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

By Ran Abramitzky, Leah Boustan, Katherine Eriksson, and Stephanie Hao*

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 United States changed their own names and/or chose more American sounding names for their children as they adapted to the country (Biavaschi, Giulietti, and Siddique 2017; Abramitzky, Boustan, and 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 an economic return to cultural assimilation.

We start by constructing a large sample of the children of immigrants observed in their childhood households in the 1920 census linked to adult outcomes in the 1940 census. Parents of these children immigrated to the United States during the Age of Mass Migration from Europe (1850–1913), one of the largest migration episodes in American history. During this era, about 30 million immigrants from countries like Ireland, Germany, Italy, and Russia moved across the Atlantic (Abramitzky, Boustan, and Eriksson 2012, 2014, 2016 and Abramitzky and Boustan 2017).

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 themselves less culturally assimilated (a foreign spouse or a spouse with a more foreign name). 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. 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

*Abramitzky: Department of Economics, Stanford University, and NBER (email: ranabr@stanford.edu); Boustan: Department of Economics, Princeton University, and NBER (email: lboustan@princeton.edu); Eriksson: Department of Economics, University of California-Davis, and NBER (email: kaeriksson@ucdavis.edu); Hao: Department of Economics, Princeton University (email: slhao@princeton.edu).

those that did not. In particular, comparisons between brothers rule out discrimination in the labor market on the basis of first names, but this null result could still 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 and other demographic groups, which is often measured using differential treatment by name. The challenge in this literature (recognized by the authors) is that names or name changes could be correlated with other aspects of family background or individual motivation (Arai and Thoursie 2009; Moser 2012; Goldstein and Stecklov 2016; Carneiro, Lee, and Reis 2016; Biavaschi, Giulietti, and Siddique 2017; in another context, see Rubinstein and Brenner 2014). Our findings suggest that the negative association between ethnic names and outcomes at least partially reflects the return to cultural assimilation or household selection into assimilation rather than external discrimination. Consistent with our results, Oreopoulos (2011) does not find lower callback rates for workers with foreign-sounding first names in a resume audit study but does find lower rates for workers with foreign-sounding last names.[1] We complement Oreopoulos by looking at real labor market success as measured by wages and employment, considering a variety of other outcomes like education and matching in the marriage market, and studying an important period in US history during which ethnicity is said to have played a large role in employment (the classic "no Irish need apply" signs) and before antidiscrimination laws were passed.

---

[1] In a related literature, Bertrand and Mullainathan (2004) documents lower callback rates for resumes assigned a distinctively black name, whereas Fryer and Levitt (2004) instead finds 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) finds a mortality *advantage* for men who received distinctively black names in the nineteenth and early twentieth centuries and points to suggestive evidence that receiving a black name is correlated with cultural factors at the household level.

I. 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. This data has been made available to researchers by the Minnesota Population Center and *Ancestry.com*. To develop a systematic measure of name foreignness, we use the newly digitized complete-count 1920 and 1940 US censuses 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). The foreignness index, or *F*-index, is defined as

$$\textit{Foreignness Index}_{name}$$

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

and ranges from 0 to 100, with a value of 0 reflecting the fact that no men in the United States 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 20 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 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 in the 1920 census and adulthood in the 1940 census. Siblings are observed in the same household in 1920. 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).[2] 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 standard for historical matched samples (see Abramitzky et al. 2019).

Online 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. As is typical for matched samples, we find that men in our linked dataset have higher socioeconomic status than those 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 OLS association between name foreignness and economic outcomes is similar in the matched sample and the full population (compare online Appendix Table 2 to Table 1, which will be described in more detail hereafter).

II. Foreign Names and Outcomes in the 1940 Labor and Marriage Markets

Having a foreign name may 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 may perceive lower returns to education and exert less effort in school (Jensen 2010). Second, teachers or employers may 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.

