Discrimination and the Returns to Cultural Assimilation in the Age of Mass Migration

Ran Abramitzky        Leah Boustan        Katherine Eriksson        Stephanie Hao
Stanford and NBER     Princeton and NBER  UC-Davis and NBER         Princeton

November 2019

Abstract: We document that, in the early twentieth century, children of immigrants who were given more foreign first names completed fewer years of schooling, earned less, and married less assimilated spouses. However, we find few differences in the adult outcomes of brothers who were given more foreign versus more American-sounding first names. This pattern suggests that the negative association between ethnic names and adult outcomes in this era did not stem from discrimination on the basis of first names (although teachers and employers may have discriminated using other ethnic cues), but instead reflects household differences associated with cultural assimilation.
I. Introduction

Immigrants who are identifiably foreign due to their accent, dress, name or other ethnic markers may be subject to discrimination at school or in the labor and housing markets. Because of this cost to retaining their cultural identity, immigrants often take steps to culturally assimilate into their adoptive countries. Indeed, in the early twentieth century, many immigrants to the U.S. changed their own names and/or chose more American sounding names for their children as they adapted to the country (Biavaschi, Giulietti, and Siddique, 2016; Abramitzky, Boustan, Eriksson, Forthcoming). In this paper, we study whether there is a negative association between foreign sounding first names and outcomes, and whether this negative association is indicative of a causal effect of having an ethnic-sounding first name, or may instead reflect the return to cultural assimilation.

We start by constructing a large sample of the children of immigrants observed by linking observations of their childhood households in the 1920 Census to their adult outcomes in the 1940 Census. The parents in our sample immigrated to the United States during the Age of Mass Migration from Europe (1850-1913), one of the largest and most formative migration episodes in American history. During this era, which we discuss in more detail elsewhere (Abramitzky, Boustan and Eriksson 2012, 2014, 2016 and Abramitzky and Boustan 2017), about 30 million immigrants from countries like Ireland, Germany, Italy and Russia moved across the Atlantic.

In a raw comparison, we find that men who were given more foreign-sounding first names attained fewer years of schooling, earned less, and experienced lower levels of employment than those with more American-sounding names. Furthermore, these men were more likely to marry women who were also less culturally assimilated (a foreign spouse or a spouse with more foreign name herself). These baseline comparisons could reflect two channels: differences in the way that
teachers, employers and others treat a Luigi versus a John (discrimination, both taste-based and statistical), and/or the family attributes that are correlated with the choice of an Italian name (particularly, the degree of cultural assimilation). Our paper tries to disentangle the role of family attributes from the role of discrimination against students or workers with foreign sounding names.

We find that the associations between name foreignness and adult outcomes remain economically meaningful and statistically significant when adding an increasingly detailed set of fixed effects – starting with country-of-origin, and then moving to surname, and childhood county of residence (in 1920) – implying that second generation immigrants with foreign first names fared worse than others who shared similar backgrounds.

Yet, the relationship between name foreignness and adult outcomes diminishes substantially when adding household fixed effects. We find small differences between brothers with more or less foreign-sounding first names, even when focusing on brothers born a few years apart who otherwise experienced a similar process of cultural assimilation at home. This pattern suggests that observed disparities in economic outcomes by name foreignness were largely driven by differences between households that took steps to assimilate (e.g., by giving their children more American names) and those that did not. In particular, comparisons between brothers rules out discrimination in the labor market on the basis of first names but this null result still could be consistent with ethnic discrimination on the basis of other ethnic signals (including last name, accent, neighborhood of residence, etc.).

Our paper contributes to the literature on discrimination against immigrants, which is often measured using differential treatment by name. Our results are consistent with Goldstein and Stecklov (2015), who find that men with foreign-sounding first names had lower occupation-based earnings in 1930 even after controlling for surname, one (coarse) measure of family background.
Immigrants who choose to change their own first or last name seem to benefit economically from more American-sounding names, although these name changes might be one of many steps that a worker takes in order to get ahead (Arai and Thoursie, 2009; Biavaschi, Giulietti, and Siddique, 2016; Carneiro, Lee, and Reis, 2015).\(^1\)

