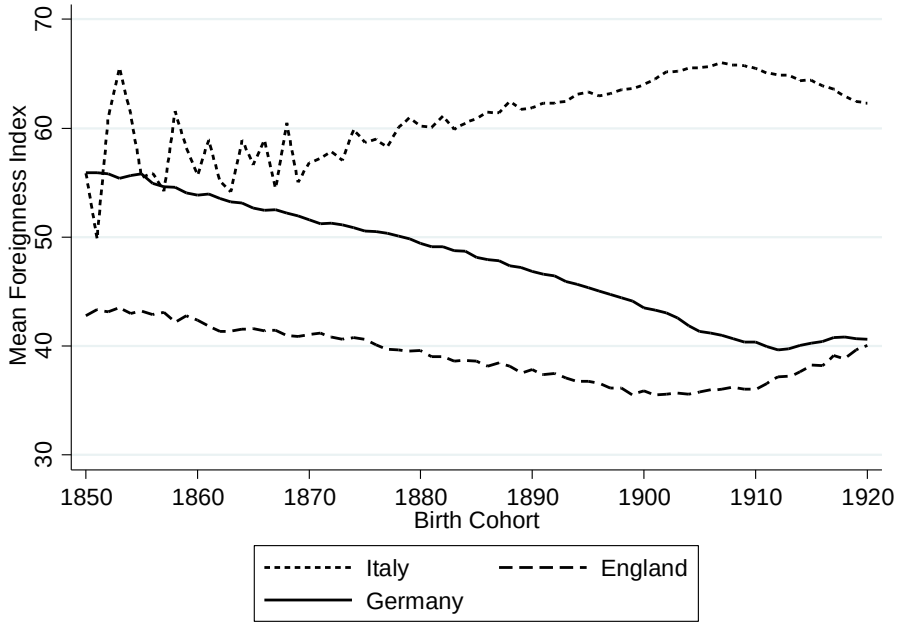
Zelinsky, Wilbur. “Cultural Variation in Personal Name Patterns in the Eastern United States.” *Annals of the Association of American Geographers* 60, no. 4 (1970): 743-769.

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

A. Sons of immigrant and native-born fathers

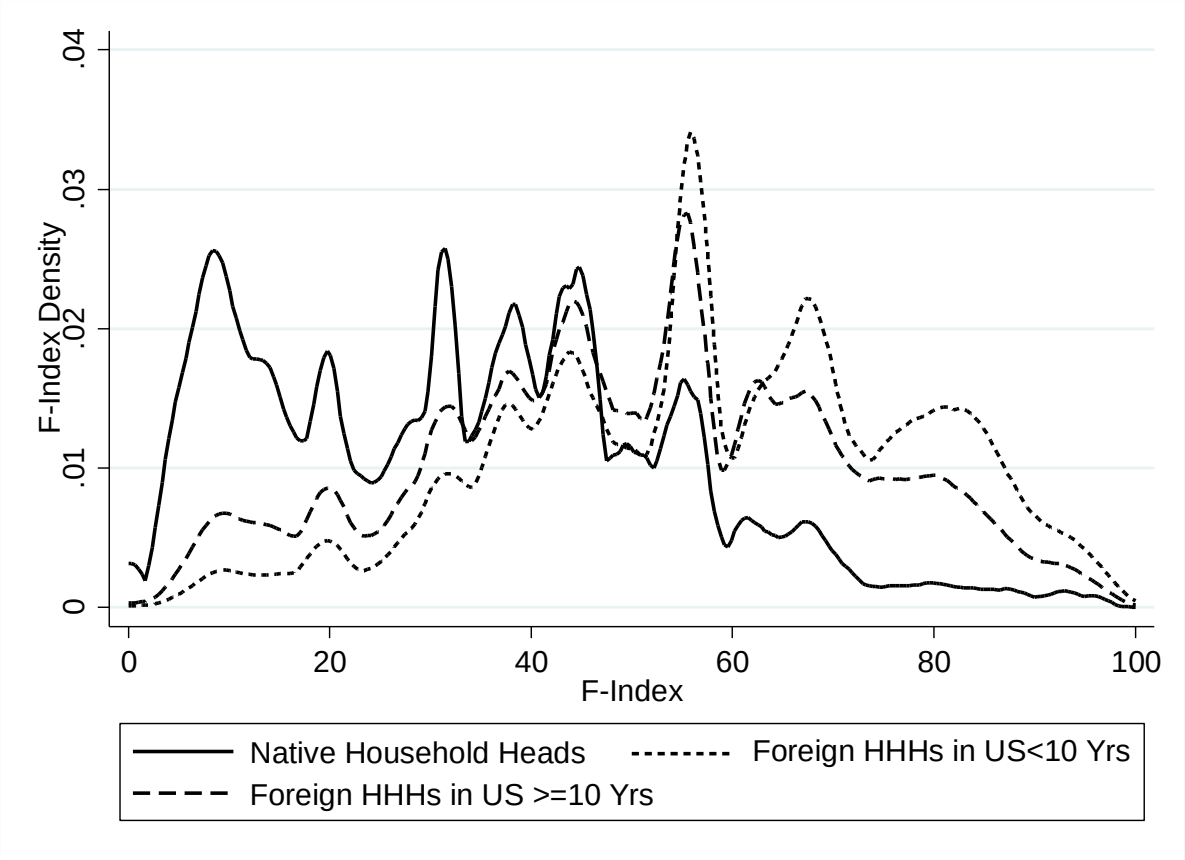


B. Examples of countries with large changes in naming (Germany), little changes in naming (Italy) and little differentiation from natives in naming (England)



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*. Belgium, Denmark, France, Ireland, Sweden and Switzerland follow the pattern in Panel A; Austria, Finland and Portugal follow the pattern in Panel B and Wales follows the pattern in Panel C.

