ISSN 2042-4116

ESRC Centre for Population Change • Working Paper 43 • January 2014

CPC

centre for population change

The economic payoff of name Americanization

Costanza Biavaschi
Corrado Giulietti
Zahra Siddique

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

ABSTRACT

We examine the impact of the Americanization of names on the labor market outcomes of migrants. We construct a novel longitudinal data set of naturalization records in which we track a complete sample of migrants who naturalize by 1930. We find that migrants who Americanized their names experienced larger occupational upgrading. Some, such as those who changed to very popular American names like John or William, obtained gains in occupation-based earnings of at least 14%. We show that these estimates are causal effects by using an index of linguistic complexity based on Scrabble points as an instrumental variable that predicts name Americanization. We conclude that the tradeoff between individual identity and labor market success was present since the early making of modern America

JEL CLASSIFICATION

J61, J62, Z1, N32

KEYWORDS

Americanization, names, assimilation, migration

EDITORIAL NOTE

Dr Costanza Biavaschi is a Research Associate and Deputy Program Director for the Migration Area at the Institute for the Study of Labor, Germany. Costanza's research interests are in labor economics and applied econometrics, with a focus on return migration, the dynamics of migration choices, and the selection of migrants.

Dr Corrado Giulietti is Director of Research at the Institute for the Study of Labor, Germany. Corrado is also a Research Associate of the ESRC Centre for Population Change. His research interests are labor economics and applied econometrics, with a focus on the determinants of migration, the labor market and welfare effects of migration, the assimilation of immigrants, and the estimation of migration flows.

Dr Zahra Siddique is a Lecturer in economics at the University of Reading, UK. Zahra's research interests are in micro-econometrics, labor economics and development economics.

ACKNOWLEDGEMENTS

The authors would like to thank participants at the 10th IZA Annual Migration Meeting, the 4th TEMPO conference on international migration, the Essex Fresh meeting as well as seminar participants at IZA, the University of Reading, Tsinghua University and the University of Southampton for their helpful comments. This draft has greatly benefitted from comments by Benjamin Elsner, Lidia Farré, Jesús Fernández-Huertas Moraga, Daniel Hamermesh, Carolyn Moehling, Andrew Oswald, Núria Rodríguez-Planas, Todd Sorensen, Derek Stemple, Christopher Taber, and Konstantinos Tatsiramos. Lena Knops and Eleni Kostopoulou provided excellent research assistance. We are also grateful to John R. Logan who shared the crosswalk between enumeration districts and Census tracts with us.

The ESRC Centre for Population Change Working Paper Series is edited by Teresa McGowan.

THE ECONOMIC PAYOFF OF NAME AMERICANIZATION

TABLE OF CONTENTS

1. INTRODUCTION

And while you bring all countries with you, you come with a purpose of leaving all other countries behind you - bringing what is best of their spirit, but not looking over your shoulders and seeking to perpetuate what you intended to leave in them. [...] You cannot dedicate yourself to America unless you become in every respect and with every purpose of your will thorough Americans.

Woodrow Wilson, May 10, 1915[1]

Americanization, the process by which immigrants strive to assimilate into American society, encompasses several dimensions. One such element is the Americanization of migrants' first names, a key aspect of the desire – or need – to conform to the American norm. The importance of first names has long been stressed by sociologists (Lieberson, 2000), and it could serve as a crucial marker to understanding the returns and tradeoffs from Americanization. In this paper, we provide the first evidence on the economic consequences of the Americanization of first names during a pivotal period in American history.[2]

Most Americans have heard stories of migrant ancestors Americanizing their names in the early half of the twentieth century. However, such anecdotes are typically stored solely in familial memories, with no study having measured the extent and implications of name Americanization. Table 1 provides preliminary facts about the magnitude of this phenomenon. Defined throughout as the custom of adopting a first name that was more popular in the U.S. born population than the original migrant's name, name Americanization was a widespread practice. Almost a third of naturalizing immigrants abandoned their first names by 1930 and acquired popular American names such as William, John or Charles. These popular destination names are shown in Panel A. Panel B shows substantial variation in name Americanization by country of birth, highlighting that migrants from Italy, Russia and Germany were all very likely to abandon their "foreign-sounding names" and adopt names popular among the U.S. born population.

[1] Speech given to a group of naturalized Americans in Philadelphia.

[2] Surname Americanization – as we define it in the next section – was in place but appears to be less common, with only 7% of migrants Americanizing their surnames. Details on surnames changes are provided in the online Appendix. The rest of the paper focuses on first names.

A) Popular American Names

Name	Frequency	Cumulative
William	0.090	0.090
John	0.089	0.179
Charles	0.065	0.244
George	0.061	0.306
...	...	...
Otto	0.001	0.791
...	...	...
Angelo	0.000	0.932

B) Name Americanization

Country of origin	% Americanized	N
Italy	0.189	729
Russian Empire	0.567	598
Other new migrants	0.467	683
Germany	0.258	365
Ireland	0.038	316
Other old migrants	0.130	385
Other	0.235	277
Total	0.303	3353

Table 1: Popular American Names and Name Americanization.

Notes: Panel A): Own tabulations from IPUMS Census, 1930. Name frequencies are calculated using US-born population aged 50+ in 1930 New York.
Panel B): Own tabulations from sample of naturalizing immigrants in 1930, Ancestry.com. Name Americanization is defined as the custom of adopting a first name that was more popular in the U.S. born population than the original migrant's name. For details see data section.

Widespread name Americanization prompts the question of whether it had an impact on migrants' economic success. Figure 1 provides a preliminary answer to this question. Name Americanization into the most popular names - e.g. the top quartile - was associated with an occupation-based earnings increase of above 10%. These gains were larger than those experienced by migrants Americanizing into less popular names - e.g. the first quartile - and even more so than those experienced by migrants who kept their original name or changed to a more distinctive name.

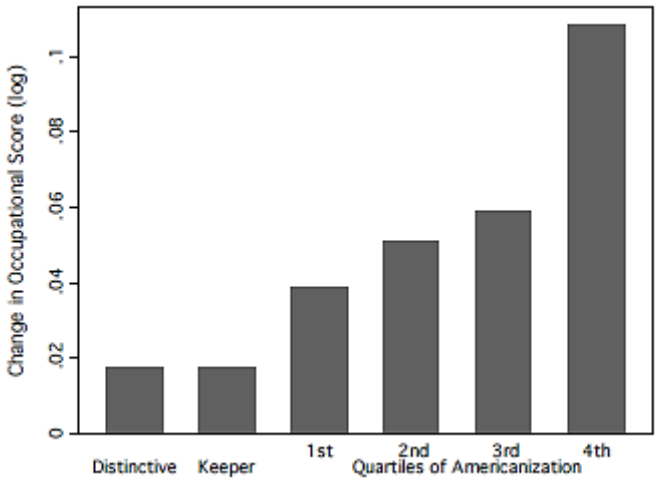


Figure 1: Change in Log-Occupational Score and Americanization

Notes: The x-axis represents quartiles of the change in the Americanization. The category "Distinctive" indicates individuals who changed names into less frequent names. The category "Keepers" indicates individuals who did not change name or change into equally frequent names. Name Americanization is defined as the custom of adopting a first name that was more popular in the U.S. born population than the original migrant's name. The dependent variable is the change in log-occupational score.

We devote much of our analysis to verifying that the positive impact of name Americanization on occupation-based earnings represents a credible causal effect and is not driven by observable and unobservable differences across migrants. We account for characteristics that are often unavailable, even in modern data sets, allowing us to control for time varying socio-demographic traits as well as nationality-specific and local labor market-specific time trends. More importantly, we examine the causal effect of changing names by exploiting the longitudinal nature of our data and using an instrumental variable approach. We instrument name Americanization with an index based on Scrabble points, which captures the degree of linguistic complexity of names upon arrival compared to the linguistic complexity of names at destination.

To understand the link between name Americanization and labor market outcomes we have digitalized a novel data set in which we can observe an entire random sample of migrants who completed their naturalization papers by 1930 in New York City. We are able to follow the full set of individuals over time, due to the nature of the naturalization process and documentation procedure, which required migrants to first file a declaration and later a petition

for naturalization. By exploiting this two-step procedure we obtain information on name Americanization and a wide range of migrant characteristics at different points in time, allowing us to examine economic outcomes following name Americanization. This strategy differs from the common methodology used in constructing historical panel data, which, resorting to record linkage based on name and age, only delivers partial matching and does not allow the detection of name changes.

Throughout our analysis and across numerous specifications, we find a substantial pay- off of name Americanization. Further tests suggest that name Americanization was more common among migrants that were likely low skilled, more discriminated against, or with less alternative means for socio-economic improvement. Therefore, our study adds to recent evidence provided in Abramitzky et al. (2012a), which concludes that migrant occupational upgrading was rather limited for low skill migrants between 1900 and 1920, by interpreting name Americanization as a way of circumventing negative occupational shocks and climbing up the occupational ladder.

More generally, our results highlight the tradeoff between maintaining one's individual identity and labor market success, suggesting that the process of cultural assimilation at the dawn of the modern "melting pot" was instrumental for migrants' economic advancement. Therefore, such a tradeoff is not only present in recent times (e.g. Bertrand and Mullainathan, 2004, Fryer and Levitt, 2004, Arai and Thoursie, 2009, Algan et al., 2012) but was also in place during the early 1900s.

We are therefore able to add evidence to a growing literature that links names to economic success during a period that was pivotal in laying the foundations of modern America. Much of the literature on the relationship between names and outcomes focuses on the second half of the twentieth century. For example, psychologists have shown that first names closer to those in host societies are associated with positive attitudes among host populations in general (Kang, 1971, Drury and McCarthy, 1980) and more specifically among employers, co-workers and customers (Laham et al., 2011). Economists have shown that more common names result in better educational (Figlio, 2005) and labor market success, due to reduced

discrimination (e.g. Bertrand and Mullainathan, 2004).[3] The association between names and economic outcomes, however, might also stem from unobserved factors that correlate with parents naming choices and economic success. For instance, Fryer and Levitt (2004) find that after controlling for background characteristics white names are no longer associated with better outcomes. On the other hand, recent work by Algan et al. (2012) shows that parents in France are willing to give up one year of expected average income by giving their children an Arabic name; Brenner and Rubinstein (2013) find evidence that the Israeli labor market discriminates based on ethnic sounding surnames.