The challenge in this literature (recognized by authors) is that names could be correlated with other aspects of family background or, in the case of optional name changes, with individual motivation. Our findings suggest that the negative association between ethnic names and outcomes at least partially reflects the return to cultural assimilation, as the children of families that choose to assimilate might be different from those that retain their cultural markers. Oreopoulos (2011) addresses possible selection on family background and individual motivation by sending out resumes with randomized foreign- and Anglo-sounding names in response to job postings in Toronto, Canada. Consistent with our results, he does not find lower callback rates for workers with foreign-sounding first names, but does find lower rates for workers with foreign-sounding last names.\(^2\) We complement Oreopoulos by looking at real labor market success as measured by wages and employment, by considering a variety of other outcomes like education and matching in the marriage market, and by studying an important period in US history, during which ethnicity

---

\(^1\) In a different context, Rubinstein and Brenner (2014) show that Israeli children with an Ashkenazi mother and a Sephardic father (and thus a Sephardic-sounding last name) are penalized relative to Israeli children with a Sephardic mother and an Ashkenazi father.

\(^2\) In a related literature, Bertrand and Mullainathan (2004) document lower call-back rates for resumes assigned a distinctively black name, whereas Fryer and Levitt (2004) instead find that having a black name is not associated with lower levels of education or early childbearing after controlling for measures of family background. Cook, Logan and Parman (2016) find a mortality advantage for men who received distinctively black names in the nineteenth and early twentieth centuries and point to suggestive evidence that receiving a black name is correlated with cultural factors at the household level.
is said to have played a large role in employment (the classic “No Irish Need Apply” signs) and before anti-discrimination laws were passed.

II. Data and definitions

A. Measuring the foreignness of given names

Historical Census data contain individual records with information 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 an index of the relative probability that a given name was held by a foreigner versus an American (used also in Abramitzky, Boustan and Eriksson (Forthcoming) and based on Fryer and Levitt’s (2004) index of black names). In particular, the Foreignness Index is defined as:

\[
\text{Foreignness Index}_{name} = 100 \cdot \frac{\# \text{ foreigners}_{name}}{\text{total} \# \text{ foreigners}} - \frac{\# \text{ foreigners}_{name}}{\text{total} \# \text{ foreigners}} + \frac{\# \text{ natives}_{name}}{\text{total} \# \text{ natives}}
\]

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 American name) and a value of 100 assigned to a child whose first name is distinctively foreign.

In our main analysis, the F-index is calculated by contrasting the names of all foreign-born residents to all US-born residents for the twenty years prior to an individual’s birth year (to account for changes in naming trends). We also consider alternative measures that instead create country-specific name indices (e.g., all German-born versus all non-German-born). To address the entry of second-generation immigrants into the pool of the US born over time, we also present results using an F-index that is fixed at a point in time (calculated for the 1895 to 1905 birth cohorts), rather
than varying by birth year. In all cases, we focus on the sample of men with first names held by at least 100 others in the relevant years to calculate a reliable F-index.

**B. Creating a linked Census sample**

We create a matched dataset of the US-born sons of immigrants observed in childhood and adulthood. We first identify children living at home with their parents in the 1920 Census, and then link these individuals to their adult outcomes in the 1940 census. By creating this matched sample, we are able to compare siblings in 1940 who shared the same household in 1920. We use the Complete Count Census data accessed via the NBER server and provided by the Minnesota Population Center. We link men across censuses by first and last name, age, and state of birth using the Abramitzky, Boustan and Eriksson algorithm (2012, 2014), and we explore alternative matching algorithms in the online appendix (see Abramitzky, et al. 2019). We note that the presence of false positive links would bias the coefficients of interest in regressions with household fixed effects toward the OLS coefficients because picking two random people due to bad matches and calling them brothers would tend to recover the population-wide estimates. We restrict our attention to men between the ages of 3 and 15 in 1920, who were young enough to be living at home with their parents in 1920 and would have been of prime labor market age in 1940. Our primary linked sample contains more than 800,000 men, around 300,000 of whom are in sets of matched brothers. We achieve a match rate of 25.5 percent, which is slightly higher than the standard for historical matched samples (e.g., Ferrie, 1996; Abramitzky, Boustan and Eriksson, 2012).