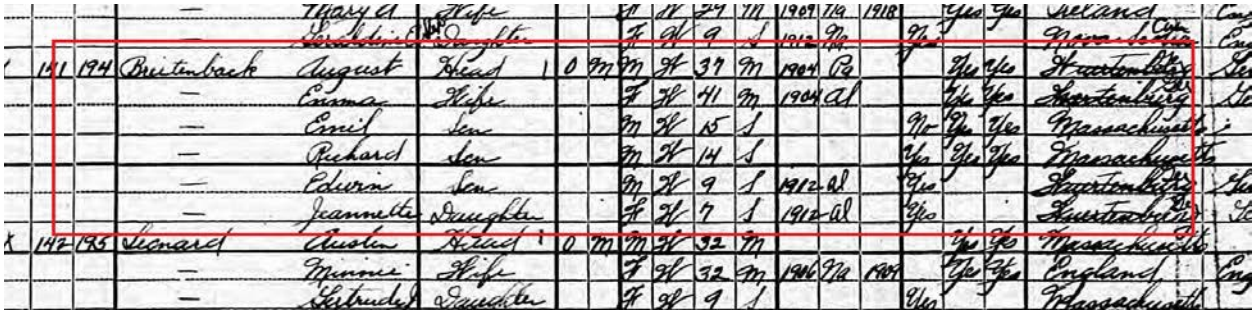
Figure 2: Kernel density estimates of name Foreignness Index in 1920, Children of native-born or foreign-born in the US more/less than 10 years



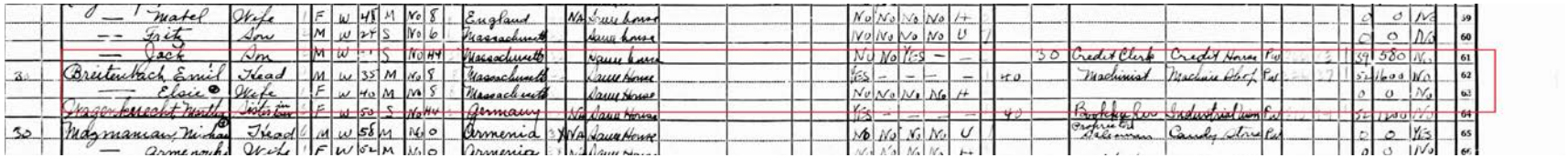
Notes: Sample includes non-black children born in the US (outside of the South), living with their parents, and between the ages of 0-18 in the 1920 Census. Households are classified as native- or foreign-born based on the place of birth of the household head. Foreign-born households are further divided by time spent in the US (more/less than 10 years).