We start by regressing adult outcomes 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, controlling for a vector of dummies for the son's birth year (γ_{ij}):

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

We control for parental years in the United States and the child's place in the birth order to absorb the association between name foreignness and other aspects of family structure or parental cultural assimilation (see Abramitzky, Boustan, and Eriksson forthcoming). We then augment this core specification with sets of fixed effects included in the vector X, starting with parental country of origin, then surname, and then childhood county of residence. Finally, we add household fixed effects (α_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 is negatively related to educational attainment, employment, and earnings. 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 ($= -1.169/5 = 0.23$ of a school year); 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

---

[2]We note that the presence of false positive links would bias the coefficients from 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.

TABLE 1— EFFECT OF GIVEN-NAME FOREIGNNESS ON LABOR AND MARRIAGE MARKET OUTCOMES

Outcomes	Mean	(1)	(2)	(3)	(4)	(5)	(6)
Years of education	10.16	−1.169	−0.905	−0.814	−0.797	−0.215	−0.180
		(0.0165)	(0.0175)	(0.0226)	(0.0222)	(0.0527)	(0.0488)
Observations	802,519	709,727	709,727	709,723	709,723	709,727	113,239
Employed	0.865	−0.0534	−0.0260	−0.0203	−0.0159	−0.00409	−0.0172
		(0.00190)	(0.00203)	(0.00268)	(0.00269)	(0.00798)	(0.00757)
Observations	820,294	725,300	725,300	725,295	7252,95	725,300	115,601
log annual income	6.846	−0.111	−0.0929	−0.0769	−0.0944	−0.0234	−0.0302
		(0.00495)	(0.00534)	(0.00742)	(0.00725)	(0.0230)	(0.0213)
Observations	629,969	555,406	555,406	555,403	555,403	555,406	87,238
F-index of spouse's name	0.450	0.122	0.0571	0.0479	0.0399	0.0118	0.0122
		(0.00145)	(0.00154)	(0.00226)	(0.00227)	(0.00954)	(0.00892)
Observations	435,802	381,440	381,440	381,438	381,438	381,440	60,076
Foreign-born spouse	0.0522	0.0410	0.0269	0.0251	0.0228	0.00729	0.00228
		(0.00167)	(0.00180)	(0.00262)	(0.00265)	(0.0102)	(0.00978)
Observations	459,209	401,810	401,810	401,808	401,808	401,810	63,088
Country of origin fixed effects		No	Yes	Yes	Yes	No	No
Surname fixed effects		No	No	Yes	Yes	No	No
1920 county of residence fixed effects		No	No	No	Yes	No	No
Household fixed effects		No	No	No	No	Yes	No
Brothers two or fewer years apart		No	No	No	No	Yes	Yes

Notes: Standard errors are shown in parentheses. Sample includes native-born men matched between the 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 United States, and child's rank in the birth order.

culturally distant from the United States, 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 2019).

We then compare brothers who were given names with different ethnic content. Table 1 (column 5) demonstrates that including household fixed effects substantially reduces the estimated effects of name foreignness on adult outcomes (estimates become statistically insignificant in most cases).[3] Foreign names had a small and

significant effect on educational attainment within brother pairs (20 points of F-index associated with 2 fewer weeks of schooling) and no effect on earnings and unemployment. The population-based estimates pick up other differences between families that chose foreign or American sounding names rather than the causal effect of foreign names themselves.[4]

The ethnic signal of names that parents select for their children at birth can be attenuated or augmented as the name becomes more/less

[3] In an earlier version of this paper, we found an association between name foreignness and adult outcomes, even within pairs of brothers (Abramitzky, Boustan, and Eriksson 2016). These results were driven by a few households with coresident children who were likely not related and could not be easily identified with census data provided by Family Search (e.g., two families living within a single household

or boarders living with a family). The updated results in this draft are instead based on census data from the Minnesota Population Center, 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 foreignness and adult outcomes within brother pairs.

[4] Estimates are little changed by weighting the distribution of father's country of origin in the linked sample to match the full population in 1920 or 1940; see online Appendix Figure 4.

popular among certain groups. We find similar results when using 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 (online Appendix Figure 1). More relevant to an employer's perception of a worker's ethnic identity might be the foreignness index of his name at the time of labor market entry. In Appendix Table 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), so no clear pattern emerges. Note that our matched sample contains disproportionate numbers of men with uncommon names. One possibility is that employers cannot discern the ethnic content of rare names. However, we find similar results if we split our sample by name frequency (see online Appendix Tables 4 and 5).