---

3 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,
Appendix Table 1 compares the men in our matched sample to all white second-generation immigrant men in 1940 on a number of baseline characteristics (note that parents’ birthplace is a sample-line characteristic in 1940 available for only 5 percent of the population). Men with common names are less likely to match, as are men with lower numeracy who are more likely to misreport their age. Thus, as is typical for matched samples, we find that men in our linked dataset have higher socio-economic status than in the full population, as measured by employment and income. We present results that reweight the matched sample to match the population on family background. Furthermore, we find that the core association between name foreignness and economic outcomes is similar in the matched sample and the full population (compare Appendix Table 2 to Table 1, which will be described in more detail below).

III. Foreign names and outcomes in the 1940 labor and marriage markets

Having a foreign name might have a negative effect on one’s life trajectory for a number of reasons. First, ethnic identity might be more salient to children with foreign names, and if they expect to be treated less favorably because of their foreign identity, they might perceive lower returns to education and exert less effort in school (Jensen, 2010). Second, teachers or employers might use names as a signal of ethnicity and discriminate against children or workers with foreign-sounding names (Figlio, 2005). A similar dynamic may play out in the marriage market: men with foreign names might identify more strongly with their ethnic group and prefer to find a spouse within their own ethnic community, or these men may have been overlooked or rejected by spouses from other backgrounds.

and improvements in life expectancy. Furthermore, we match a younger sample that would have lower mortality rates than adult samples.
We start by regressing the adult outcome of son $i$ from family $j$ ($y_{ij}$), namely education, earnings, employment, and spouse characteristics on the foreignness of the son’s name at birth ($FIndex_{ij}$), controlling for a vector of dummies for the son’s birth year ($\gamma_{ij}$):

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

We also control for parental years in the US and child’s place in the birth order to focus on the effect of names themselves, rather than the association between name foreignness and other aspects of family structure or parental cultural assimilation (see Abramitzky, Boustan and Eriksson, Forthcoming). We then sequentially augment this core specification with sets of fixed effects included in the vector $X$, starting with parental country of origin, and then surname, and childhood county of residence. Finally, we add household fixed effects ($\alpha_j$) in order to compare brothers who were raised in the same family but received names with a different foreignness index.

Table 1 documents that name foreignness was negatively related to educational attainment, employment and earnings in the full matched sample. In OLS, a 20-point shift in the F-index, the typical gap between the children of immigrants and US born, is associated with 2 fewer months of schooling (relative to mean years of schooling of 10 years); a 2.2-percent decline in annual earnings; and a 1-percentage point increase in the probability of unemployment (column 1). The estimated effects decline by 15 to 70 percent but remain economically meaningful and statistically significant when adding country of origin, surname, and county fixed effects (columns 2-4). These controls address the fact that immigrants from some regions, particularly those more culturally distant from the US, gave their children more foreign-sounding names and also may have had worse economic outcomes. Furthermore, more recently arrived and less assimilated immigrants clustered in particular locations, which may itself have disadvantaged their children (Eriksson and Ward, Forthcoming).
We then compare brothers who were given names with different ethnic content. Table 1 (column 5) demonstrate that the estimated effects of name foreignness on adult outcomes are substantially reduced (and become statistically insignificant in most cases) after including household fixed effects. Foreign names had a small and significant effect on educational attainment within brother pairs (20 points of F-index associated with two weeks less schooling) and no effect on earnings and unemployment. The population-based estimates seem to be picking up other differences between immigrant families that chose foreign or American sounding names for their children, rather than the causal effect of foreign names themselves.

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 US born parents today. To address the changing popularity of names over time, we use an F-index calculated for the 1895 to 1905 birth cohorts, which measures name foreignness for a fixed set of birth cohorts born immediately before the men in our sample. The results, presented in Appendix Figure 1, are similar to those obtained using the birth cohort-specific F-index above. More relevant to an employer’s

---

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

5 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 (or, alternatively, the full population in 1920); see panels (a) and (b) of Appendix Figure 4. Weighting adjusts for potential variation in match rates by ethnic group. Results are also robust to weighting the linked sample so that it matches the full population in 1920 on father characteristics including age, literacy, occupational income score, home ownership status, and urban status; see panel (c) of Appendix Figure 4.
perception of a worker’s ethnic identity might be the Foreignness Index of his name at the time of labor market entry. The idea is to compare two immigrants who received an equally foreign sounding name at birth but where one of the names became less foreign sounding by the time the immigrants reached the labor market. In Appendix Table 3, we try a specification that includes two F-indices on the right-hand side – one calculated at birth and the other at labor market entry – however the two measures are highly correlated (corr. = 0.98) and so no clear pattern emerges.