Figure 3: Census manuscripts for the Breitenbach family

A. Childhood household in 1920



B. Emil Breitenbach in 1940



C. Richard Breitenbach in 1940

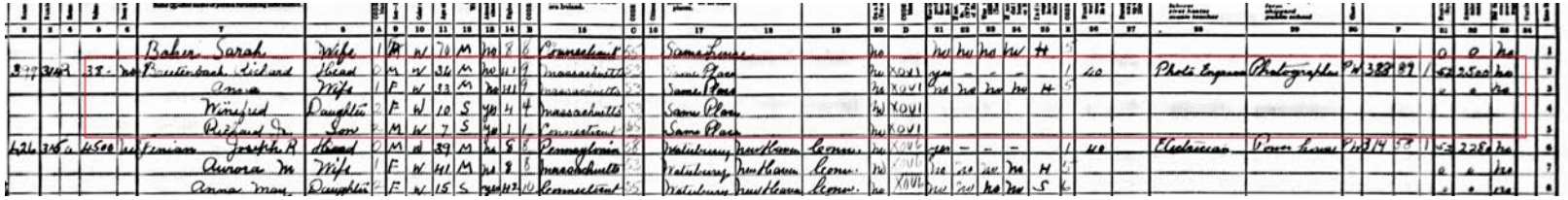
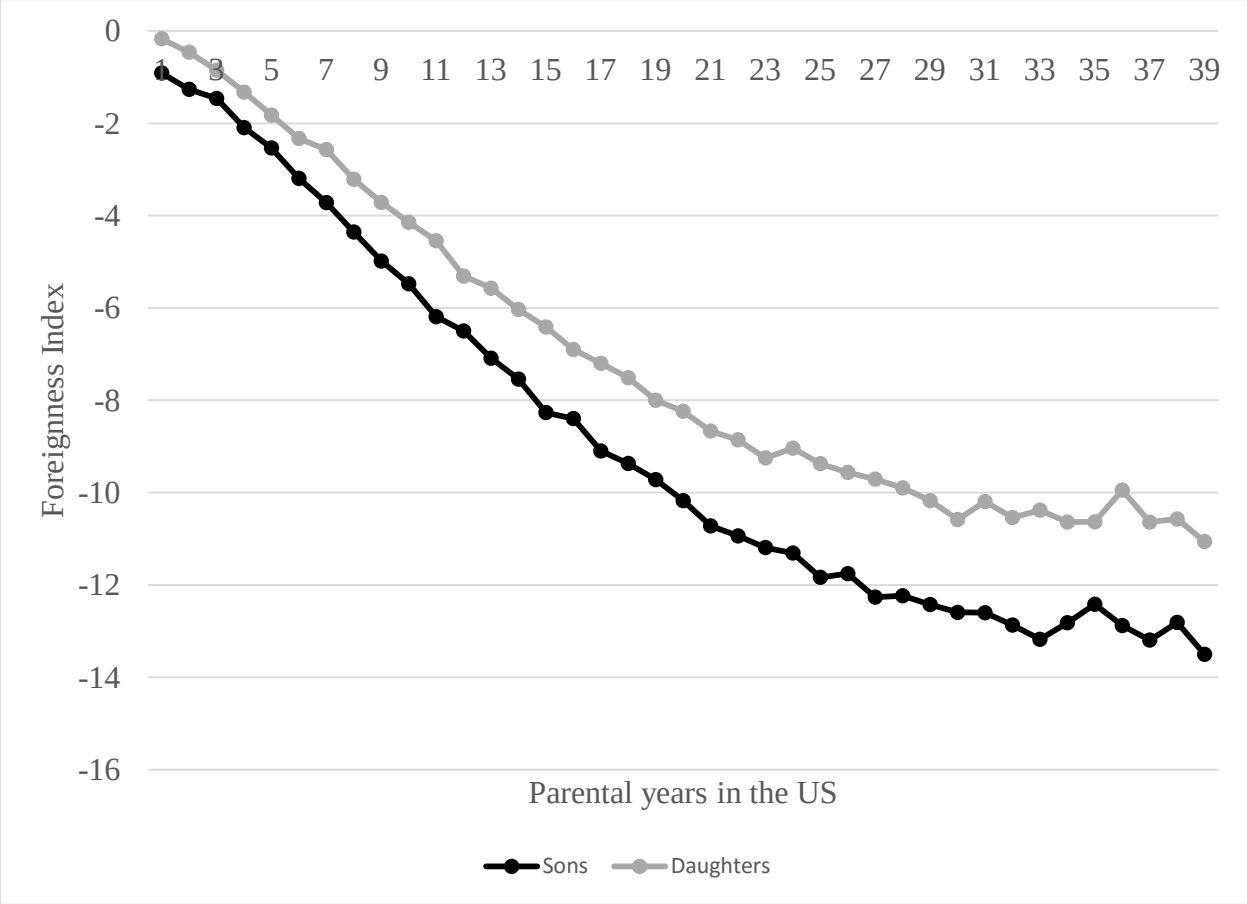
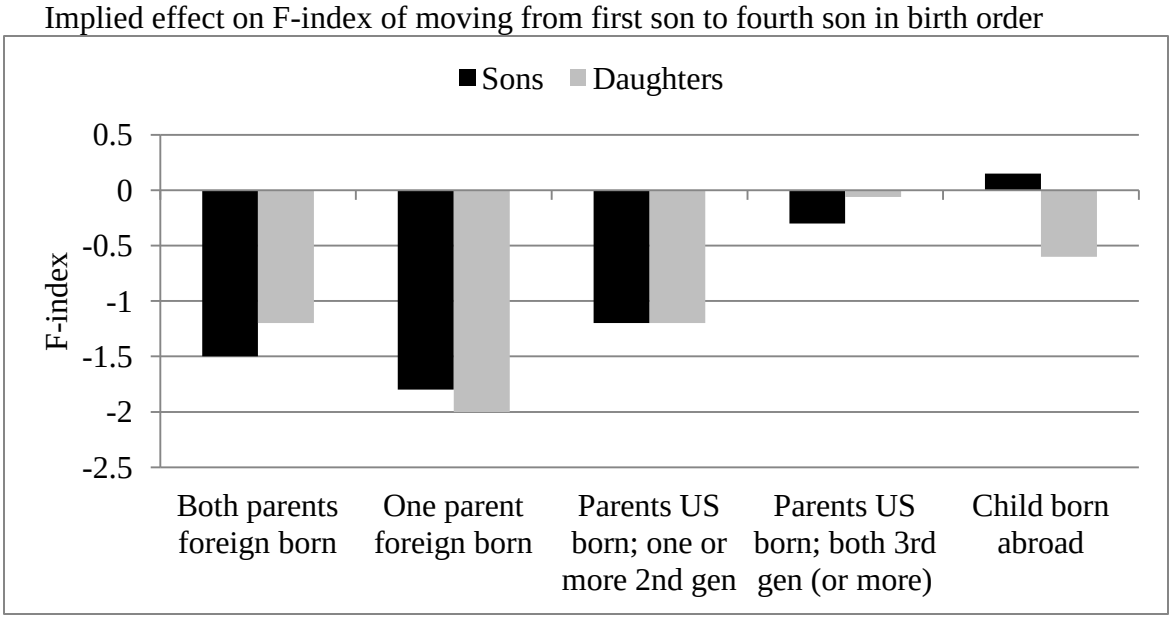