To our knowledge, only two studies have looked at name changes, both focusing on renouncing surnames. The first analysis relates names to gender identity, focusing on women who decide to keep their maiden name instead of acquiring their husband's surname upon marriage (Goldin and Shim, 2004).[4] While this work analyzes surname choice rather than its effect on outcomes, it provides compelling evidence on reversion to the custom of taking the husband's name during the second half of the twentieth century. More importantly, it documents the existence of a negative correlation between surname changes and skills. The second analysis looks at the effect of surname changes made by immigrants from Asian and Slavic countries living in Sweden in the 1990s (Arai and Thoursie, 2009). In this context, in which 0.4% of migrants changed surnames, fixed-effects estimates show a substantial increase in earnings. The authors conclude that this result is likely driven by reduced labor market discrimination associated with having a Swedish-sounding surname. The scant literature on this topic is the result of a scarcity of data sets containing sufficient information to empirically test whether changing names improves economic outcomes. Furthermore, in a historical setting like ours, nailing the consequences of name Americanization is particularly challenging due to the non-availability of longitudinal data sources together with difficulties in tackling endogenous name choices. Our data and estimation techniques overcome such limitations, contributing to a better understanding of the pressures migrants faced to conform to American norms.

[3] For instance, racial or ethnic names are frequently used in audit and correspondence studies to test for racial discrimination in hiring. Bertrand and Mullainathan (2004) use a correspondence study to provide evidence that résumé's with African-American names had significantly lower callback for recruitment by hiring firms than did resumes with white names. Similar studies in Europe and developing countries use ethnic first names to find evidence of labor market discrimination against immigrants, ethnic minorities and/or socio-economically disadvantaged groups.

[4] To the best of our knowledge, the only precedent on a similar topic is the sociological study by Broom et al. (1955). The author looks at the characteristics of 1,107 petitions for name changes filed at the Los Angeles Superior Court, in 1946-1947. However, this study is very descriptive in nature, is unclear on how the petitions have been selected, does not focus on migrants and does not look at labor market outcomes.

The paper is organized as follows. We begin by providing a description of the data set we have collected and the empirical strategy adopted. We then present and discuss the estimated effects of name Americanization before providing robustness checks and concluding.

The online Appendix referred to in this article can be accessed at the following link: https://www.dropbox.com/s/mcfkzh5riur63mx/appendix.pdf

2. DATA

In order to understand its magnitude and consequences, it is necessary to define and measure appropriately the term 'name Americanization'. An ideal data set would consist of individuals who are observed before and after Americanizing their names. Historical data sources such as the censuses do not include identifiers that allow tracking of individuals over time. Hence, much of the existing literature (e.g. Ferrie, 1996, Abramitzky et al., 2012a) resorts to linking individual records, often matching individuals by name, age and birthplace. This strategy is clearly not an option for us, as we would be unable to match individuals who have changed names. In fact, name changes are one of the challenges in successfully link individuals – especially immigrants – over time.

Instead, we exploit the rich records stemming from the naturalization procedure.[5] With only marginal changes since 1795, the naturalization procedure consisted of two steps. Free white aliens residing in the U.S. for at least two years (one year of which was spent in the state or territory in which the application was made) were required to first file a declaration of intention. The second step involved filing a petition for admission to citizenship which could be undertaken a minimum of five years following the initial declaration. At the moment of filing citizenship papers, the Bureau of Immigration and Naturalization checked ship manifests and issued a certificate of arrival, a document which included the name held at arrival. Finally, migrants took a naturalization oath or oath of allegiance. After having completed all citizenship requirements, migrants were issued a certificate of naturalization. Following 1906, the declaration of intention, the petition for naturalization and the certificate of naturalization all

[5] Naturalization was considered to be the ultimate act of identifying with the American culture. In fact, before the 1924 restrictions - but also after the imposition of quotas - direct benefits for citizenship were rather limited per se. Immigrants during this time were permanent residents and, before the New Deal, social benefits were too small to reflect a motivation for naturalization (Lleras-Muney and Shertzer, 2012).

had a standardized format prescribed by the Bureau of Immigration and Naturalization, forming the rich records that we have digitalized as part of our research.

To compile our data set we extracted naturalization records from Ancestry.com, a website providing genealogical and family history records. We accessed the complete naturalization records filed at the U.S. District Court for the Southern and Eastern districts of New York City in 1930.[6] More than 30% of all certificates granted in the United States pertain to the district of New York (Annual Report of the Commissioner of Naturalization, 1930, p.15), highlighting the importance of these records in understanding immigrant assimilation in the U.S. Ancestry.com provides access to 26,113 of the official 30,361 petitions that were filed in the naturalization district of New York (Annual Report of the Commissioner of Naturalization, 1930, p.15), corresponding to more than 85% of all the records thought to be available for that year. The remaining records are likely to have been granted by other district courts of New York City or within the State, yet are not available in electronic form. Nonetheless, the overwhelming share of naturalizations granted by the Eastern and Southern courts highlights the importance of the two courts, making their records representative of nearly the entire population of naturalizing immigrants residing in the state of New York. While our sample of study represents only a part of all immigrants living in the U.S. in 1930, this subset is of particular relevance. First, it captures immigrants who arrived on average in the late 1910s, during the last surge of migration before U.S. doors were shut. Second, while return migration was common in the early 1900s (Bandiera et al., 2013), these migrants settled permanently, truly contributing to 'the making of modern America'. Third, the cultural and economic tradeoffs faced by naturalized migrants in New York City should be less pronounced than those of non-naturalized migrants in the city or of migrants in other states less exposed to the new migration waves, likely providing a lower bound on the magnitude and returns to name Americanization.[7]

We selected and manually transcribed a 20% random sample of the available records for

[6] The Southern District Court jurisdiction comprises the counties of New York (Manhattan), Bronx, Westchester, Putnam, Rockland, Orange, Dutchess, and Sullivan. The Eastern District court's territorial jurisdiction includes the counties of Kings (Brooklyn), Queens, Richmond (Staten Island), Nassau and Suffolk.

[7] We compared the characteristics of our sample with random samples of migrants (naturalized and non-naturalized) from the 1930 IPUMS. See the online Appendix for details.

1930. While our randomization procedure involved the collection of naturalization records for both males and females, in the paper we use the records for male immigrants only[8]. We also restrict the sample to individuals who filed petition papers two to seven years after declaration, which gives us an estimation sample of 3,353 migrants.[9]

Given the two steps involved in the naturalization process, we can observe migrants' characteristics at two different points in time without resorting to record linkage. Besides being obvious by simple observation of the individual characteristics, we are confident that the same migrant is being observed for two reasons: first, except in rare cases, documents are ordered by petition number and sequenced by document type for each migrant; and second, the petition number is usually printed on all documents.[10] In this paper, we refer to these two pieces of information as "declaration" and "petition", respectively. In order to test our main hypothesis – whether name Americanization influenced migrants' economic outcomes – we derived variables for both time periods. In the following paragraphs, we describe the steps taken to derive our key variables, which include popular American names, occupational scores and geographical characteristics.

American norm and Americanization. We measure conformability with American norms by exploring how immigrants' names compare with American ones. The first step is to rank American names by the frequency of their appearance in the U.S.-born population compiled from the 5% Integrated Public Use Microdata Serie of the 1930 Census (IPUMS, Steven Ruggles et al., 2010). We focus on American individuals born before 1880 and living in the state of New York at the time of the census. The reason behind restricting our attention to individuals aged 50+ is to capture naming patterns for a period that encompasses earlier migration waves to the U.S. only, hence abstracting from 1930s fashions and naming patterns that might be influenced by the contemporary migration waves of that

[8] The reasons for doing this are twofold. First, females could naturalize through marriage, and not file papers. Therefore, the female sample is a somewhat selected group likely to be systematically different from the population of all female migrants. Second, after the Act of 1922, females married to a U.S. citizen were not required to file first papers, hence we cannot observe them over time.

[9] Only 15 individuals are observed outside the interval 2-7 years since declaration. Note that immigrants only appear in our sample if their naturalization procedure has been completed - but independent of whether or not citizenship was granted. For about 2.6% of the sample citizenship was refused. We keep these records in the sample.

[10] The only exception is when we cannot find a whole set of documents for a particular petition number. This has occurred only in 0.2 per cent of our cases.

time. We also focus on naming patterns in the state of New York only to narrow down the pool of names that migrants in the sample were likely exposed to. Panel A of Table 1 shows common American names. Names including William, John, Charles and George were most common in the 50+ American-born population, which we consider representative of the American norm.

Based on our definition of the American norm, we derive a simple metric to capture name Americanization. For a name held by an individual i and observed at time t we define our metric as a normalized frequency of this name in the 50+ American-born population living in New York state:

$$A_{it} = \frac{S_{itk}}{max\,(S_{it1},\ldots,S_{itK})} \qquad (1)$$

$$\text{where} \quad S_{itk} = \sum_k \mathrm{I}(\mathrm{Name}_{it} = \mathrm{Name}_k) \qquad \text{for } k \in \{1, \ldots, K\} \text{ US-born 50+, NY}$$

where A_{it} is our metric, $\mathrm{I}(\mathrm{Name}_{it} = \mathrm{Name}_k)$ is an indicator variable that takes the value one if the name of a native-born individual is the same as the name of individual i; thus, S_{itk} counts the number of native-born individuals holding the same name as individual i. The denominator represents the maximum frequency across all names held by U.S.-born individuals, bounding our metric between zero and one.

In other words, A_{it} measures how frequent individual i's name at time t is in the American-born population on a scale from 0 to 1. In fact, names that are unique to our migrants are observed in the American-born population with a frequency equal to zero, so the metric associated with these names will also be zero. On the other hand, migrants called William and George will have a metric of, respectively, 1 and 0.68 (given by 0.0612/0.0899, see Table 1). What is name Americanization? This occurs whenever a migrant changes his name into one that more frequently occurs in the U.S. population, corresponding to an increase in A_{it} over time. As an example of name Americanization, a migrant called Giovanni who changes his name into William would be Americanizing his name, with an initial value of A_{it} equal to zero and a subsequent value of A_{it} equal to one. On the other hand, A_{it} would take a value of one for any t for a British migrant called William who does not change his name. Similarly, if Giovanni were to change his name into Salvatore, A_{it} will take a value zero

for any *t*, given that neither of these names were common American names. In both of these cases there is no change in A_{it} over time. Finally, while Americanizing one's name corresponds to an increase in A_{it}, A_{it} could also decrease over time if migrants were to change into more distinctive foreign names.

This simple metric allows us to define "name Americanizers" as individuals who choose different levels of A_{it}. The purpose of our index is to capture the distribution of names that is not affected by migration. It differs from other name indices, such as Fryer and Levitt (2004) in one fundamental aspect: while Fryer and Levitt (2004) are interested in a relative index that is invariant to name popularity across minority groups, we aim to exclusively measure the popularity of American names only.[11] Moreover, as shown in Section 6, the overall findings are robust to re-defining popularity using different times and including migrants in the sample of names. This is not surprising given that naming patterns were rather stable during the late 1800s and early 1900s and that the distribution of names was not substantially affected by immigration.

