From Patrick to John F.: First-name choice as a measure of assimilation and predictor of occupational achievement in historical U.S. censuses

Extended abstract for 2013 PAA proposed paper

Joshua R. Goldstein (Max Planck Institute for Demographic Research)
Guy Stecklov (Hebrew University)

September 21, 2012

Abstract

The ethnic distinctiveness of first-names of the children of immigrants can measure assimilation and predict occupational attainment in early the United States in the early 20th century.

Some names like "Patrick" or "Antonio" are ethnically distinctive from the Anglo majority. Others like "John" or "Charles" are less so. In this research, we develop a measure of first-name ethnic distinctiveness using the relative frequencies of father's birthplace for each name. We then show (as with recent research on the "Blackness" of names) that ethnically distinct names are correlated with lower socioeconomic achievement for some groups but not others. We plan to use multivariate models to sort out various mechanisms of selection (e.g., ethnically distinctive names due to lower assimilation of the parental generation) and causation (the signalling effect of an ethnically distinctive name).

Our results inform the debate on the advantages and disadvantages of cultural assimilation of the second generation for socio-economic achievement.

1 Background and Motivation

The advantages and disadvantages of assimilation for contemporary U.S. immigrant populations is hotly debated (Portes and Zhou 1993; Xie and Greenman 2011, Perlmann and Waldinger 1997). In this proposed paper, we take a historical look at this question, taking a new look at the contours of assimilation of European immigrants to America at the turn of the last century and the extent to which cultural assimilation was associated with and even caused improvements in socio-economic status.

Our new measure of assimilation is based on the ethnic distinctiveness of first names – the degree of over-and-under representation of a name according to birthplace of parents. This measure allows us to do investigate comparatively the degree of cultural assimilation of different European immigrant groups as well as the association of cultural assimilation with socio-economic achievement.

We think of names both as a consequence and potential cause (or signal) of assimilation. Giving a child an ethnically-distinct or an Americanized name was a choice made by immigrant parents. This choice can be used to measure the cultural context and degree of assimilation of the parental generation. In this sense, first-names give us information about the degree of assimilation of the parental generation doing the naming.

First names may also have a causal effect on socio-economic outcomes (Fryer and Levitt 2004), acting as a *signal* (intentional or not) to others about the degree of "American-ness," or ethnic distinctiveness, a signal that could itself have consequences for socio-economic achievement.

Finally, the name reported by adults is not only the consequence of the name-choice made by parents but also includes the subjective choice of how to report the name as an adult. Thus, both an Anthony and an Antonio can become "Tony", a Michael can become a "Mike", or a Vincenzo could become a Vincent or a Vinny.

In each of these pathways, the ethnic distinctiveness or lack thereof can be thought of as a measure of cultural assimilation. Each of these mechanisms suggests in its own way that first-name choice may be linked to socio-economic status.

The goals of the current research are :

1) To develop measures of ethnic distinctiveness of first names that can be directly derived from the statistical patterns with micro-census data.

2) To measure the association of ethnic distinctiveness with socio-economic

attainment.

3) To use multivariate analysis to try to begin to separate the different causal pathways that drive the association between occupational achievement and ethnic distinctiveness of first names

Our research builds on earlier work by Watkins and London (1994), who studied name choice among Jews and Italians in the U.S. census of 1910 (but did not look at occupational achievement). The public availability of large IPUMS USA census samples (and modern computing power) makes it possible to work with large numbers of first (and potentially *last*) names, using "string" variables in a quantitative way that past researchers were only able to do easily with numeric measures. First names have long been an object of sociological investigations, notably by Alice Rossi (1965) and Stanley Lieberson (2000). The study of first name choice for contemporary occupational achievement is a hotbed of research activity, including work by Fryer and Levitt (2004) on the effects of having a "Black"-sounding name and innovative studies such as that done by Mahmood and Thoursie (2009) on the effects of name changes of immigrants on earnings and employment in contemporary Sweden. Recent work by Gerhards and Hans (2009) has introduced the study of first-naming practices as a measure of assimilation of Turkish immigrants to Germany.

2 Study Design

2.1 Data

We use the publicly available census micro-samples (IPUMS) available for the US censuses of 1880-1930, all of which include both first and last names as well as occupational coding. The status and average earnings of occupations are also coded, enabling summary measures about the relative occupational achievement of different individuals and sub-populations. The IPUMS samples also include individual measures of own birth place, mother's birthplace, and father's birthplace.

In the preliminary analysis we focus only on the census of 1930, for which there is a 5 percent sample. For ease of analysis, we restrict ourselves to males aged 35 to 44. We plan to expand the analysis to include the census of 1920, as well as additional age groups.