Figure 4: Immigrants selected less foreign names for children after spending time in US, (Dependent variable = F-index)



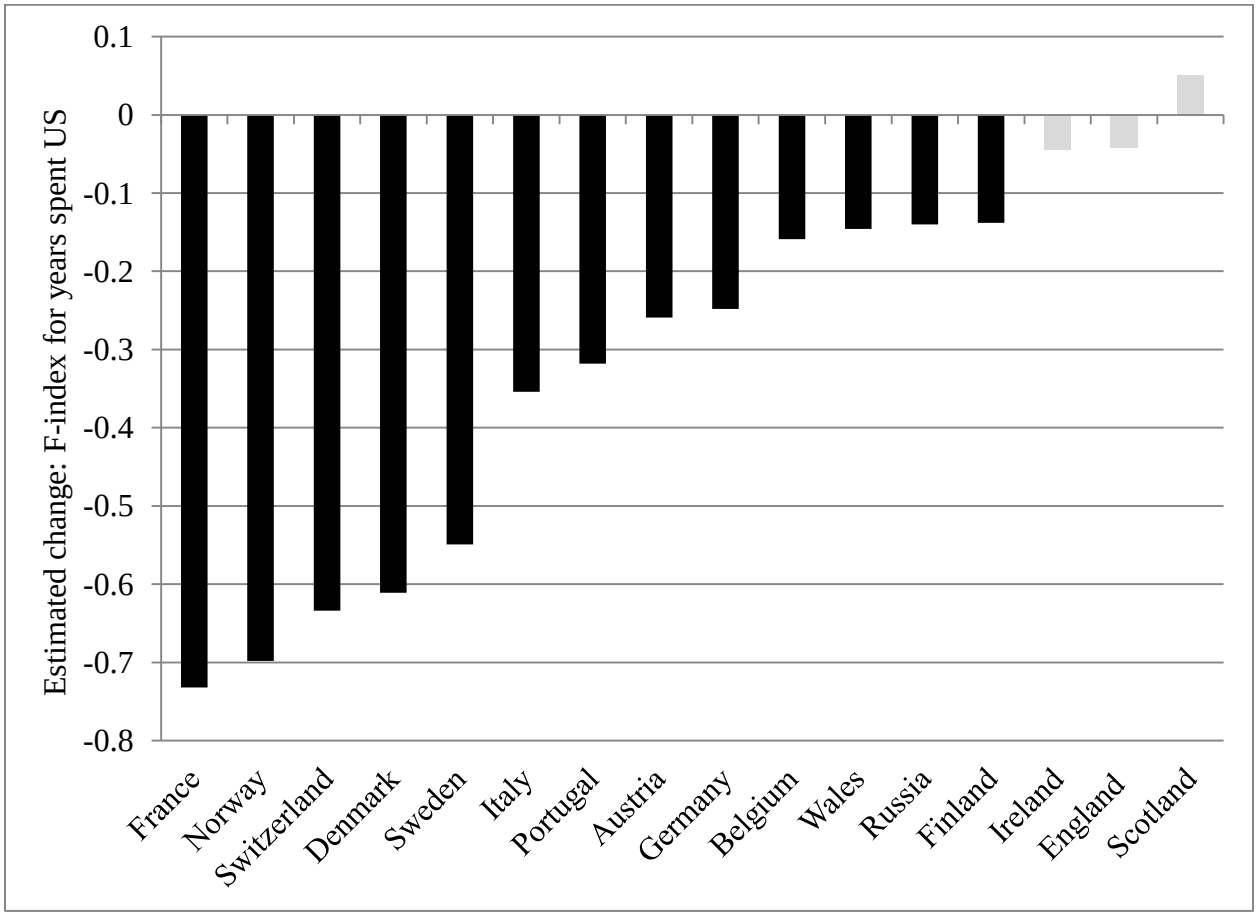
Notes: The graph reports coefficients from estimates of Equation 1, a regression of the F-index on a set of dummy variables for years that the household head had spent in the US by the time of the child's birth. Regressions also include dummy variables for child's age and a set of household fixed effects. Data from 1920 complete-count Census. Sample includes children aged 0-18 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) = 2,130,352; *N* (daughters) = 2,081,724).