Occupations and earnings. One challenge in studying historical labor market outcomes is the lack of earning measures prior to the 1940 Census. We rely on an indirect, well-established measure of "earning potential" that assigns income scores to each occupation. To obtain this "occupational score", we first collected the occupation string from the naturalization papers and standardized occupation titles to match those identified in IPUMS. While all occupations were standardized during this process, we flagged those occupations for which some imputation was made for assignment of an occupation. Next, we attributed the "occupational score" provided by IPUMS to migrants in our data. Such an income score indicates the median total income (in hundreds of 1950 dollars) of a person in a given occupation. It should be noted that although the occupational score has a well-established use in economics, its limitations are also well recognized (Abramitzky et al., 2012a). For

[11] In our context, the index of Fryer and Levitt (BNI) would be defined as

$$\frac{\sum k^{1}(\text{Name}_{it} = \text{Name}_{k})}{(\sum k^{1}(\text{Name}_{it} = \text{Name}_{k}) + \sum l^{1}(\text{Name}_{it} = \text{Name}_{k})}$$

(for *k* in US-born and *l* in foreign-born), which varies between 0 and 1 as our A*it*. As it can be seen, the BNI index weights the frequencies of names by the overall distribution of names across groups. When we apply the BNI index in our analysis, we find very similar results.

instance, we are unable to measure within-occupation changes in earnings related to the Americanization of a migrant's name. For an effect to be found, it is necessary that changing one's name pushes an individual into a different mean-wage occupation, which annihilates the actual variation in wages. It is possible that our results would be stronger if individual-level information on earnings were available.

Local Labor Markets. An additional and somewhat unique piece of information contained in the naturalization records is the residential address of the migrant and his dependents at declaration and at petition. We exploit this information in several ways: first, we use it to identify the migrants' local labor market; second, we use it to detect the mobility of internal migrants, i.e. those individuals who at the time of petition moved to a local labor market different than at declaration; and third, at an even more disaggregated level, we control for the potential role of ethnic enclaves.

We implement a geocoding procedure through which we use migrants' addresses to derive coordinates, subsequently assigning local labor markets to each individual. Local labor markets were defined comparing individuals' labor market outcomes across alternative geographic entities. Our preferred choice for local labor market coincides with community districts.[12]

Summary Statistics. Figure 1 in the introduction shows a positive association between name Americanization and occupational scores. Table 2 further explores the differences across migrants who Americanize their names and those who do not, both at the time of declaration and between declaration and petition.

[12] Exploiting the detailed geographical information, it is easy to note that Brooklyn and Manhattan were the two most populated boroughs. Distinguishing the place of residence by major country of birth, some clusters by country of origin are evident, including those of Italians in the lower east side of Manhattan and Russians in the Brownsville area of Brooklyn, yet different ethnic groups coexisted in the same areas. Detailed explanations of the construction of these local labor markets, as well as results and additional tables are given in the online Appendix.

Variable	All	Distinctive	Keepers	Americanize	1^{st}	2^{nd}	3^{rd}	4^{th}
	At declaration							
Log-Occupational Score	3.1509	3.2185	3.1446	3.1579	3.1522	3.1825	3.1668	3.1313
	(0.4679)	(0.4351)	(0.4339)	(0.5376)	(0.5345)	(0.5311)	(0.5781)	(0.5073)
Age	29.9039	28.6471	30.4471	28.8396	28.7556	28.3049	27.8560	30.4177
	(8.8615)	(7.7927)	(8.8184)	(8.9557)	(9.4028)	(9.1017)	(8.5392)	(8.5540)
Years since migration	6.8542	5.7981	5.8920	9.0737	8.2659	8.8560	10.6829	8.5787
	(7.1428)	(6.5587)	(6.7178)	(7.6019)	(7.4536)	(7.6848)	(7.7642)	(7.3253)
Married	0.4408	0.3846	0.4458	0.4356	0.4045	0.4080	0.4390	0.4921
	(0.4966)	(0.4889)	(0.4972)	(0.4961)	(0.4917)	(0.4924)	(0.4973)	(0.5009)
Spouse present	0.3242	0.2885	0.3060	0.3677	0.3521	0.3280	0.4024	0.3898
	(0.4681)	(0.4552)	(0.4609)	(0.4824)	(0.4785)	(0.4704)	(0.4914)	(0.4887)
Has children	0.3108	0.2404	0.3163	0.3058	0.3146	0.2760	0.3008	0.3307
	(0.4629)	(0.4294)	(0.4651)	(0.4610)	(0.4652)	(0.4479)	(0.4595)	(0.4714)
Birthplace								
Italy	0.2174	0.0481	0.2625	0.1357	0.0524	0.1240	0.1301	0.2402
	(0.4126)	(0.2150)	(0.4401)	(0.3426)	(0.2233)	(0.3302)	(0.3371)	(0.4280)
Russian Empire	0.1783	0.3365	0.1004	0.3333	0.3970	0.3960	0.4146	0.1260
	(0.3829)	(0.4748)	(0.3005)	(0.4716)	(0.4902)	(0.4900)	(0.4937)	(0.3325)
Other new migrants	0.2037	0.2308	0.1523	0.3137	0.3483	0.2840	0.2967	0.3228
	(0.4028)	(0.4234)	(0.3594)	(0.4642)	(0.4773)	(0.4518)	(0.4578)	(0.4685)
Germany	0.1089	0.1346	0.1151	0.0924	0.0637	0.0600	0.0772	0.1693
	(0.3115)	(0.3430)	(0.3193)	(0.2898)	(0.2446)	(0.2380)	(0.2675)	(0.3757)
Ireland	0.0942	0.0385	0.1344	0.0118	0.0037	0.0280	0.0081	0.0079
	(0.2922)	(0.1932)	(0.3412)	(0.1080)	(0.0612)	(0.1653)	(0.0900)	(0.0886)
Other old migrants	0.1148	0.0962	0.1456	0.0492	0.0749	0.0480	0.0203	0.0512
	(0.3189)	(0.2962)	(0.3528)	(0.2163)	(0.2637)	(0.2142)	(0.1414)	(0.2208)
Other	0.0826	0.1154	0.0896	0.0639	0.0599	0.0600	0.0528	0.0827
	(0.2753)	(0.3210)	(0.2857)	(0.2447)	(0.2378)	(0.2380)	(0.2242)	(0.2759)
	Difference petition-declaration							
Log-Occupational Score	0.0315	0.0174	0.0173	0.0640	0.0388	0.0510	0.0590	0.1083
	(0.5216)	(0.4327)	(0.4937)	(0.5847)	(0.6523)	(0.5833)	(0.6025)	(0.4862)
Age	5.2965	5.8824	5.3378	5.1460	5.3434	5.0738	5.0166	5.1317
	(2.1923)	(2.7622)	(2.1379)	(2.2324)	(2.6741)	(1.9252)	(2.1889)	(2.0224)
Years since migration	4.6982	5.0865	4.7325	4.5831	4.6779	4.4760	4.4472	4.7205
	(1.7170)	(1.6901)	(1.7086)	(1.7307)	(1.7921)	(1.6627)	(1.7436)	(1.7116)
Married	0.2177	0.3365	0.2191	0.2026	0.2060	0.2160	0.2195	0.1693
	(0.4128)	(0.4748)	(0.4137)	(0.4021)	(0.4052)	(0.4123)	(0.4148)	(0.3757)
Spouse present	0.2228	0.3846	0.2267	0.1976	0.2022	0.2160	0.2033	0.1693
	(0.4439)	(0.4889)	(0.4437)	(0.4362)	(0.4382)	(0.4406)	(0.4419)	(0.4251)
Has children	0.1721	0.2019	0.1747	0.1632	0.1573	0.2040	0.1585	0.1339
	(0.3775)	(0.4034)	(0.3798)	(0.3698)	(0.3648)	(0.4038)	(0.3660)	(0.3412)
N	3353	104	2232	1017	267	250	246	254

Table 2: Characteristics by level of Americanization: at Declaration and differences between Petition and Declaration.

Notes: Standard deviations in parentheses. Distinctive refers to migrants who changed their names into less frequent names; Keepers are migrants who did not change name, or changed name into an equally frequent name; Americanize refers to migrants who changed names into more frequent names; quartiles refer to migrants who Americanize their names.

"Other new migrants" include migrants born in: Finland, Albania, Greece, Malta, Portugal, Azores, Spain, Austria, Austria-Hungary, Bulgaria, Czechoslovakia, Hungary, Poland, Austrian Poland, Russian Poland, Romania, Yugoslavia, Montenegro, Estonia, Latvia, Lithuania, Armenia.

"Other old migrants" include migrants born in: Denmark, Norway, Sweden, Scotland, United Kingdom, Belgium, France, Alsace-Lorraine, Netherlands, Switzerland.

"Other migrants" include migrants born in: Newfoundland, Mexico, Guatemala, Cuba, Jamaica, British West Indies, Barbados, British Virgin Islands, Argentina, Brazil, Colombia, Guyana/British Guiana, East Indies, Philippines, Cyprus, Palestine, Syria, Turkey, Tunisia, Australia.

As shown in the first column, migrants were about thirty years old at the time of declaration, had been in the U.S. for about seven years and around 45% were married (although only 32% had been joined by their spouse in the U.S.), while about 30% had children. A large fraction were born in Italy, Russia or other major sending countries of this time, such as Poland or the Czech Republic. We call these migrants "new migrants" in contrast to the groups originating from Northern Europe. The classification between "old" and "new migrants" distinguishes between ethnic groups who already had an established history of successful settlement in the U.S. and new waves of migrants which ignited xenophobic reactions in the early 1900s. As we expected (given the time period of analysis), approximately 60% of the naturalized migrants in 1930 came from the new sending regions.

The other columns of Table 2 show statistics by name Americanization, which was clearly quite common, marked by 1,017 individuals in the sample Americanizing their name. Both groups of migrants, namely those who changed and those who kept their original names, exhibit essentially the same average characteristics in terms of age, marital status, probability of having children and occupational score measured at declaration time. Differences arise primarily in three traits: first, at the time of petition, migrants who Americanized their names had stayed longer in the U.S. compared to name keepers; second, they had a higher probability of having their spouse present in the country and third, the distribution by country of birth varied substantially. New migrants were far more likely to Americanize their names than old migrants. Part of this is likely driven by the names of new migrants being less common in the U.S. than names of old migrants from the UK or Ireland. This can be seen, for instance, from the fact that among the new migrants Russians are far more likely to Americanize their names than Irish. All these patterns persist when considering each quartile of the Americanization index.