2.2 Methods

We measure ethnic distinctiveness in the following manner. Taking each country of origin separately, we produce a list of the most popular names. For example, in the preliminary results below, we show the top 30 names for men with fathers born in Ireland, Poland, Russia, and Italy. The ethnic distinctiveness of each name relative to a country of origin is then estimated by calculating the log-odds that a person with that name had a father born in the country of interest. Thus, for example, the frequency of the name "Patrick" in the 1930 census among men aged 35-44 was 5 per 10,000 for those whose father was *not* born in Ireland and 455 per 10,000 for those whose father was born in Ireland, an increase of some 90-fold. The log-odds, which we use as scores, are 4.5-times higher. We estimate the log-odds for each of the top names for each immigrant origin and use these as our scores of "Irish-ness", "Italian-ness", and so forth.

In order to measure occupational achievement, we then consider the average occupational earnings by name within each immigrant group, again defined by father's birth place. We thus see for example that men named Patrick with Irish-born fathers have an occupational earnings of 26.1 on a scale from 0 to 80. Comparable men named "John" with Irish-born fathers, however, have an occupational earnings score of 27.4. The magnitude of the "Patrick" penalty can be be put into perspective by noting that the average occupational earnings for all of the men with Irish-born fathers was about 2.5 points (compared to men with fathers born in the US and elsewhere besides Ireland). Thus the "Patrick" penalty among the Irish is equal to about 50 percent of the effect of being Irish vs. non-Irish.

The preliminary results section below gives bi-variate results of these analyses by ethnic group. For the PAA, we plan to include multivariate analysis of each name, including background variables that will help us to distinguish the extent to which first-name distinctiveness is associated with different background variables for the families of origin of the adults we observe in 1920 and 1930. For example, we can look at *children* named "Patrick" in 1880 and measure the characteristics of the parents and household: time since immigration, occupation, and perhaps even first-name of father. These can then be used as control variables in order to see if the characteristics of the families of origin can explain the later life differentials we see in 1930. Such results will help separate selection effects from any additional causal effect carrying a name might have.

3 Preliminary Results

In the attached figures we show the results of our preliminary investigations using the IPUMS 5 percent sample for 1930 for men aged 35-44 from Ireland, Italy, Poland, and Russia. For graphical clarity we restrict ourselves to the top-30 names for each group. However, we find similar quantitative results when we include the top 100 names.

The x-axis of each figure shows the log-odds of the name being held by someone with a father from the country under consideration. The y-axis shows the regression coefficient on occupational earnings of holding that particular name (the omitted category includes all names other than those shown in the plots). The lines shown in the figures are regression lines, weighted by the frequency of names. The size of text used for the names are proportional to their frequencies.

Our preliminary results provide evidence in support of the proposition that assimilation was indeed a path to economic success for the Italians and Irish. However, we also see signs that for some groups – notably the Jews of Russian and Polish origin – that having a very distinctive name (e.g., "Abraham") does not appear to have hindered economic success, at least for those who retained the name. Another sign in support of the segmented-assimilation framework in which some contexts of assimilation lead to downward mobility is that those who take on Americanized nicknames (e.g., "Pat" instead of "Patrick", "Mike" instead of "Michael") have the lowest occupational achievement.

Our preliminary results are fairly conclusive in showing that the distinctiveness of names can be measured. Furthermore they are suggestive of different paths of assimilation among different populations. It appears, at this early stage of our work, that the generalizations made by many scholars about assimilation for European immigrants receive support among the Italian and Irish immigrant populations, but that there are other populations in which distinctive-naming is not such a clear disadvantage. We work in the coming months will aim to verify these claims and sort out the mechanisms at work.

4 Research Plans

As discussed above, our research plans for this project include at least the following three extensions:

1. Expanding the samples to include a larger age-range, the census of 1920, and perhaps additional countries-of-origin.

2. Multivariate analysis to try to separate the influences of cultural assimilation of the parental generation giving the name from the life-course experience of the individuals carrying the name. These would include childhood conditions by name, such as father's occupation, geography, and even the ethnic distinctiveness of father's name.

3. Including last-names in the analysis (both in terms of ethnic distinctiveness and occupational achievement) as a way of enriching the information we have and to some extent helping us to understand the potential "signalling" information content of a first-name. (For the socio-economic effects of ethnically distinct first names might be larger for people with Anglo-sounding last names than for people whose last name already connotes ethnicity: Patrick Brown and John Brown might differ more than Patrick O'Connor and John O'Connor.)

References

Arai, M., & Thoursie, P. S. (2009). Renouncing personal names: An empirical examination of surname change and earnings. *Journal of Labor Economics*, *27*, 127-147.

Gerhards, J., & Hans, S. (2009). From Hasan to Herbert: Name-giving patterns of immigrant parents between acculturation and ethnic maintenance. *American Journal of Sociology*, *114*, 1102-1128.