Figure 5: Children of immigrants received less foreign names later in birth order, but children of other household types did not



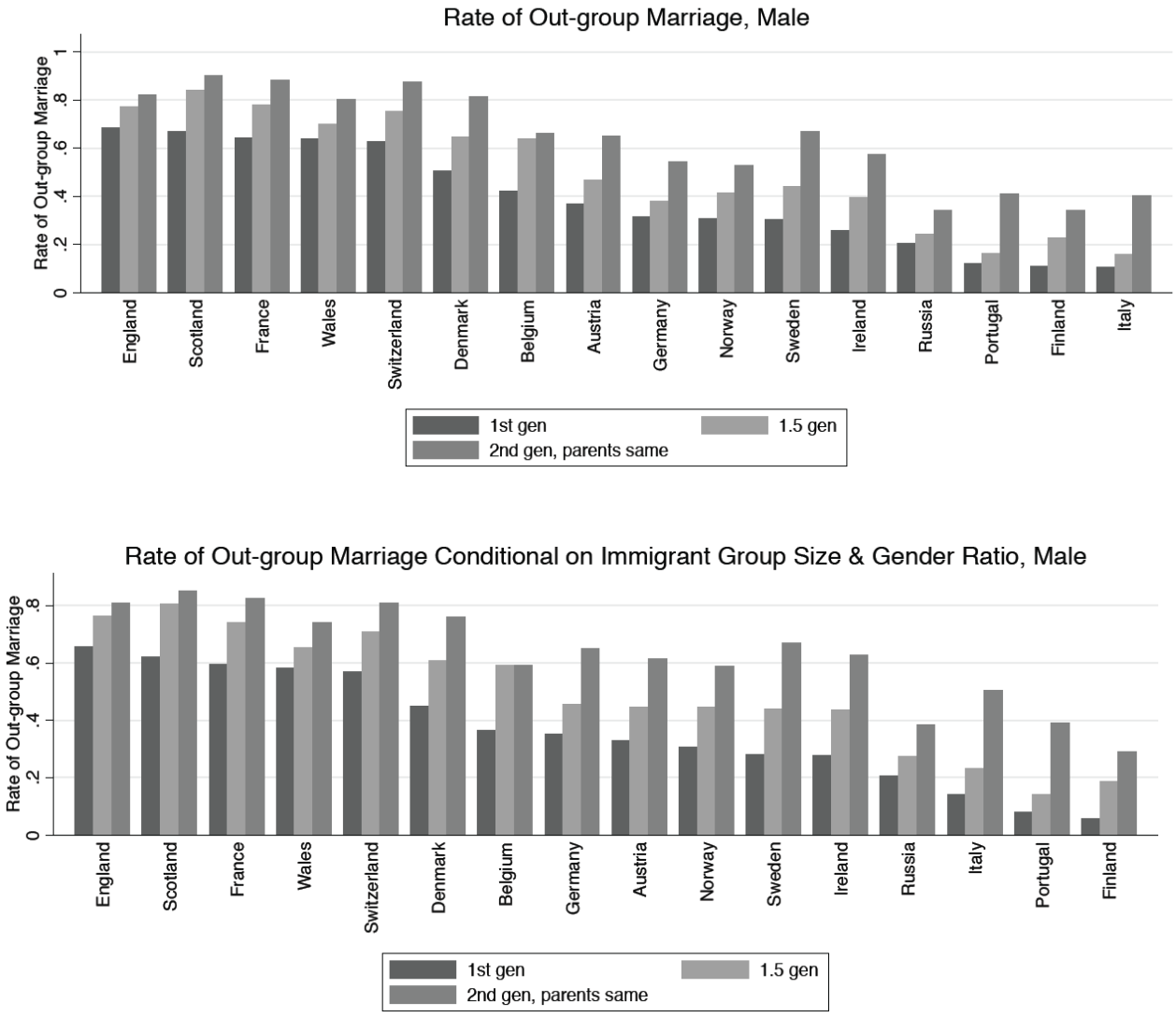
Notes: Data from the complete-count 1920 census. Bars report the implied effect on F-index of moving from first son to fourth son in birth order. Sample includes non-black children aged 0-18 living with their parents in a non-southern state. Spouse of household head (mother) must be less than 43 years old. Underlying regression also controls for household fixed effects, child's birth year, whether child has same name as father, and the overall frequency of the name.

Figure 6: Effect of parental years in US on name foreignness, by sending country



Notes: Reported coefficients from estimates of Equation 1, controlling for household fixed effects, child's birth year, whether child has same name as father, and overall name frequency. Dark bars are statistically significant at the 5 percent level. Data is from the complete-count 1920 census. Sample includes children aged 0-18 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 7: 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.

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

Most foreign (F-Index >0.90)	Most neutral (0.5 < F-Index < 0.52)	Most native (F-Index <0.025)
	A. Male names	
Vito	Orlando	Gaylord
Mario	Benjiman	Doyle
Hyman	Murray	Clay
Pasquale	Otto	Lowell
Isidor	Theodor	Dale
Nick	Herman	Wayne
	B. Female names	
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	48.43	53.35	-4.824***
	(21.72)	(14.91)	(0.130)
Foreignness Index, 1940	48.26	53.08	-4.875***
	(21.29)	(15.64)	(0.112)
# characters, first name	5.767	5.713	0.054***
	(1.254)	(1.296)	(0.002)
Age	8.809	8.858	-0.048***
	(3.699)	(3.697)	(0.005)
Number of Siblings	3.969	3.972	-0.002
	(1.966)	(1.978)	(0.002)
Number of Brothers	2.505	2.521	-0.015***
	(1.324)	(1.348)	(0.002)
Rank in brother order	1.782	1.743	0.038***
	(0.977)	(0.954)	(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; we restrict to men in the relevant age range with fathers born in one of our sixteen countries.