The second panel of Table 2 shows changes in characteristics of our sample over time. There are very few differences across groups. Migrants who Americanize and those who do not Americanize their names do not differ in terms of the timing of filing petition; and therefore, despite the original gap in years since migration, both groups file petitions within five years of having filed first papers. This behavior is also reflected in the change in age. However, importantly, the two groups show substantially different raw means in terms of labor market outcomes. Specifically, changes in occupational score between declaration and petition are essentially zero for name keepers and for those who change into a more distinctive

name. However, for migrants who Americanize their name, the change in occupational score is significantly positive (0.064 log points on average). This payoff is observed across nearly all quartiles of the Americanization index and is graphically presented in Figure 1.

3. EMPIRICAL MODELS

The summary statistics highlight a correlation between name Americanization and occupational standing, although part of this correlation might be driven by observable differences across individuals with different levels of Americanization. More importantly, those who Americanize and those who do not might differ in terms of other - perhaps unobservable - dimensions.

Going beyond a simple descriptive analysis, we present below a number of empirical strategies used to determine whether the correlation persists once compositional differences are taken into account.

OLS estimator. Our first model considers a regression equation relating occupational scores and name Americanization:

$$y_{it} = \beta_0 + \beta_1 A_{it} + \beta_2 t_{it} + x_{it}'\gamma + \beta_3 (t_{it} * COb_i) + \beta_4 (t_{it} * LabMkt_i) + c_i + \varepsilon_{it}. \quad (2)$$

where y_{it} is the log-occupational score of individual i observed at time t. A_{it} is our key variable, as defined in equation (1), representing the normalized frequency of individual i's name at time t in the U.S. born population.

We also include a time variable, t_{it}. Given that not all migrants file their first papers at the same time, a linear time trend is individual specific and captures the different time periods of observation in the data. This trend has both economic and econometric relevance, given that it controls for different immigrant cohorts and changes in the occupational distribution that are driven by the business cycle.

x'_{it} is a vector of time varying socio-economic variables, such as marital status, presence of the spouse in the U.S. or having children.

Equation (2) also includes two interaction terms: one between the time trend and indicators for the country of birth $(t_{it} * COB_i)$ and one between the time trend and indicators

for the local labor market ($t_{it} * LabMkt_i$). Controlling for country of origin and time trends in occupational scores is important in our context. For instance, Abramitzky et al. (2012a) report considerable variation in occupational score patterns across nationalities over time. Furthermore, several documents of this time describe episodes of exclusion of aliens from engaging in particular occupations.[13] Additionally, Moser (2012) provides evidence of taste-based discrimination towards individuals with German-sounding names. Hence, occupational changes might have followed different trajectories depending on the individual's country of birth. Similarly, the interaction term $t_{it} * LabMkt_i$ controls for different labor market patterns that might simultaneously affect occupation scores and name changes.

Finally, the term c_i is a vector of time-invariant and unobservable personal traits that have an impact on occupational scores (such as innate ability) and might be correlated with other regressors in the model. ε_{it} is an error term satisfying the usual assumptions.

This first estimation strategy simply pools the data and applies OLS estimation to equation (2). By doing so, we control for observable variables such as personal and labor market attributes to examine the existence and extent of payoffs associated with acquiring popular American names. If name Americanization were fully explained by observable differences across individuals, controlling for them would eliminate the potential bias in our estimates, which stems from individual self-selection, with the OLS results pinning down the extent to which name Americanization and occupational scores are related.

First-difference estimator. There might be several unobservable traits (such as individual time-invariant ability) represented by c_i in equation (2), that are correlated with name Americanization and occupational scores and might drive our OLS results. We therefore proceed by adopting a second estimation method.

Re-writing equation (2) for occupational scores observed at the time of declaration ($t = Decl$) as:

$$y_{iDecl} = \beta_0 + \beta_1 A_{iDecl} + \beta_2 t_{iDecl} + x'_{iDecl}\gamma +$$

[13] For example, in 1914, the Supreme Court upheld a New York statute excluding non-naturalized Italians from engaging in building the New York subway (Higham, 2002). Occupational mobility might have had different patterns following different pre-naturalization conditions.

$$+ \beta_3 \; t_{iDecl} * COB_i \; + \beta_4 \; t_{iDecl} * LabMkt_i \; + c_i + \varepsilon_{iDecl}, \qquad (3)$$

and for occupational scores observed at the time of petition (t = P et) as:

$$y_{iPet} = \beta_0 + \beta_1 A_{iPet} + \beta_2 t_{iPet} + x'_{iPet}\gamma + \\ + \beta_3 \; t_{iPet} * COB_i \; + \beta_4 \; t_{iPet} * LabMkt_i \; + c_i + \varepsilon_{iPet}. \qquad (4)$$

Taking the difference between these two equations yields a model in first differences:

$$(y_{iPet} - y_{iDecl}) = \beta_1(A_{iPet} - A_{iDecl}) + \beta_2(t_{iPet} - t_{iDecl}) + (x_{iPet} - x_{iDecl})'\gamma + \\ + \beta_3 \; (t_{iPet} - t_{iDecl}) * COB_i \; + \beta_4 \; (t_{iPet} - t_{iDecl}) * LabMkt_i \; + \\ + (\varepsilon_{iPet} - \varepsilon_{iDecl}). \qquad (5)$$

Here, $(A_{iPet} - A_{iDecl})$ is a continuous variable bounded between -1 and 1, indicating how the migrant's name becomes more or less frequent over time in comparison to the U.S. population. From our previous examples, this variable would take a value of one for Giovanni who Americanized his name to William, and zero for both the British William who did not change name and the Italian Giovanni who changed his name to Salvatore.

Due to the structure of our data we need a modification of (5). Equation (5) correctly measures the association between changing names and occupational scores if all migrants change their names between declaration and petition. If migrants were to Americanize their names prior declaration, the previous equation would consider these migrants as name keepers. For instance, if Giovanni were to become William after arrival yet before declaration, $(A_{iP\,et} - A_{iDecl})$ would take a value of zero as we would observe Giovanni holding a common American name in both time periods. However, if an association exists between names and occupational scores, y_{iDecl} would have already been influenced by the new name, i.e. Giovanni's y_{iDecl} would have changed in response to the new name. To avoid confounding name changers and name keepers, we substitute A_{iDecl} with $A_{iArrival}$, one of the few information reported in the certificate of arrival. Through this implementation, the "will-be-William" Giovanni is considered as having Americanized his name, as opposed to being a name keeper. Using "name at arrival" is arguably more exogenous to unobserved U.S. shocks that might occur after arrival and potentially provides a more conservative estimate of the relationship of interest if name Americanization acts as a one-time change in occupational scores. In estimating equation (6) we implicitly assume that there is no dynamic selection into name Americanization. In other words, we assume that individuals who change names between arrival and declaration are observationally similar to those changing names between

declaration and petition, after controlling for relevant characteristics. This assumption is used by Arai and Thoursie (2009) and allows us to substitute A_{iDecl} with $A_{iArrival}$. We perform several robustness checks with respect to this definition in Section 6 with further details available in the online Appendix.

To summarize, our second estimation method is a first-difference technique (FD) based on the following model:

$$(y_{iPet} - y_{iDecl}) = \beta_1(A_{iPet} - A_{iArrival}) + \beta_2(t_{iPet} - t_{iDecl}) + (x_{iPet} - x_{iDecl})'\gamma + \\ + \beta_3 \ (t_{iPet} - t_{iDecl}) * COB_i \ + \beta_4 \ (t_{iPet} - t_{iDecl}) * LabMkt_i \ + \\ + (\varepsilon_{iPet} - \varepsilon_{iDecl}). \qquad (6)$$

Compared to a simple pooled OLS estimator, equation (6) has the advantage of being able to purge out time-invariant and unobservable personal traits (such as innate ability) that might have an impact on occupational scores and the propensity of name Americanization. At the same time it allows controlling for specific trends in earning growth by country of birth and local labor markets, thus in effect reducing potential biases due to occupational or geographical sorting.

If these characteristics were sufficient to guarantee conditional independence of name Americanization from the unobserved component of the model, we could conclude that these results estimate a causal effect of name changes on occupational scores. However, it is possible that several relevant factors might still be omitted in the regression analysis, that reverse causality might be at work and that time varying self-selection might be present. For example, migrants who Americanized might have accumulated different levels of human capital or language skills. Additionally, name Americanization might occur before occupational change for some individuals, while occupational change might lead to name changes for others. The next two estimators aim to control for time-varying factors that might be correlated with name Americanization and occupational upgrading.

Name-changers only. We estimate equation (6) on the subsample of name changers only (NC), thus exploiting the different timing of name Americanization across individuals. With this strategy, the control group for those who Americanize their names at the time of declaration are those who change their names at the time of petition. This should better control for time-varying individual factors that might cause a spurious correlation between our key variables (a similar strategy is followed in Arai and Thoursie, 2009). For example, as long as those who Americanize their names have similar human capital accumulation patterns - unrelated with the timing of name Americanization - this strategy will net out the effects of

these variables.

IV estimator. While focusing on name changers reduces endogeneity concerns, identification still relies on the assumption of parallel trends between current and future name changers (after conditioning on the full set of characteristics). Furthermore, this method does not rule out the possibility of reverse causality. To recover a causal link between name Americanization and economic success, we apply an instrumental variable (IV) strategy on equation (6).

In order to understand the causal link between changing names and economic success, we need a variable that predicts the former without directly influencing the latter. We look for a measure of linguistic complexity that is unrelated to the specific content of a name, given that the latter might be correlated with socioeconomic outcomes and, hence, ultimately correlated with occupational scores.

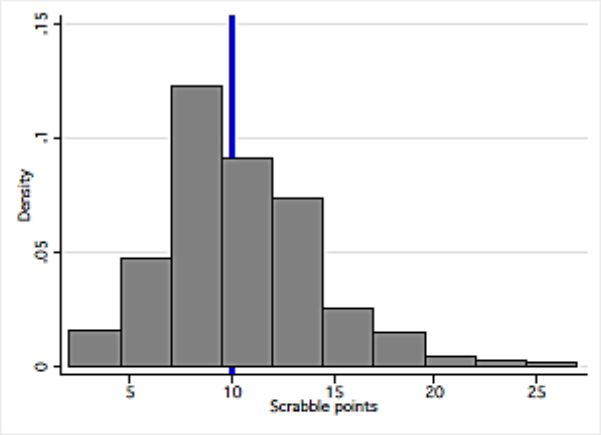
We calculate the "Scrabble points" for each name at arrival by summing the scores attributed to each letter in the popular board game and use these points to predict name Americanization. The origin of the Scrabble point system dates to 1938 and is attributed to the architect Alfred Moscher Butts who performed a frequency analysis of letters from the front page of various newspapers. A name associated with relatively high points would correspond to a linguistically complex name whereas a name with relatively low points would correspond to a relatively simple phoneme or to a euphonious name.