Beyond the labor market, having a foreign name may have influenced men's marriage market outcomes. 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 the spouse's first name, an indicator of being either born abroad or raised in a less culturally assimilated family in the United States.[5] 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 in the *F*-index of his spouse's first name (on a base of 45 points).[6] Yet, as with the labor market outcomes, neither relationship remains statistically significant

after adding household fixed effects (although confidence intervals are large).

Online 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.[7] 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.[8]

We present separate results in Figure 1 by country of origin for all sending countries with at least 1,000 brother pairs in order to differentiate between groups like the Brits who may have easily assimilated with new immigrant groups and like Italians and Russians who may have been targets of discrimination in the labor market. We do find one outcome—education—that seems to be associated with name foreignness even within pairs of brothers. The negative association between name foreignness and education (panel A) is statistically significant or nearly so in regressions with household fixed effects for immigrants 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 or marriage market

---

[5] Until 1930, the census asked all respondents about parents' birthplace, which would have allowed 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 5 percent of the population, and so we instead focus on whether a spouse is a first-generation immigrant herself.

[6] Results on spouse characteristics are restricted to the subsample of men who were 25 years or older in 1940 and married in that year. Men with a more foreign-sounding name are less likely to be married by 1940, but this effect is economically small. Twenty 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).

[7] The 1940 census only contains information on wage and salary income. As a result, results on annual earnings exclude the self-employed. Online Appendix Figure 2 and online Appendix Table 6 show that shifts in the *F*-index have only a very small effect on the probability of self-employment.

[8] See online Appendix Figures 5–6 and online Appendix Tables 7–10. See Abramitzky et al. (2019) for a detailed description of alternative matching approaches.
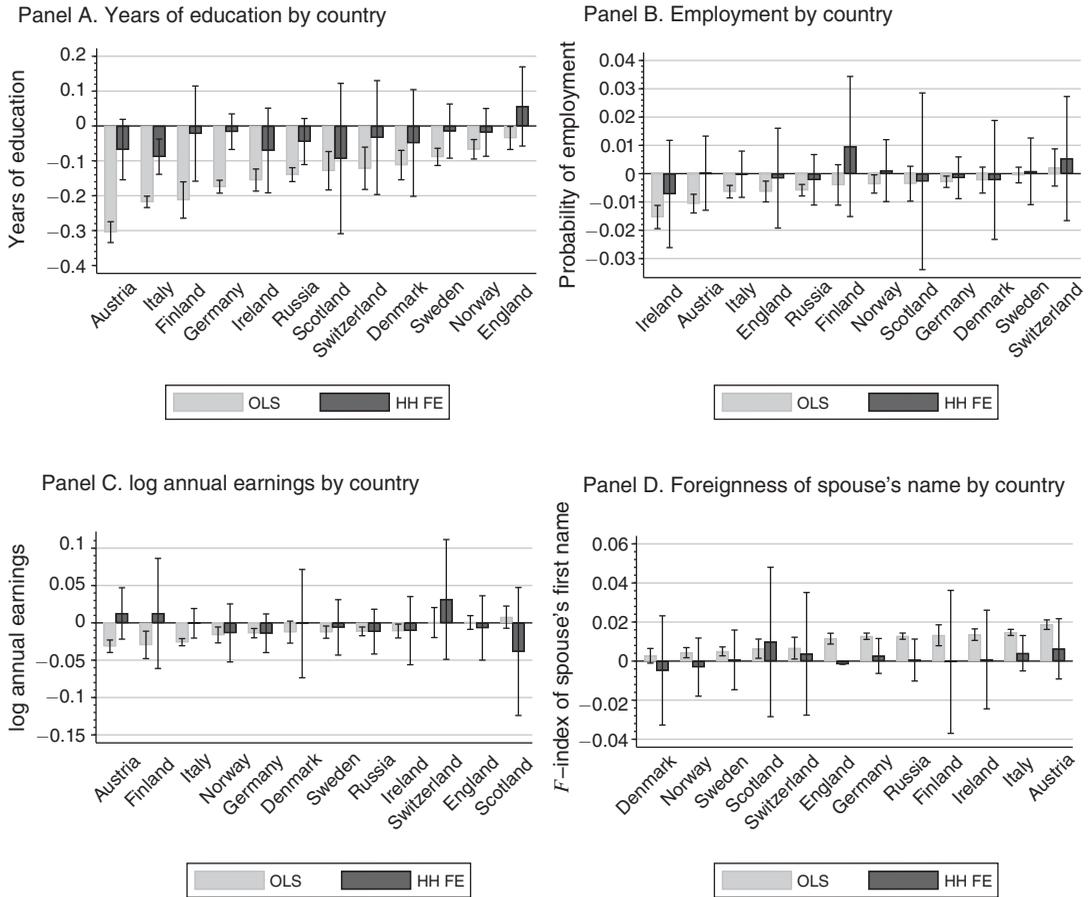
Panel A. Years of education by country