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	-0.435***	-0.290***	-0.226***	-0.095***
	(0.010)	(0.013)	(0.013)	(0.002)
Birth order controls	N	Y	Y	N
Name control	N	N	Y	Y
N	1,993,899	1,993,899	1,993,899	1,993,899

Notes: Data is from the complete-count 1920 census. Sample includes sons aged 0-18 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, as well as a set of household 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	-0.418***	-0.484***	-0.435***	-0.419***
	(0.011)	(0.027)	(0.014)	(0.014)
N	1,809,006	297,657	809,715	1,294,202

Notes: Data from complete-count 1920 census. Sample includes sons age 0-18 living with their parents who were born in a non-southern state. Regressions control for dummy variables for child's age, as well as a set of household 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

Years in US	-0.323***
	(0.031)
Years in US *x* F-index of last name	-0.234***
	(0.043)
N	795,453

Notes: Data from the complete-count 1920 census. Sample includes sons with a foreign-born head of household who were between the ages of zero and 18 who were born outside the South and lived with their parents. Regressions control for dummy variables for child's age, as well as a set of household fixed effects. To observe complete birth order, sample restricted to households in which mother is less than 43 years old.

Table 6: The effect of name foreignness on education, earnings and unemployment

	No household fixed effects			With household fixed effects	
	(1) Baseline	(2) Add controls	(3) Add F-index at 20	(4) Full Sample	(5) Brothers 1-2 years apart
	Panel A				
	Dependent variable: Highest grade (Mean = 10.26)				
F-index	-0.009***	-0.009***	---	-0.006***	-0.008***
	(0.0001)	(0.0001)		(0.0003)	(0.0007)
N	1,054,765	972,211		972,211	168,515
	Panel B				
	Dependent variable: =1 if unemployed *x* 100 (Mean = 9.5)				
F-index	0.026***	0.026***	0.015***	0.027***	0.017***
	(0.001)	(0.002)	(0.005)	(0.003)	(0.008)
F-index at 20			0.012**		
			(0.005)		
N	988,383	910,936	910,936	910,936	157,531
	Panel C				
	Dep. Variable: Annual earnings (Mean=$21,057)				
F-index	-10.61***	-12.65***	-2.67	-6.51**	-13.67**
	(0.910)	(0.958)	(2.88)	(2.70)	(61.48)
F-index at 20			-11.04***		
			(3.00)		
N	673,810	620,413	620,413	620,413	107,045

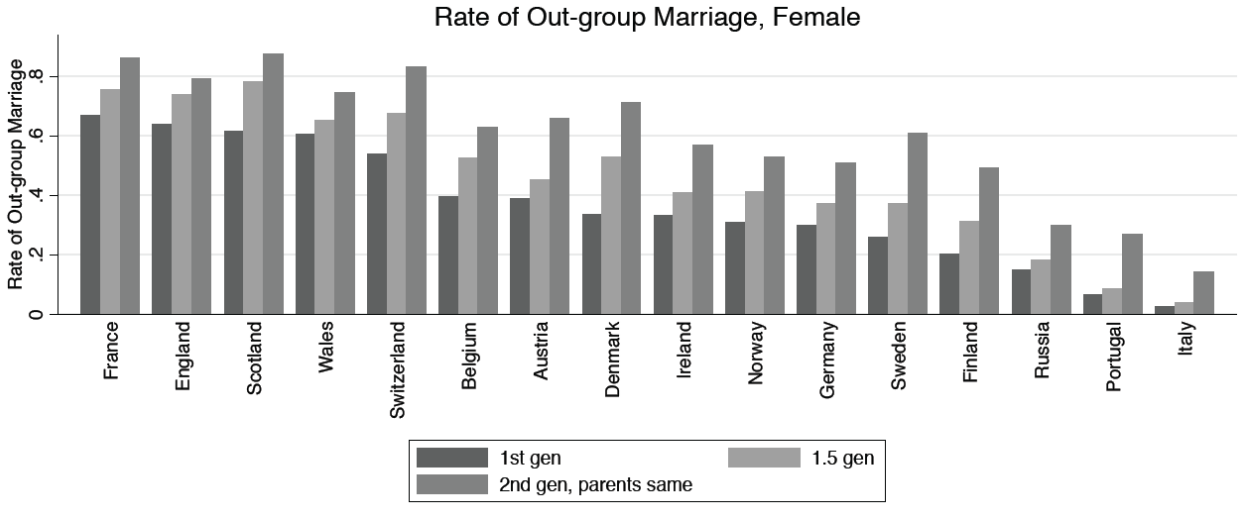
Note: Sample includes men matched between 1920 and 1940 complete-count Censuses. Men must be 3-18 in 1920, born outside the South and living at home with parents in 1920 in a household whose head was foreign-born. Panel C is further restricted to men with non-zero earnings who were not self-employed in 1940. All regressions control for a vector of dummies for child's age in 1940. Columns 2-5 control for parental years in the US and child's rank in the birth order. Columns 4-5 add household fixed effects.