In practice, we create a measure of distance between the Scrabble points of the migrant's name at arrival and the Scrabble points associated with the "American norm". Our Scrabble index $S_{jArrival}$ is defined as:

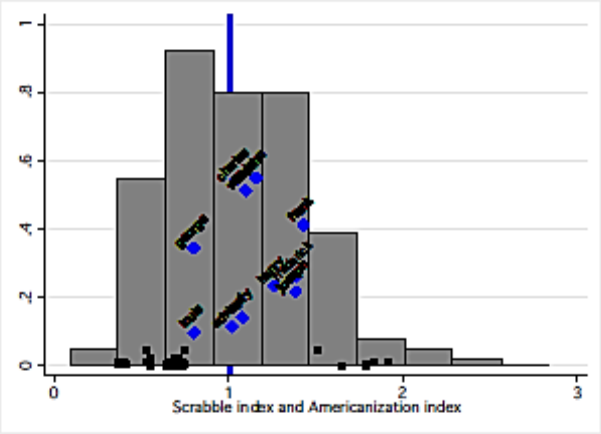
$$S_{jArrival} = \frac{SP_{jArrival}}{\Sigma_{w \neq j} SP_w / \Sigma_{w \neq j} 1} \quad (7)$$

where $SP_{jArrival}$ is the Scrabble points of name j, divided by the median Scrabble points across all American individuals born before 1880 and currently living in the state of New York, excluding name j from the computation of the median. Figure 3(a) shows the distribution of Scrabble points in the U.S. born population. Values of S_j substantially above or below 1 indicate a name that is "phonetically" distant from the American norm. This normalization allows us to measure the degree of complexity of the arrival name with respect to the available linguistic complexity at destination.

(a) Scrabble Points, U.S. born population 1880



(b) Scrabble index and Americanization Intensity



(c) Difference in Scrabble index over time

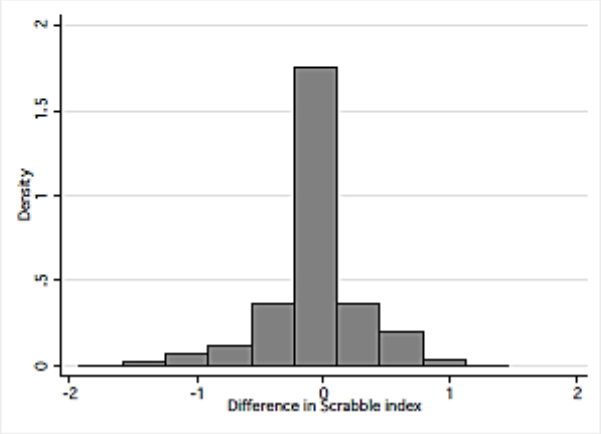


Figure 3: Scrabble points, Scrabble index and Americanization index.

The Scrabble index seems to be a suitable instrument for several reasons. Given that points are based on the frequency of letters, the Scrabble index does not make use of the semantics associated with names. For the same length, two names score differently depending on their complexity with respect to the American norm. For this reason, it is unlikely that this index is associated with individuals' labor market performance and thus our exclusion restriction should be valid. While we cannot directly test for the validity of the exclusion, we report summary statistics of our sample in the online Appendix by separating our sample into those who have below and above median values of S_j. The table suggests that both groups of individuals are observationally similar. In particular, there are no differences in terms of years since migration, which means that the instrument is capable of purging any channel linked to human capital accumulation, including the acquisition of language skills. Furthermore, Aura and Hess (2010) show that Scrabble points are generally unrelated with lifetime outcomes, and, particularly, do not predict occupational prestige during 1994 and 2002.

The remaining panels of Figure 3 show how our Scrabble index predicts name Americanization. Figure 3(b) shows the graphical distribution of $S_{jArrival}$; the vertical line indicates the median, which is very close to one. In the same graph, we report selected values of the changes in the Americanization index associated with the Scrabble index (for ease of representation, we calculated the means of Scrabble index and Americanization index by aggregating for each name at petition). The blue markers refer to values of name Americanization that are above the mean, while the remainder are values pertaining to observations from the lowest decile. The first noteworthy aspect is that the distribution of S_j at both points in time is bell shaped, indicating that a larger mass of people have names of average complexity in the U.S. Second, and more importantly, there is a clear relationship between the distribution of $S_{jArrival}$ and name Americanization. The highest values of Americanization are found within an interval of the Scrabble index corresponding to the American norm. On the other hand, the lowest values of Americanization are scattered further away, towards both tails of the distribution. This means that the Scrabble index could be considered as a functional predictor of Americanization. Another way of interpreting this is to observe the "dynamics" of the Scrabble index for individuals who Americanize. In Figure 3(c), we show the distribution obtained by subtracting $SP_{jArrival}$ from the Scrabble points calculated at the time of petition, when migrants changed name. We omit from the graph name keepers (i.e. those for whom $SP_{jArrival}$ did not vary over time). The distribution shows that the vast majority of changes in the Scrabble index happen in an interval around 0.5 points from the American norm. This means that only those close to the "Scrabble norm" find it beneficial to change name. For those who are too far from the norm, the "costs" of Americanizing their name are so high that we observe very few instances of name changes, as well as a negligible increase in the Americanization index. Remarkably, the same pattern

emerges when we look at this "difference" distribution by country, as well as by aggregating across popular names of migrants at petition (see the online Appendix for additional details). These graphical representations suggest the existence of a non-linear relationship between S_j and name Americanization and convex costs in name Americanization.

While the IV estimator has the advantage of showing whether the positive association found in the summary statistics persists after controlling for time-varying, individual-specific factors, it is important to remember that these estimates only measure a local average treatment effect (providing the assumptions given in Angrist et al. (1996) are satisfied). This is the average effect of changing names on occupational scores for complier migrant men only, i.e. those migrant men who would change their name because of particular values of the instrument, but who would not change their name otherwise. The LATE estimates could differ from the average treatment effect of name changes for the entire population of naturalizing migrant men.

The OLS estimator, the modified first-difference estimator (FD), focusing on name changers only (NC) and the instrumental variable method (IV) are all designed to examine whether name Americanization had a return in the labor market, i.e. the consequences of such choices.

4. THE ECONOMIC PAYOFF OF NAME AMERICANIZATION

We start the analysis in Table 3, with the first three columns showing the estimates in levels (i.e. pooling equations 3 and 4), ignoring individual heterogeneity, while the last three columns show the estimates of equation (6). Across columns we vary the set of regressors given in our empirical framework. These range from migrant and household characteristics (whether or not married, whether or not children are present and whether or not a spouse is present) to trends by country of birth and by local labor market. For notational simplicity, we use A_{it} throughout to indicate the Americanization index, irrespective of the type of estimator applied.

	OLS			First differences		
	I	II	III	IV	V	VI
A_{it}	0.0426**	0.0375**	0.0474**	0.1178***	0.1344***	0.1412***
	(0.0185)	(0.0190)	(0.0192)	(0.0376)	(0.0378)	(0.0385)
Marital status	0.0255	0.0305*	0.0366**	0.0332	0.0327	0.0228
	(0.0172)	(0.0174)	(0.0172)	(0.0269)	(0.0265)	(0.0272)
Has children	−0.0041	0.0030	−0.0076	0.0028	0.0027	0.0076
	(0.0127)	(0.0127)	(0.0127)	(0.0202)	(0.0209)	(0.0213)
Spouse present	0.0407***	0.0310**	0.0105	−0.0258	−0.0147	−0.0089
	(0.0142)	(0.0141)	(0.0145)	(0.0260)	(0.0265)	(0.0277)
Italy		−0.0617**	−0.1024***		0.0028	0.0086
		(0.0263)	(0.0298)		(0.0055)	(0.0065)
Russian Empire		−0.0251	−0.0629		0.0061	0.0134
		(0.0361)	(0.0428)		(0.0067)	(0.0089)
Other new migrants		0.0073	−0.0162		−0.0001	0.0055
		(0.0254)	(0.0300)		(0.0044)	(0.0054)
Germany		0.0483*	0.0199		−0.0028	0.0014
		(0.0261)	(0.0300)		(0.0048)	(0.0055)
Ireland		−0.0781***	−0.1154***		0.0186**	0.0219***
		(0.0271)	(0.0310)		(0.0075)	(0.0082)
Other		−0.1162**	−0.1448***		0.0355***	0.0424***
		(0.0497)	(0.0551)		(0.0111)	(0.0121)
Trend	Yes	Yes	Yes	Yes	Yes	Yes
Lab. Mkt. Ind.	No	No	Yes	No	No	Yes
R^2	0.01	0.02	0.08	0.01	0.01	0.03
N	6706	6706	6706	3353	3353	3353

Table 3: Effect of Name Americanization on Log-Occupational Score, Benchmark Specification

Notes: Robust standard errors in parenthesis. A_{it} = Americanization index. See explanation on page 6. Labor market indicators refer to dummy variables for each of the NYC community districts. See text for explanation. In first difference models, country of birth and labor market indicators are interactions between these variables and the time trend. All regressions include a time trend. Reference category for country of birth: Other old migrants. Married, Has children, Spouse present in the U.S. are all indicators. See text for explanation. See Table 2 for explanation of the country of birth categories

OLS estimates. In the pooled OLS regression, name Americanization is associated with a 4% increase in occupational score, a payoff that remains stable across the three specifications. Having a spouse present in the U.S. is associated with higher occupational scores, while marital status or having children does not seem to have an impact on the migrant's economic outcomes. Note that these variables aim to capture time-varying characteristics that might affect assimilation incentives - and hence occupational upgrading. In column (II), we condition on indicators for nationality-specific trends to capture different labor market patterns across origin groups. We find that our results become statistically weaker and point estimates slightly decrease. Nonetheless, the association between Americanization and occupational upgrading remains positive even within nationality-specific occupational score patterns. Comparing

nationalities with the group of other old migrants (including British, Swedish and Norwegian migrants), Germans exhibit higher labor market success, while all the other nationalities under-perform.

Lastly, we explore the influence of local labor markets by including in Column (III) trend indicators for community districts in which the immigrants live at the time of declaration. The pattern of our baseline results remains unchanged and we find a statistically significant return to name Americanization. Despite some evidence of sorting shown in the online Appendix, this seems to indicate that the impact of local labor market conditions is relatively mild and does not drive our main findings.

First-difference estimates. Exploiting the panel structure of the data, the last three columns of the table report the estimates in first differences. The positive association between name Americanization and occupational-based earnings persists. Estimates become larger and statistically stronger compared with the OLS estimates. The positive association between name Americanization and occupational upgrading remains once unobserved heterogeneity is controlled for. Furthermore, payoffs are larger for individuals with uncommon names who choose very common American names.