Beyond the labor market, having a foreign name may have influenced men’s marriage market outcomes. 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, US-born spouses may discriminate against men who they perceive to be “too foreign.” Table 1 considers two measures of the foreignness of a man’s spouse: whether the spouse, herself, was born abroad and the (gender-specific) 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.\(^6\) 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 5.2 percent) and a 2.4-point increase the F-index of his spouse’s first name (on a base of 45 points).\(^7\) Yet, as with the labor market outcomes,\

---

\(^6\) Until 1930, the Census asked all respondents about parents’ birthplace, which would allow us to classify whether spouses were 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. Note that birthplace of spouse is a variable available for all married men with spouse present in the sample, whereas F-index of spouse is only available for men whose spouse’s name was held by at least 100 other women in the relevant period.

\(^7\) 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).
neither relationship remains statistically significant after adding household fixed effects (although confidence intervals are large).

Appendix Figure 2 considers a series of additional labor market outcomes, including hours and weeks worked per year and the log weekly wage, as well as whether or not an individual was self-employed or employed by the government in a public works job.\(^8\) Consistent with the negative association between name foreignness and employment at the time of the Census, men with more foreign names work less time during the year in both hours and weeks and receive a lower weekly wage in the OLS specifications. Men with foreign names were also more likely to hold a public works job through the New Deal, which could be an indication of weak attachment to the labor force. In all cases, the relationship between name foreignness and economic outcomes is reduced when comparing brothers within households and nearly disappears in all cases except weekly wages. These results are robust to alternative linking algorithms that increasingly introduce more conservative requirements on what is considered a successful match.\(^9\)

Thus far, our analysis has included immigrants from all sending countries, combining high-skilled groups like the Brits that may have easily assimilated with lower-skilled groups like Italians and Russians that may have been targets of discrimination in the labor market. We present separate results in Figure 1 by country of origin for all sending countries with at least 1000 brother pairs.

The negative association between name foreignness and education (Panel \((a)\)) is statistically significant or nearly so in regressions with household fixed effects for immigrants

\(^8\) The 1940 Census only contains information on wage and salary income. As a result, results on annual earnings exclude the self-employed. Appendix Figure 2 and Appendix Table 4 show that shifts in the F-index have only a very small effect on the probability of self-employment.

from Italy, Russia, Ireland, and Austria, countries that by historical account may have suffered more discrimination and worse outcomes.

The negative association between name foreignness and other economic outcomes is not statistically significant for any country of origin in specifications with household fixed effects, even for immigrants from newer sending countries (panels (b)-(c)). However, because the confidence intervals are large when we subdivide the sample by country and include household fixed effects, we note that, in almost all cases, we cannot reject the hypothesis that the coefficients from OLS regressions and regressions with household fixed effects are the same. A similar set of countries exhibit a significant relationship between name foreignness and marrying a foreign-born spouse in OLS (Russia, Ireland, Austria, Italy, and also Finland), however associations between name foreignness and spouse attributes decline and become statistically insignificant when adding household fixed effects for all sending countries. These patterns generally hold when we use country-specific foreignness index measure instead (see panels (a)-(e) of Appendix Figure 3).

VI. Conclusion

In the early twentieth century, receiving an American-sounding name was associated with a series of positive outcomes for the children of immigrants, including more education, increased income, and a higher probability of employment. These results persist even after including fixed effects for country of origin, surname, and county of residence. We also find that the negative relationship between name foreignness and labor market outcomes is more prominent for men with fathers from newer sending countries. However, within households, brothers with more and less foreign names did not seem to attain very different levels of schooling, earnings, and employment. Our findings suggest that the negative association between ethnic names and outcomes may not
be indicative of a causal effect of having an ethnic-sounding first name, but may instead reflect the return to cultural assimilation, as the children of families that chose to assimilate had better outcomes than those that retained their cultural markers.