Lieberson, S. (2000). *A matter of taste: How names, fashions, and culture change.* Yale University Press.

Perlmann, J., & Waldinger, R. (1997). Second generation decline? Children of immigrants, past and present-a reconsideration. *International Migration Review*, *31*(4), pp. 893-922.

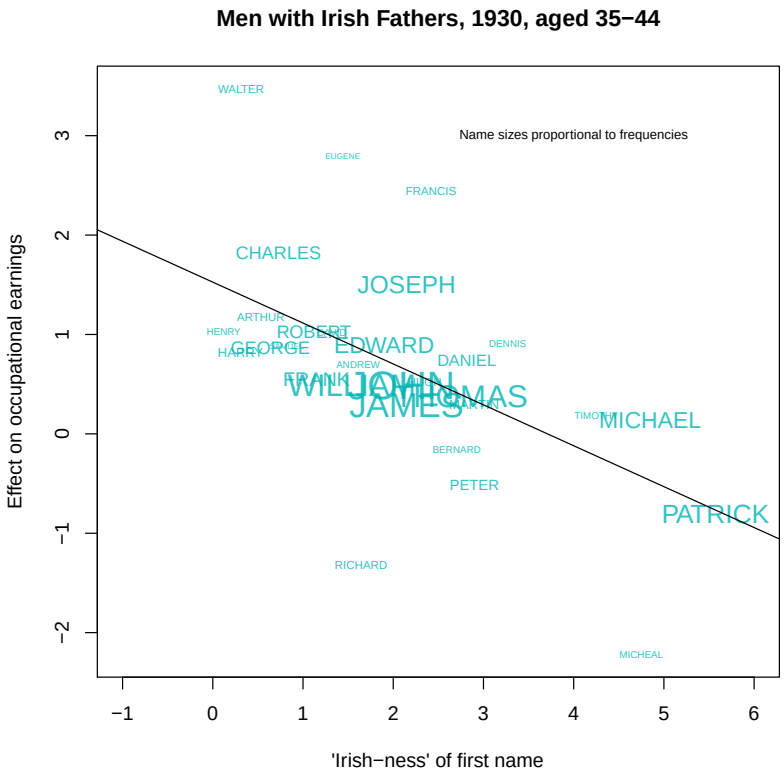
Roland G. Fryer, J., & Levitt, S. D. (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, *119*(3), pp. 767-805.

Rossi, A. S. (1965). Naming children in middle-class families. *American Sociological Review*, *30*(4), pp. 499-513.

Watkins, S. C., & London, A. S. (1994). Personal names and cultural change: A study of the naming patterns of italians and jews in the united states in 1910. *Social Science History*, *18*(2), pp. 169-209.

Xie, Y., & Greenman, E. (2011). The social context of assimilation: Testing implications of segmented assimilation theory. *Social Science Research*, *40*, 965-2984.

Figure 1: Irish and Italian 2nd Generation



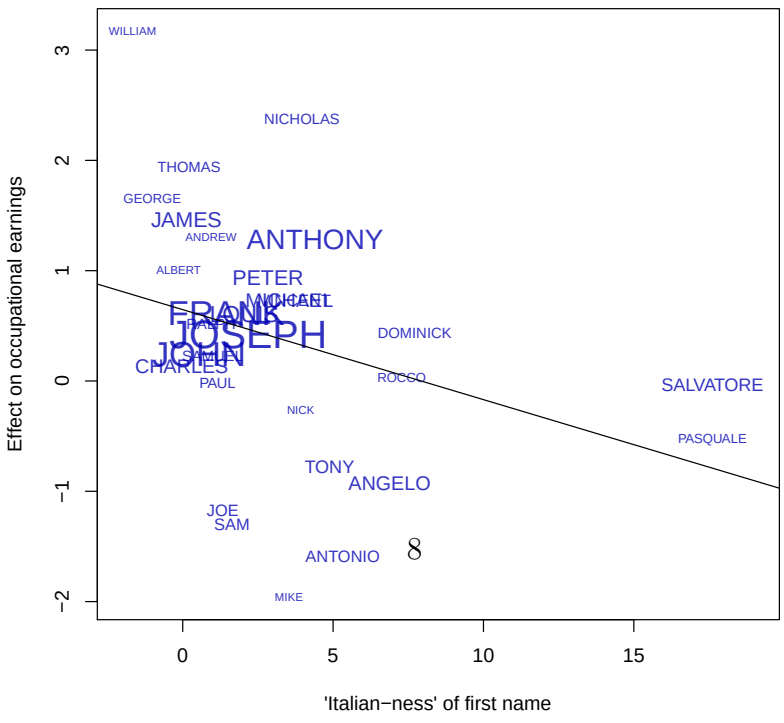


Figure 2: Polish and Russian 2nd Generation