Conditioning on a number of variables, we find that changing from a purely foreign name to a very common American name is associated with a 14% return, while household characteristics have little impact on occupational upgrading. Additionally, nationality specific trends indicate the presence of little occupational convergence, except for the Irish and the "Other" category migrant group.

Assuming that individual self-selection is purely driven by either observable traits that we control for or time-invariant characteristics, these estimates indicate that the positive association found in our OLS analysis is not explained by selection but rather is likely driven by the market rewarding migrants holding an American name. In fact, name Americanization results in a 14% increase in occupational scores when individual productivity remains unchanged.

Estimates for name-changers only. Within the first-difference estimator, we restrict our sample to name changers only, i.e. we exploit the different timing of the name choice. We estimate the model progressively adding regressors as in our baseline specification. Results are shown in the first three columns of Table 4. While the sample sizes are smaller, patterns are quite remarkable as we still find a positive relationship between acquiring popular American names and economic success in the U.S. As long as name changers in one year have similar trends to those in a different year, these estimates capture the causal effect of name Americanization on occupational scores. Furthermore, as long as time-varying

unobservable characteristics are shared within this group, these estimates implicitly control for variables such as English language acquisition and human capital accumulation which are unobservable to us

	Name changers			Instrumental variable			Instrumental variable		
	I	II	III	IV	V	VI	VII	VIII	IX
A_{it}	0.1091***	0.1136***	0.1313***	0.8083**	0.5604*	0.6850**	0.8495***	0.5977**	0.7116**
	(0.0417)	(0.0424)	(0.0440)	(0.3497)	(0.3156)	(0.3458)	(0.3005)	(0.2727)	(0.2894)
N	1538	1538	1538	3353	3353	3353	3353	3353	3353
Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
C. of birth Ind.	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Lab. Mk. Ind.	No	No	Yes	No	No	Yes	No	Yes	Yes
				First stage			First stage		
$S_{jArrival}$				0.0499***	0.0550***	0.0543***	0.1266***	0.1486***	0.1530***
				(0.0069)	(0.0072)	(0.0074)	(0.0196)	(0.0196)	(0.0201)
$(S_{jArrival})^2$							−0.0438***	−0.0532***	−0.0557***
							(0.0096)	(0.0095)	(0.0098)
Marital status				0.0054	0.0080	0.0078	0.0023	0.0042	0.0036
				(0.0152)	(0.0150)	(0.0149)	(0.0152)	(0.0150)	(0.0149)
Has children				−0.0133	−0.0158	−0.0147	−0.0130	−0.0151	−0.0137
				(0.0101)	(0.0100)	(0.0101)	(0.0101)	(0.0100)	(0.0101)
Spouse present				−0.0187	−0.0203	−0.0196	−0.0155	−0.0168	−0.0159
				(0.0150)	(0.0149)	(0.0148)	(0.0150)	(0.0149)	(0.0148)
Italy					0.0067***	0.0070***		0.0060***	0.0063**
					(0.0022)	(0.0025)		(0.0022)	(0.0025)
Russian Empire					0.0061***	0.0075***		0.0065***	0.0081***
					(0.0022)	(0.0027)		(0.0022)	(0.0027)
Other new migrants					0.0158***	0.0170***		0.0164***	0.0177***
					(0.0025)	(0.0029)		(0.0025)	(0.0029)
Germany					0.0142***	0.0140***		0.0147***	0.0146***
					(0.0028)	(0.0030)		(0.0028)	(0.0030)
Ireland					−0.0082***	−0.0078***		−0.0086***	−0.0082***
					(0.0019)	(0.0022)		(0.0019)	(0.0022)
Other					0.0028	0.0030		0.0023	0.0026
					(0.0026)	(0.0029)		(0.0025)	(0.0029)
F 1st stage				51.737	58.416	53.685	30.473	37.773	36.411
Partial R^2				.013	.016	.015	.018	.023	.023
Wooldridge test p-value				.038	.163	.103	.008	.074	.038
Sargan test							.859	.871	.915
N				3353	3353	3353	3353	3353	3353

Table 4: Effect of Name Americanization on Log-Occupational Score, Causality Tests

Notes: Robust standard errors in parenthesis. A_{it} = Americanization index. See explanation on page 6. See Table 2 for explanation of the country of birth categories. $S_{jArrival}$ refers to the scrabble index as explained in the text. Wooldrdige test refers to a robust score test of endogeneity (Wooldridge, 1995). In all models, country of birth and labor market indicators are interactions between these variables and the time trend.

IV estimates.To further account for any other source of endogeneity, we conclude this section by presenting the IV estimates of equation (6). In the first stage, we start by using our Scrabble index as a simple linear predictor of Americanization in columns IV to VI of Table 4. In columns VII to IX, we model the non-linear relationship, as visually represented in Figure 3, adding a quadratic term. Throughout our models, $S_{jArrival}$ is positively associated with name Americanization. The quadratic term is also highly significant, hinting at a concave relationship between the Scrabble index and name Americanization, even after controlling for all covariates. This suggests that individuals who Americanize have higher Scrabble points, although incentives to Americanization decrease as names differ more from the norm. Intuitively, the cost of name Americanization is smaller for individuals around the norm. This panel also reports the impact of other characteristics and nationality-specific trends on name Americanization. Compared to the other old migrants, all other nationalities - with the exception of the Irish - seem to be more likely to adopt American names over time. The instruments continue to perform well independently of the controls added and remain relevant predictors, as shown by the first stage F-statistics, which indicate the validity of our instrument. A quadratic model in the Scrabble index allows us to test the validity of our exclusion restriction via a Sargan test for overidentifying restrictions. We cannot reject the null hypothesis of excluding either instrument from the first stage. Finally, we tested whether there is evidence of endogeneity. While we cannot fully reject exogeneity of name Americanization in the first model, we reject this hypothesis at the five per cent significance level in the second.

After testing the validity of the instruments, we find in the second stage that the returns to name Americanization increase once we instrument for A_{it}. It appears that controlling for the possible time-varying self-selection of individuals as well as any other form of measurement error into the decision to change names increases the returns to name Americanization to about 70%. As mentioned earlier, this should be interpreted as the average effect of name Americanization on occupational scores for complier migrant men, i.e. for those who change their name because they have a rather high Scrabble index, but who would not have changed their name had they had a low Scrabble index. While potentially different from the average treatment effect, the local average treatment effect further corroborates the existence of a causal relationship between the variables of interest.

In summary, the positive relationship between name Americanization and occupational upgrading persists after directly tackling the endogeneity of name Americanization with our IV approach. Our results suggest that it is the act of Americanization per se that causes higher occupational upgrading, and not (only) the selection of migrants into such a choice, since the IV estimates can be interpreted as the occupational payoff consequent to “random” name Americanization.

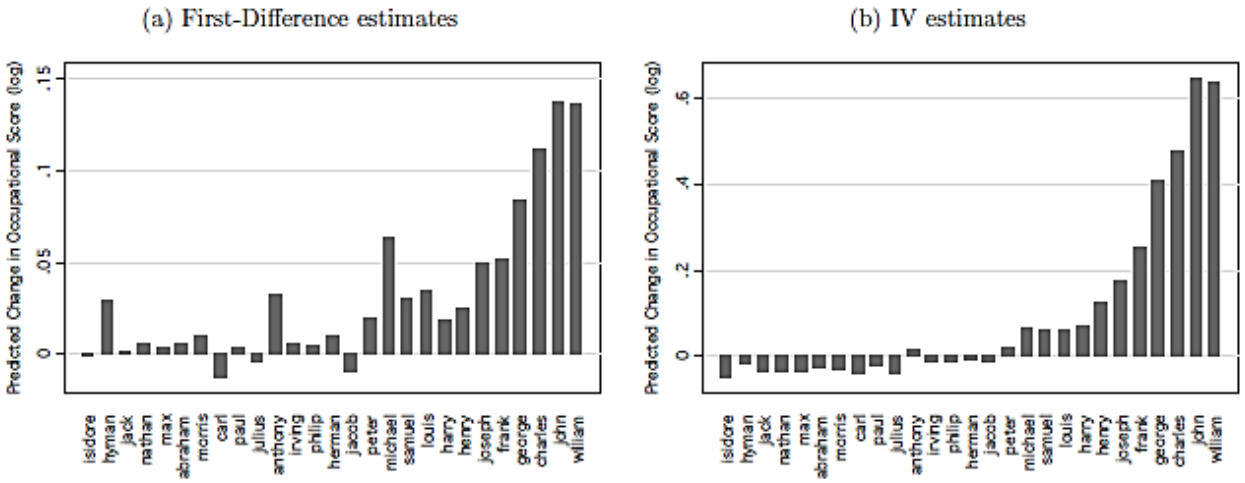


Figure 2: Effect of Name Americanization on Log-Occupational Score, by Name

Figure 2 further summarizes this finding, highlighting the positive association between name Americanization and occupational upgrading, with larger payoffs the more popular American the new name. Figure 2(a) shows the returns from the FD model, while Figure 2(b) shows the same returns in the IV model. These payoffs were calculated using the estimates of equation (6) from the two models and show the returns for those who Americanized their names. Changing a name into William or John is associated with larger occupational upgrading compared to keeping one's name or changing into a less common American name. On the other hand, adopting names such as Hyman or Isidore, which were less common albeit present in the U.S. born population, has a much lower - statistically zero - payoff. Interestingly, as these names are distinctively Jewish, it appears that name Americanization that occurred while embracing a cultural heritage had little return.

5. CHANNELS

The existence of positive returns to name Americanization poses two important questions: first, why some individuals Americanized their names and others did not; second, concerning the expected direction of self-selection into name Americanization.

There might be various reasons why only some migrants Americanized their names. Fryer and Levitt (2004) list several possible frameworks that help explain the persistence of distinctive names. First, individuals might not have full information on the returns to name Americanization. Second, within their price theory model of names, incentives for name Americanization may differ by neighborhood composition. For example, within ethnic enclaves, the benefits from name Americanization might be lower while the costs might be higher. Third, individuals might have utility gains if they follow a prescribed behavior of a

particular group (Akerlof and Kranton, 2000). Hence, only those individuals subject to few identity prescriptions will Americanize their names.

While these frameworks suggest that migrants who Americanize their names should be observationally different than those who do not, it is unclear *a priori* whether Americanization occurs among migrants, who face occupational mobility or those who do not. For instance, in the presence of imperfect information, migrants with higher ability would probably be more likely to Americanize their names. In the context of a price theory model, individuals in more segregated neighborhoods might face a higher cost of Americanization, if there is a cost to holding an American name when interacting with other migrants. Yet, whether the individual Americanizes his name or not depends on whether the particular ethnic enclave consists of mostly high- or low-skilled workers. Lastly, within an identity framework, it is uncertain *a priori* whether behavioral prescriptions would be stronger among those who are successful in the labor market. Overall, the type of migrant who Americanizes cannot be pinned down unequivocally.