Table 7: The effect of name foreignness on spouse characteristics

	No household fixed effects			With household fixed effects	
	(1) Baseline	(2) Add controls	(3) Add F-index at 20	(4) Full Sample	(5) Brothers 1-2 years apart
	Panel A				
	Dependent variable: =1 if spouse foreign born *x* 100 (Mean = 5.4)				
F-index	0.040***	0.036***	0.037***	0.039***	0.032**
	(0.002)	(0.002)	(0.005)	(0.005)	(0.013)
F-index at 20			-0.002		
			(0.005)		
N	488,918	446,327	446,327	488,918	76,771
	Panel B				
	Dep. Variable: F-index of spouse (Mean = 43.5)				
F-index	0.078***	0.073***	0.066***	0.082***	0.070***
	(0.001)	(0.001)	(0.005)	(0.005)	(0.011)
F-index at 20			0.008*		
			(0.005)		
N	475,331	433,995	433,995	475,331	74,664

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 age in 1940. Columns 2-5 control for parental years in the US and child's rank in the birth order. Columns 4-5 add household fixed effects.

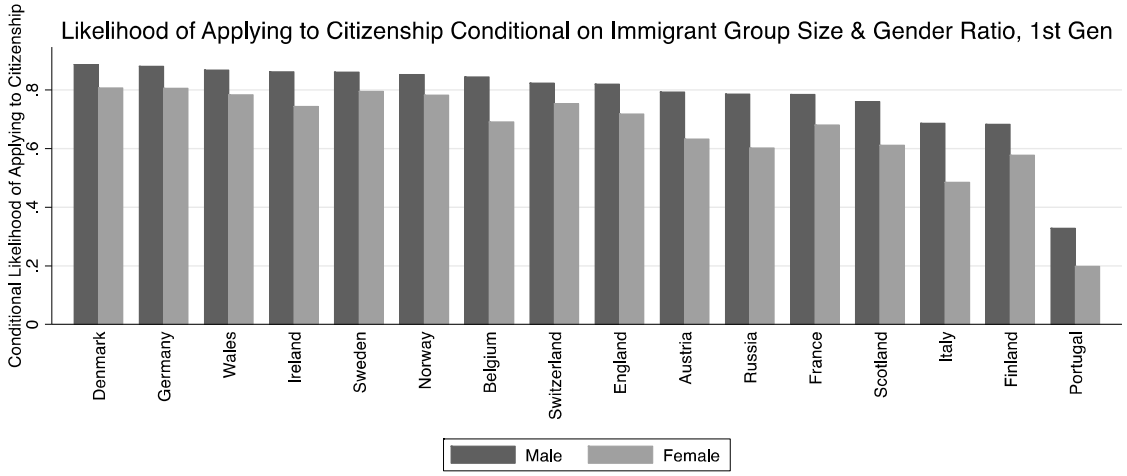
Appendix Figure 1:
Share of first and second generation immigrant women in out-group marriage, by country of origin, 1930



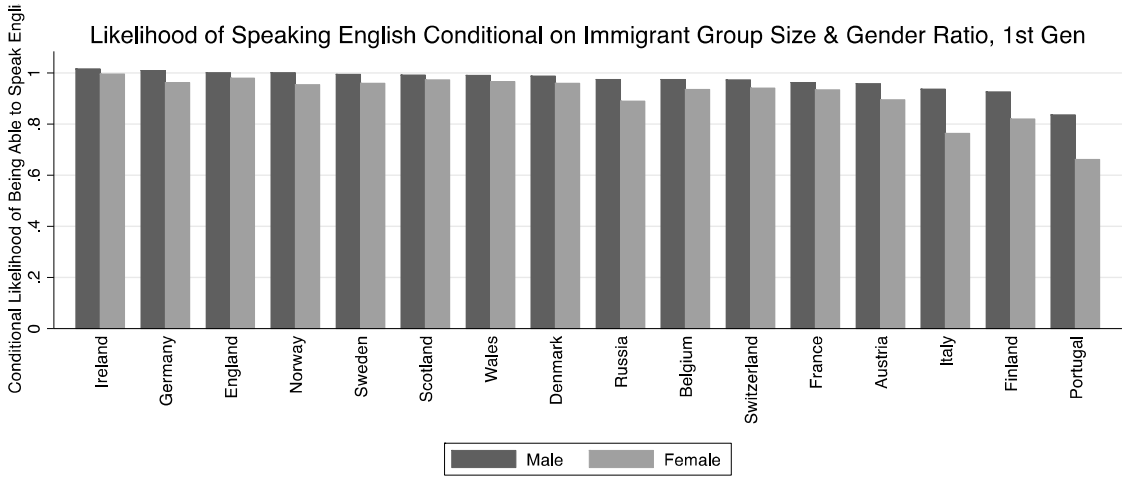
Note: Figure based on women in IPUMS 5% sample of 1930 census who are currently married and whose age at first marriage occurred after arrival in the US. Women whose spouse (or spouse's parents) were born in the same country of origin as she (or her parents) are considered to be in an 'endogamous' marriage. 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 7 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 7 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	-0.366***	-0.270***	-0.326***	-0.049***
	(0.013)	(0.017)	(0.013)	(0.002)
Birth order control	N	Y	Y	Y
Name controls	N	N	Y	Y
N	1,935,780	1,935,780	1,935,780	1,935,780