References


<table>
<thead>
<tr>
<th>OUTCOMES</th>
<th>Mean</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of education</td>
<td>10.16</td>
<td>-1.169***</td>
<td>-0.905***</td>
<td>-0.814***</td>
<td>-0.797***</td>
<td>-0.215***</td>
<td>-0.180***</td>
</tr>
<tr>
<td>N</td>
<td>802519</td>
<td>709727</td>
<td>709727</td>
<td>709723</td>
<td>709723</td>
<td>709727</td>
<td>113239</td>
</tr>
<tr>
<td>Employed</td>
<td>0.865</td>
<td>-0.0534***</td>
<td>-0.0260***</td>
<td>-0.0203***</td>
<td>-0.0159***</td>
<td>-0.00409</td>
<td>-0.0172*</td>
</tr>
<tr>
<td>N</td>
<td>820294</td>
<td>725300</td>
<td>725300</td>
<td>725295</td>
<td>725295</td>
<td>725300</td>
<td>115601</td>
</tr>
<tr>
<td>Log annual income</td>
<td>6.846</td>
<td>-0.111***</td>
<td>-0.0929***</td>
<td>-0.0769***</td>
<td>-0.0944***</td>
<td>-0.0234</td>
<td>-0.0302</td>
</tr>
<tr>
<td>N</td>
<td>629969</td>
<td>555406</td>
<td>555406</td>
<td>555403</td>
<td>555403</td>
<td>555406</td>
<td>87238</td>
</tr>
<tr>
<td>F-index of spouse’s name</td>
<td>0.450</td>
<td>0.122***</td>
<td>0.0571***</td>
<td>0.0479***</td>
<td>0.0399***</td>
<td>0.0118</td>
<td>0.0122</td>
</tr>
<tr>
<td>N</td>
<td>435802</td>
<td>381440</td>
<td>381440</td>
<td>381438</td>
<td>381438</td>
<td>381440</td>
<td>60076</td>
</tr>
<tr>
<td>Foreign-born spouse</td>
<td>0.052</td>
<td>0.0410***</td>
<td>0.0269***</td>
<td>0.0251***</td>
<td>0.0228***</td>
<td>0.00729</td>
<td>0.00228</td>
</tr>
<tr>
<td>N</td>
<td>459209</td>
<td>401810</td>
<td>401810</td>
<td>401808</td>
<td>401808</td>
<td>401810</td>
<td>63088</td>
</tr>
</tbody>
</table>

| Country of origin FE | No | Yes | Yes | Yes | No | No |
|Surname FE | No | No | Yes | Yes | No | No |
|1920 county of residence FE | No | No | No | Yes | No | No |
|Household FE | No | No | No | No | Yes | Yes |
|Brothers 2 or fewer years apart | No | No | No | No | No | Yes |

Note: *p < 0.1; **p < 0.05; ***p < 0.01; standard errors are shown in parentheses. Sample includes native-born men matched between 1920 and 1940 complete-count Censuses. Men must be white, aged 3-15 in 1920, born outside the South, and living at home with parents in 1920 in a household whose head was foreign-born. All regressions control for a vector of dummies for child’s age in 1920, parental years in the US, and child’s rank in the birth order.
Figure 1

(a) Implied effect of a 20 point shift in F–index on years of education, by country

(b) Implied effect of a 20 point shift in F–index on employment, by country

Note: In these figures we estimate regressions of each outcome listed on the foreignness index of an individual’s name for men from countries with more than 1000 observations at the household level. We report 0.2x the coefficient value of an individual’s F-index calculated for his year of birth. All regressions control for a vector of dummies for child’s age in 1920, parental years in the US, and child’s rank in the birth order. Sample includes native-born men matched between 1920 and 1940 complete-count Censuses. Men must be white, aged 3-15 in 1920, born outside the South, and living at home with parents in 1920 in a household whose head was born in one of the countries listed.
Note: In these figures we estimate regressions of each outcome listed on the foreignness index of an individual’s name for men from countries with more than 1000 observations at the household level. We report 0.2x the coefficient value of an individual’s F-index calculated for his year of birth. All regressions control for a vector of dummies for child’s age in 1920, parental years in the US, and child’s rank in the birth order. Sample includes native-born men matched between 1920 and 1940 complete-count Censuses. Men must be white, aged 3-15 in 1920, born outside the South, and living at home with parents in 1920 in a household whose head was born in one of the countries listed.