Our results highlight that individual self-selection played a role in name Americanization beyond simple observable differences across individuals. A comparison of the OLS estimate (4%) with the FD (13%), NC (13%) and IV (70%) estimates highlights an interesting and consistent direction of the bias in the OLS estimator: there is negative selection on time-invariant (FD) and time-varying (NC, IV) unobservable characteristics. Hence, it was those migrants facing the greatest barriers to occupational mobility who Americanized their names.

This result hints at imperfect information not being a driving factor of selection into name Americanization. This evidence is also consistent with Goldin and Shim (2004), who find that keeping the maiden name is more common among highly skilled women. It is likely that high achieving individuals did not need to change their names as their quality was already signalled in the market, as they were subject to less discrimination or could find alternative ways of occupational improvements such as migration or human capital accumulation.

We further corroborate the hypothesis that name Americanization was a way of overcoming labor market barriers by examining the sub-populations in which name Americanization had a higher return. We would expect higher payoffs for groups that were low skilled, potentially more discriminated against, or with less alternative means for socio-economic improvement.

	Old Migrants				New Migrants			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0460	0.0911	0.0988	−0.5149	0.0544**	0.1526***	0.1434***	0.6963**
	(0.0290)	(0.0566)	(0.0776)	(0.7953)	(0.0265)	(0.0467)	(0.0515)	(0.3171)
N	2070	1035	251	1035	4636	2318	1287	2318
	Internal Migrants				Internal Stayers			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0193	0.0559	0.0567	0.2148	0.0608*	0.2325***	0.1713**	1.2913**
	(0.0201)	(0.0353)	(0.0424)	(0.3116)	(0.0362)	(0.0743)	(0.0852)	(0.5298)
N	3830	1915	881	1915	2876	1438	657	1438
	High Exposure				Low Exposure			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0273	0.0081	0.0005	−0.0737	0.0429**	0.1550***	0.1595***	0.5619***
	(0.0387)	(0.0384)	(0.0444)	(0.3237)	(0.0208)	(0.0399)	(0.0484)	(0.2141)
N	1400	700	414	700	4202	2101	858	2101

Table 5: Effect of Name Americanization on Log-Occupational Score, Channels

Notes: Robust standard errors in parenthesis. First Stage F-tests: 3.849, 33.232, 12.815, 17.67, 7.449, 25.537. A_{it} = Americanization index. See explanation on page 6. All models include the covariates in the last column of Table 3. OLS = Ordinary Least Squares estimates; FD = first-difference estimates; NC = name changers only estimates; IV = instrumental variable estimates. In FD, NC, IV models, country of birth and labor market indicators are interactions between these variables and the time trend.

Table 5 divides the sample into several sub-populations. We report estimates of the payoff to name Americanization across all estimators for new (old) migrants, for internal migrants (stayers) and for individuals who live in ethnic enclaves.[14]

In the first panel, we distinguish the effects between migrants whose first settlement happened during the early wave of mass migration (Germans, Scandinavians, Irish, British) and those who came from regions with a later experience of emigration to the U.S. (Italians, Eastern Europeans, Russians). The pattern of results for the new migrants is similar to our baseline estimates. For the old migrants, the magnitude of estimates is smaller and the effect is statistically insignificant throughout. Therefore, it appears that changing into a popular American name only determined a premium for the "new" migrants. These heterogenous returns could be driven by stronger discrimination in the labor market for this group, or by their relatively low level of skills. Historical accounts (e.g. Higham, 2002) seem to confirm that there was widespread resentment and discrimination against the new migrants.

In the second panel, we focus on individuals who changed labor markets between declaration and petition (internal migrants) and those who did not. Internal migrants

[14] In the online Appendix, we perform additional analysis across physical characteristics and occupations. All results are consistent with our interpretation.

exhibited a much smaller and statistically insignificant return to name Americanization, while the payoff was particularly high for those who did not move. These results are compatible with an interpretation of migration as a form of human capital investment. Acquiring skills through migration can be seen as a strategy of earning growth that is an alternative to changing name. Therefore, returns to Americanization will be smaller for internal migrants. Further- more, if internal migrants are positively selected - as arguably only those who can afford to migrate will do so - the higher payoff for non-migrants is again indicative of name Americanization as a way of circumventing lock-in effects for those who had no alternatives means for occupational mobility.

The last panel of the table explores the role of ethnic enclaves. We calculate a measure

of exposure to ethnic enclaves using the 1% sample from the 1920 IPUMS, defined as

$$\frac{P_{kh}}{P_h}$$

Here P_{kh} represents the number of individuals living in census tract h who were born in country k and P_h is the total number of individuals – including U.S.-born – in each tract. This index provides a measure of the relative size of each immigrant group in the locality and has previously been used to study the role of enclaves (e.g. Borjas, 2000). We focus on a measure pertaining to the year 1920 in order to capture an ethnic composition that is pre-determined at the time of deciding to change name. We then distinguish between migrants living in neighborhoods with high (low) ethnic concentration, defined as those areas in the top quartile (first three quartiles) of the exposure distribution.[15]

The estimates suggest that name Americanization did not pay for individuals living in enclaves. On the other hand, migrants in less segregated neighborhoods had a higher return to name Americanization. Once again, this indicates that name Americanization paid off for those individuals who were more constrained, and faced the greatest economic and social barriers.[16]

[15] Since we are only able to match the census tracts for Manhattan and Brooklyn at the time of writing, the sample size in this exercise is limited to immigrants living in these two boroughs, which means that those living in Bronx, Queens, Staten Island and outside New York City (internal movers) at the time of declaration are excluded.

[16] The online Appendix also reports additional channels of heterogeneous effects, such as surname Americanization, employment types, age or occupations particularly hit by the Great Depression.

6. ROBUSTNESS CHECKS

Ee carry out a series of checks aiming to ascertain the robustness of our results to various data aspects and definitions. Results are reported in Table 6, and further details are available in the online Appendix.

	Type 3				Early Declarants			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0381	0.2385**	0.2162	5.0440	0.0294	0.1017**	0.1058*	0.9124*
	(0.0296)	(0.1208)	(0.1506)	(5.5015)	(0.0272)	(0.0485)	(0.0592)	(0.4771)
N	4374	2187	372	2187	3042	1521	571	1521
	Soundex				All 1930 Names			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0397**	0.1280***	0.1125**	0.9793**	0.0521***	0.1432***	0.1329***	0.6156**
	(0.0178)	(0.0412)	(0.0444)	(0.4463)	(0.0199)	(0.0406)	(0.0467)	(0.2468)
N	6706	3353	1538	3353	6706	3353	1538	3353

Table 6: Effect of Name Americanization on Log-Occupational Score, Robustness

Notes: Robust standard errors in parenthesis. First Stage F-tests: .766, 8.599, 22.041, 49.729.
A_{it} = Americanization index. See explanation on page 6. All models include the covariates in the last column of Table 3. OLS = Ordinary Least Squares estimates; FD = first-difference estimates; NC = name changers only estimates; IV = instrumental variable estimates. In FD, NC, IV models, country of birth and labor market indicators are interactions between these variables and the time trend.

Our model in equation (6) uses a definition of name Americanization based on the difference in the frequency of the name at petition and at arrival. While the petition was the document that officiated the new name, such a change could have occurred prior to the time of petition, which we can identify to some extent. As a migrant name is known at the time of arrival, at the time of declaration and at the time of petition, we can distinguish between migrants who hold the same name over time (henceforth Type 1), those who hold a certain name at arrival and another one at declaration (Type 2) and those who hold the same name up until the time of petition (Type 3). There were no cases in the data whereby the name changed twice into two different names. There are about 54.13% Type 1 individuals, while among those who change name, 75.81% were Type 2 and the rest were Type 3.[17]

The data challenge arises from the fact that, despite observing names at three different points in time, occupations are only observed at declaration and petition. In an ideal data

[17] We also observe a fourth type of individual who change name between arrival and the time of declaration yet report the original name at the time of petition (around 3.62% of the name changers). It is unclear from the data whether these individuals had been temporarily using the "new" name or whether the new name had been 'imputed' or 'misunderstood' by the clerk. Therefore we checked all these individuals' names and compared the reported names with the names in the signature. The individuals who signed the declaration with their original name were considered as name keepers while the rest were probably individuals who really changed their name multiple times. Results remain unchanged if we drop this type of individuals.

set, we would observe individual names at the first occupation after arrival (prior to the name change), compare these with the name at arrival, and then observe these individuals over time. However, our data do not contain information on the first occupation upon arrival and prior-name change. To our knowledge, no other source can be used to detect this information, as passenger records report neither the occupation at origin nor potential occupations in the U.S. (Ferrie, 1999).[18]

Our baseline model considers changes in name between petition and time of arrival, ($A_{iPet} - A_{iArrival}$), in order to capture all the name changers in the sample. As mentioned in previous sections, this definition offers the advantage of defining a name change based on a variable that is pre-determined with respect to any possible post-arrival outcome in the U.S. Therefore, this definition reduces potential concerns of reverse causality.

While the benchmark model correctly identifies Type 2 individuals as being 'treated' by the change of name, their occupational scores have already been affected by such a change at the time of declaration. If name Americanization provided a one-time increase in occupational scores, the inclusion of these individuals in the "treatment group" would bias down the baseline results, as occupational scores' trajectories would have already changed for them. Therefore, our baseline definition would provide a lower-bound effect. On the other hand, if name Americanization implied a change in both average occupational scores and growth trajectories, with the inclusion of Type 2 individuals we would still be able to detect a return to name Americanization, although we would not be able to identify the short- run change from the long-run change in occupational scores. However, under this second scenario, the association between names and occupational scores would still be detected by our analysis. These are the reasons why we prefer to keep our definition based on name at arrival in our baseline analysis, as it is potentially rather conservative and provides a larger sample size.[19]

We adopt two strategies to examine whether our preferred empirical model delivers conservative estimates. First, we drop the Type 2 individuals and therefore estimate a true first-

[18] If a larger sample size were available, it would be possible to match the migrants of our data set with the Census enumeration and obtain a third point in time. We would need, however, enough observations of individuals who arrive prior to 1920, who do not change name by the 1920 Census, and declare intention to naturalize between 1920 and 1930. For the time being, this route did not seem feasible.