Notes: Data is from the complete-count 1920 census. Sample includes daughters aged 0-18 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 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	-0.460***
N = 1,993,899	(0.005)
Adjust names with NYSIIS	-0.397***
N = 1,993,578	(0.004)
Fix F-index in 1900	-0.436***
N = 1,993,899	(0.005)
F-index calculated in 1920	-0.469***
N = 1,986,457	(0.005)
F-index, by state of birth	-0.461***
N = 1,906,120	(0.005)
F-index calculated by country	-0.715***
N = 1,477,537	(0.008)
F-index, 2nd gen foreign	-0.272***
N = 1,986,432	(0.004)
Include South	-0.440***
N = 2,096,276	(0.004)

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 zero and 18 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, 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
A. Both parents foreign born (Mean F-index = 56.1)			
=1 if second born	-0.654***	-0.649***	-0.337***
	(0.088)	(0.093)	(0.133)
=1 if third born		-1.199***	-0.905***
		(0.152)	(0.130)
=1 if fourth or more			-1.396***
			(0.182)
N	515,881	375,383	361,366
F-test: 2nd son vs. 3rd son		30.86	33.31
F-test: 3rd son vs. 4th son			22.44
B. Parents born in US; At least one parent second generation immig.(Mean F-index = 36.7)			
=1 if second born	-0.676***	-0.741***	-0.487***
	(0.093)	(0.113)	(0.135)
=1 if third born		-0.797***	-0.851***
		(0.184)	(0.173)
=1 if fourth or more			-0.774*
			(0.245)
N	413,171	237,987	199,799
F-test: 2nd son vs. 3rd son		0.23	7.35
F-test: 3rd son vs. 4th son			0.31
C. Parents born in US; both parents third generation immig. or more (Mean DV = 33.5)			
=1 if second born	-0.357***	-0.149***	-0.097
	(0.059)	(0.070)	(0.082)
=1 if third born		-0.085	-0.015
		(0.114)	(0.106)
=1 if fourth or more			0.204
			(0.150)
N	1,080,023	648,364	553,384
F-test: 2nd son vs. 3rd son		0.71	1.04
F-test: 3rd son vs. 4th son			6.47

Notes: The sample includes non-black sons aged 0-18 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)
Baseline – 2 parents foreign	***-0.493******
N = 1,664,614	***(0.049)***
Both parents arrived as adults	-0.434***
N = 1,056,891	(0.061)
One or more parents arrived as child	-0.565***
N = 607,723	(0.080)
Parents from same sending country	-0.454***
N = 1,529,520	(0.050)
Parents from different sending countries	-0.828***
N = 135,094	(0.197)
Baseline – 1 parent foreign	***-0.576******
N = 732,883	***(0.080)***
Only mother foreign	-0.408***
N = 271,991	(0.131)
Only father foreign	-0.641***
N = 460,892	(0.102)

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 zero and 18, 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, a dummy equal to one if child has same name as father, and name frequency.

Appendix Table 5:
Name foreignness and other labor market outcomes

		Coefficient on F-index	
Dependent variable	Mean	Without HH FE	With HH FE
Hourly wage (*N* = 552,058)	11.38	0.0007 (0.001)	0.002 (0.004)
Weeks worked per year (*N* = 630,290)	45.07	-0.009*** (0.0007)	-0.009*** (0.002)
Hours worked per week (*N* = 592,127)	43.79	-0.011*** (0.0008)	-0.012*** (0.002)
Employed (*N* =910.936)	0.850	-0.0003*** (0.00002)	-0.0003*** (0.00004)
Public emergency work (*N* = 774,679)	0.042	0.00007*** (0.00001)	0.00004 (0.00003)
Self employed (*N* = 744,391)	0.164	0.000001 (0.00002)	-0.000005 (0.00006)

Note: Sample includes men matched between 1920 and 1940 complete-count Censuses. Men must be between the ages of 3-18 and living at home with their parents in 1920. Sample restricted to men whose fathers were foreign-born. The first row is further restricted to men who were not self-employed in 1940. All regressions control for child's age 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 6:
Pairwise correlations of various measures of cultural assimilation

	Variables	Out-group marriage	Speaks English	Applied for citizenship
A.	**16 sending countries**			
1	Out-group marriage rate (+)			
2	Able to speak English (+)	0.642***		
3	Applied for citizenship (+)	0.476*	0.944***	
4	Average F-index of sons (-)	-0.489*	-0.767***	-0.666***
B.	**Individual (*N* = 23,043)**			
1	Out-group marriage rate (+)			
2	Able to speak English (+)	0.065***		
3	Applied for citizenship (+)	0.075***	0.258***	
4	Average F-index of sons (-)	-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 begin by identifying children living at home in the full 1920 Census index, from FamilySearch.org, whose father was born in one of 16 large European sending countries.[38] We restrict to boys between ages 3 and 18 in 1920. We select age 18 as the upper limit because more than 85 percent of the 18 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.

(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

[38] 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.

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, we use the Restricted Full Count 1940 Census available via the Minnesota Population Center to attach outcome variables in 1940 to matched cases. These outcomes include education, income, and employment status.