Panel B. Employment by country

Panel C. log annual earnings by country

Panel D. Foreignness of spouse's name by country

FIGURE 1. IMPLIED EFFECT OF A 20-POINT SHIFT IN F-INDEX

Notes: 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 1,000 observations at the household level. We report $0.2\times$ 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 United States, and child's rank in the birth order. Sample includes native-born men matched between the 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.

outcomes is not statistically significant 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. These patterns generally hold when we use country-specific foreignness index measures instead (see panels A–E of online Appendix Figure 3).

III. 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. 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

Abramitzky, Ran, and Leah Boustan. 2017. "Immigration in American Economic History." *Journal of Economic Literature* 55 (4): 1311–45.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 102 (5): 1832–56.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122 (3): 467–506.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2016. "Cultural Assimilation during the Age of Mass Migration." NBER Working Paper 22381.

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. Forthcoming. "Do Immigrants Assimilate More Slowly Today Than in the Past?" *American Economic Review: Insights.* https://doi.org/10.1257/aeri.20190079.

Abramitzky, Ran, Leah Platt Boustan, Katherine Eriksson, James J. Feigenbaum, and Santiago Pérez. 2019. "Automated Linking of Historical Data." NBER Working Paper 25825.

Arai, Mahmood, and Peter Skogman Thoursie. 2009. "Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings." *Journal of Labor Economics* 27 (1): 127–47.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Biavaschi, Costanza, Corrado Giulietti, and Zahra Siddique. 2017. "The Economic Payoff of Name Americanization." *Journal of Labor Economics* 35 (4): 1089–1116.

Carneiro, Pedro, Sokbae Lee, and Hugo Reis. 2016. "Please Call Me John: Name Choice and the Assimilation of Immigrants in the United States, 1900–30." Institute of Labor Economics Discussion Paper 9792.

Cook, Lisa D., Trevon D. Logan, and John M. Parman. 2016. "The Mortality Consequences of Distinctively Black Names." *Explorations in Economic History* 59: 114–25.

Eriksson, Katherine, and Zachary Ward. 2019. "The Residential Segregation of Immigrants in the United States from 1850 to 1940." *Journal of Economic History* 79 (4): 989–1026.

Figlio, David N. 2005. "Names, Expectations and the Black-White Test Score Gap." NBER Working Paper 11195.

Fryer, Roland G., Jr., and Steven D. Levitt. 2004. "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics* 119 (3): 767–805.

Goldstein, Joshua R., and Guy Stecklov. 2016. "From Patrick to John F.: Ethnic Names and Occupational Success in the Last Era of Mass Migration." *American Sociological Review* 81 (1): 85–106.

Jensen, Robert. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125 (2): 515–48.

Moser, Petra. 2012. "Taste-Based Discrimination Evidence from a Shift in Ethnic Preferences after WWI." *Explorations in Economic History* 49 (2): 167–88.

Oreopoulos, Philip. 2011. "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes." *American Economic Journal: Economic Policy* 3 (4): 148–71.

Rubinstein, Yona, and Dror Brenner. 2014. "Pride and Prejudice: Using Ethnic-Sounding Names and Inter-Ethnic Marriages to Identify Labour Market Discrimination." *Review of Economic Studies* 81 (1): 389–425.