[19] Further details are provided in the online Appendix.

difference model.[20] Second, we restrict our sample to individuals who declare their intention to naturalize within three years of arrival and then drop everybody else in the sample. By doing so, although we cannot fully control for the fact that Type 2 migrants have already been 'treated' compared to Type 3 migrants, we can at least limit the effects of new names to a shorter time span. Furthermore, we are able to better capture the trajectories immediately after arrival and, presumably, work with a more exogenous initial condition for the occupational distribution.[21]

The results in Table 6 are consistent with previous estimates. Point estimates are higher if only changers between declaration and petition are kept. It should be noted that given the small number of Type 3 individuals, it is perhaps unsurprising that the estimates for NC and IV are very imprecise. On the other hand, restricting the sample to those early declarants does not particularly alter our main conclusions, given that we still find a positive relationship between changing name and occupational upgrading.

Lastly, we run several analyses to ascertain the sensitivity of our results to how we define name Americanization. When looking at historical records, it is well known that many orthographic differences might appear in first and last names. Under such circumstances, it might be important to standardize names using phonetically equivalent transliterations. Common algorithms to obtain name standardizations are Soundex, NYSIIS, Metaphone, and Caverphone. In the context of our study, it could be the case that a misspelling between our data and the Census causes the Americanization level to be measured with an error. For example, if the migrant's name at arrival John is misspelled as Jon in our sample, he would be considered to have Americanized his name when in fact the name had not changed. We report our baseline results by considering a name change to be a change into a phonetically different name, thus reducing differences due to simple misspellings, as well as requiring name Americanization to be a change in the "sound" of the name. To this aim, we use the Soundex algorithm. Names that sound the same yet are spelled differently, such as John and Jon, have the same code. Therefore, we are able to purge possible orthographical errors made in the original record by the court clerks as well as during our data collection process. As can be seen in the first four columns of the second panel of Table 6, results remain within a confidence interval of our baseline analysis, even when this restrictive definition is adopted.

[20] We wish to stress, once again, that we are aware that this will not induce additional bias only under the assumption that, conditional on characteristics, the timing of name change is exogenous, or put differently, that name changers at different points in time have parallel trends.

[21] We want to emphasize once again that reporting a first-difference model on the whole sample would not be meaningful as we would then consider as "untreated" the migrants who changed their names between arrival and declaration.

The final four columns change the definition of American norm, by constructing the A_{it} variable using the name distribution of the full population living in New York in 1930. Changing the norm definition does not alter our conclusions.[22]

7. CONCLUSIONS

This paper provides the first direct evidence on the magnitude and consequences of name Americanization. Such a phenomenon, previously only known anecdotally, was not only widespread, but also had substantial impacts on upward mobility during the first half of the twentieth century. We find, in fact, that immigrants who Americanized their names experienced higher occupation-based earnings growth than those who did not. These results persist across all our specifications, including those controlling for individual heterogeneity or those based on an instrumental variable technique where we use a Scrabble index, reflecting the linguistic complexity of migrant names as an instrument for name Americanization. Hence, our results are not driven by migrant self-selection into this choice.

These stark differences stemming from name Americanization provide important insights for current research. Our unique data source is informative for current studies based on record linkage. With return migration causing non-randomness of migrant samples, the use of longitudinal data has become the gold standard in the migration literature. As Bandiera et al. (2013) show, return migration was particularly prominent in the early 1900s. A wealth of studies aiming at understanding migrants performance in that period having resorted to the creation of longitudinal datasets through linking individuals by name, age and birthplace. For example, Abramitzky et al. (2012b) estimate the return to migration by matching Norway-to-US migrants with their brothers who stayed in Norway in the late nineteenth century. Abramitzk et al. (2012a) studies immigrant assimilation by linking migrants and natives in the Censuses over time. Long and Ferrie (2013) link fathers and sons in 1850 to study occupational mobility in the U.S. and Britain. However, a name change in the U.S. - especially in the studies focusing on immigrants - might be a cause for a failed match. We have shown in our analysis that name Americanization was more prevalent among migrants facing stronger barriers in the labor market, yet occupational upgrading followed name Americanization. While these opposing features of the Americanization process - negative selection and occupational payoff - do not suggest clear implications for estimates based on matched samples, they are indeed informative in showing that immigrants who Americanized their names are not similar to

[22] In the online Appendix, we run further checks to understand the sensitivity of our analysis to coding errors. For instance, we drop the flagged occupations, we exclude the addresses that we imputed and we further explore the sensitivity of the benchmark results by changing the definition of local labor markets. None of these modifications changes our conclusions.

name keepers in terms of both observable and unobservable traits.[23] Hence, particular care should be taken when interpreting the representativeness of linked samples.

At a broader level, our results highlight the existence of a tradeoff between maintaining individual identity and enhancing labor market success, suggesting that cultural assimilation was instrumental to economic assimilation. In a historical perspective, this implies that despite migrant occupational upgrading being limited (Abramitzky et al., 2012a), migrants adopted alternative strategies to climb the occupational ladder. As one example of such a strategy, we find that returns to name Americanization were quite high. While the simple OLS estimates suggest potential explanations for the association between name Americanization and earnings to be changes in the market valuation of migrants' skill endowment, reduction in discrimination, or more rapid human capital accumulation, the first-difference and IV estimates rule out many of these reasons, with the exception of discrimination. The consequences of this are twofold. First, low occupational mobility observed in previous studies might have been caused by different attitudes and discrimination levels towards different ethnic groups. Second, from a broader perspective, the implied tradeoff between identity and labor market success suggested in several recent analyses (e.g. Bertrand and Mullainathan, 2004, Fryer and Levitt, 2004, Arai and Thoursie, 2009, Algan et al., 2012) seems to have been present since the early making of modern America.

[23] Surname could also matter. We have mentioned that surname Americanization occurred only in 7% of our sample. However, surname changes occur in 40% of the sample, while surname changes based on a Soundex ransliteration of the surname in 30% of the sample.

REFERENCES

Abramitzky, R., Boustan, L. and Eriksson K. (2012) "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Working Paper 18011,* NBER.

Abramitzky, R., Boustan, L. and Eriksson, K. (2012) "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review,* 102(5), 1832–1856.

Akerlof, G.A. and Kranton, R.E. (2000) "Economics and Identity." *The Quarterly Journal of Economics,* 115(3), 715–753.

Algan, Y., Mayer, T. and Thoenig, M. (2012) "The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France." *Unpublished manuscript.*

Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996) "Identification of Causal Effects using Instrumental Variables." *Journal of the American Statistical Association,* 91(434), 444–455.

Arai, M. and Thoursie, P.S. (2009) "Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings." *Journal of Labor Economics,* 27(1), 127–147.

Aura, S. and Hess, G.D. (2010) "What's in a Name?" *Economic Inquiry,* 48(1), 214–227.

Bandiera, O., Rasul, I. and Viarengo, M. (2013) "The Making of Modern America: Migratory Flows in the Age of Mass Migration." *Journal of Development Economics,* 102, 23–47.

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

Borjas, G.J. (2000) "Ethnic Enclaves and Assimilation." *Swedish Economic Policy Review,* 7(2), 89–122.

Brenner, D. and Rubinstein, Y. (2013) "Pride and Prejudice: Using Ethnic-Sounding Names and Inter-Ethnic Marriages to Identify Labor Market Discrimination." *The Review of Economic Studies* (forthcoming).

Broom, L., Beem, H.P. and Harris, V. (1955) "Characteristics of 1,107 Petitioners for Change of Name." *American Sociological Review,* 20(1), 33–39.

Drury, D.W. and McCarthy, J.D. (1980) "The Social Psychology of Name Change: Reflections on a Serendipitous Discovery." *Social Psychology Quarterly,* 310–320.

Ferrie, J.P. (1996) "A New Sample of Males Linked from the Public Use Microdata Sample of the 1850 US Federal Census of Population to the 1860 US Federal Census Manuscript Schedules." *Historical Methods: A Journal of Quantitative and Interdisciplinary History,* 29(4), 141–156.

Ferrie, J.P. (1999) "Yankeys Now: Immigrants in the Antebellum U.S. 1840-1860." *Oxford University Press.*

Figlio, D.N. (2005) "Names, Expectations and the Black-White Test Score Gap." *Working Paper 11195,* National Bureau of Economic Research.

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

Goldin, C. and Shim, M. (2004) "Making a Name: Women's Surnames at Marriage and Beyond." *The Journal of Economic Perspectives,* 18(2), 143–160.

Higham, J. (2002) "Strangers in the Land: Patterns of American Nativism, 1860-1925." *Rutgers University Press.*

Kang, T.S. (1971) "Name Change and Acculturation: Chinese Students on an American Campus." *The Pacific Sociological Review,* 14(4), 403–412.

Laham, S.M., Koval, P. and Alter, A.L. (2011). "The Name-Pronunciation Effect: Why people like Mr. Smith more than Mr. Colquhoun." *Journal of Experimental Social Psychology*.

Lieberson, S. (2000) "A Matter of Taste: How Names, Fashions, and Culture Change." *Yale University Press.*

Lleras-Muney, A. and Shertzer, A. (2012). "Did the Americanization Movement Succeed? An Evaluation of the Effect of English-Only and Compulsory Schools Laws on Immigrants' Education." *Working Paper 18302,* National Bureau of Economic Research.

Long, J. and Ferrie, J. (2013) "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." *American Economic Review,* 103(4), 1109–37.

Moser, P. (2012) "Taste-Based Discrimination: Evidence from a Shift in Ethnic Preferences after WWI." *Explorations in Economic History,* 49(2), 167–188.

Steven Ruggles, J., Trent, A., Genadek, K., Goeken, R., Schroeder, M.B., and Sobek, M. (2010) "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]." *Technical report,* Minneapolis: University of Minnesota.

Wooldridge, J.M. (1995) "Score Diagnostics for Linear Models Estimated by Two Stage Least Squares." In Maddala, G.S., Phillips, P.C.B., and Srinivasan, T.N. (Eds.), *Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C.R. Rao,* Chapter 3, pp. 66–87. Oxford: Blackwell.

ESRC Centre for Population Change
Building 58, Room 2001
Faculty of Social and Human Sciences
University of Southampton
SO17 1BJ

T: +44 (0)2380 592579
E: cpc@soton.ac.uk
www.cpc.ac.uk

To subscribe to the CPC newsletter and keep up-to-date with research activity, news and events, please register online: www.cpc.ac.uk/newsletter

For our latest research updates you can also follow CPC on Twitter, Facebook and Mendeley:

www.facebook.com/CPCpopulation

www.twitter.com/CPCpopulation

www.mendeley.com/groups/3241781/centre-for-population-change

The ESRC Centre for Population Change (CPC) is a joint initiative between the University of Southampton and a consortium of Scottish universities including St Andrews, Edinburgh, Stirling and Strathclyde, in partnership with the Office for National Statistics and National Records of Scotland.