THE ECONOMIC PAYOFF OF NAME AMERICANIZATION

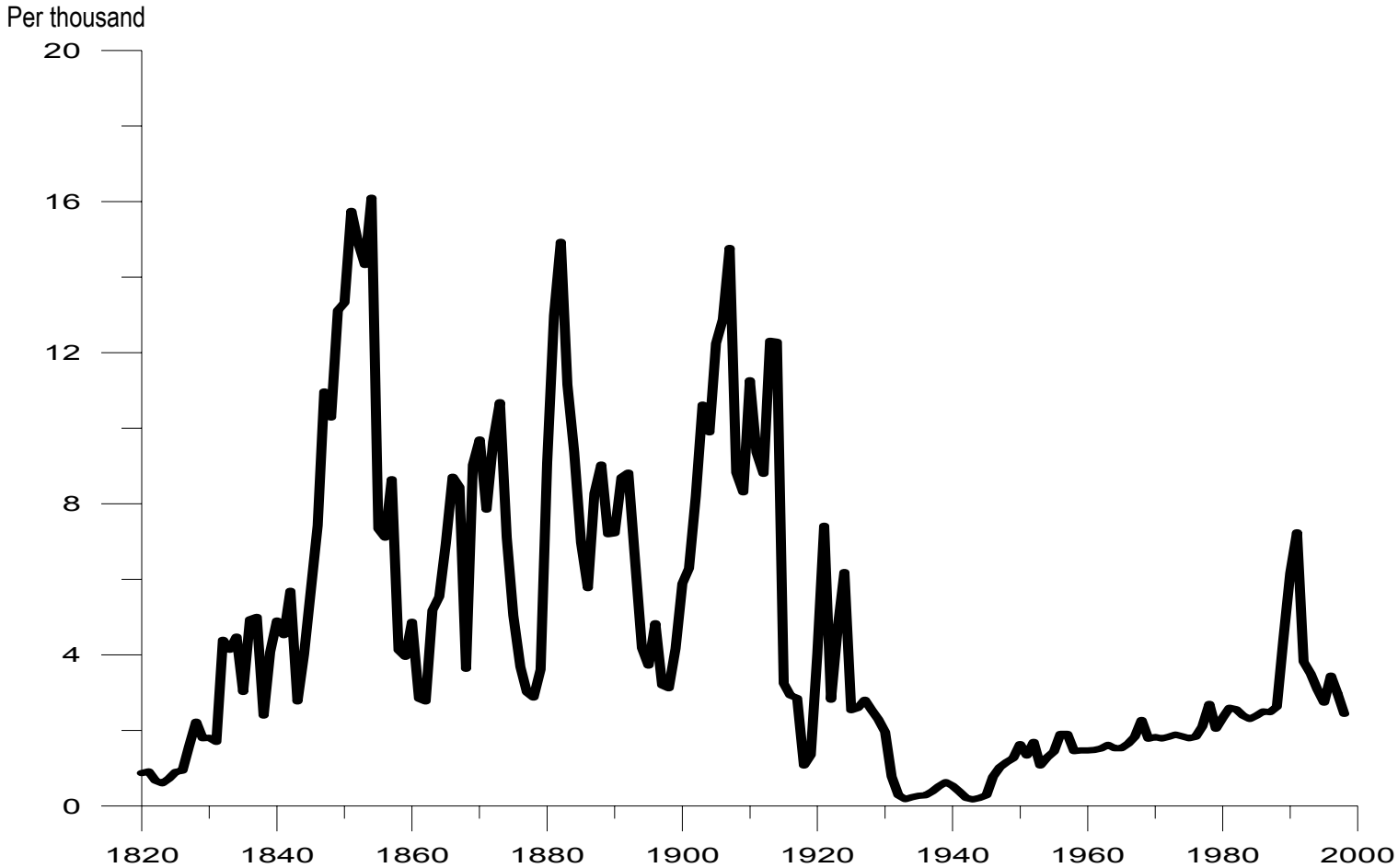
Corrado Giulietti - IZA

Costanza Biavaschi - University of Reading and IZA
Zahra Siddique - University of Reading and IZA

Knomad - WB
May 20, 2015

On migration and development

Immigration flows to the US:1820-2000



Context

Migrants often adopt native-sounding names

Context

Migrants often adopt native-sounding names

Rodolfo Alfonso Raffaello
Pierre Filibert Guglielmi
di Valentina d'Antonguolla

Arrived at Ellis Island
on December 23, 1913

Paolo Berretta

Arrived at Ellis Island
on April 28, 1921

Context

Migrants often adopt native-sounding names

Rodolfo Alfonso Raffaello
Pierre Filibert Guglielmi
di Valentina d'Antonguolla

Arrived at Ellis Island
on December 23, 1913

Rudolph Valentino

Paolo Berretta

Arrived at Ellis Island
on April 28, 1921

Paul Berretta

Research Question

Did the Americanization of male migrants' *first* names pay in the early 1900s?

Despite anecdotal evidence that name Americanization was common in the 1900s, there is no evidence up to date on its:

- Magnitude
- Consequences
- Causes (we will say more in the next paper, stay tuned!)

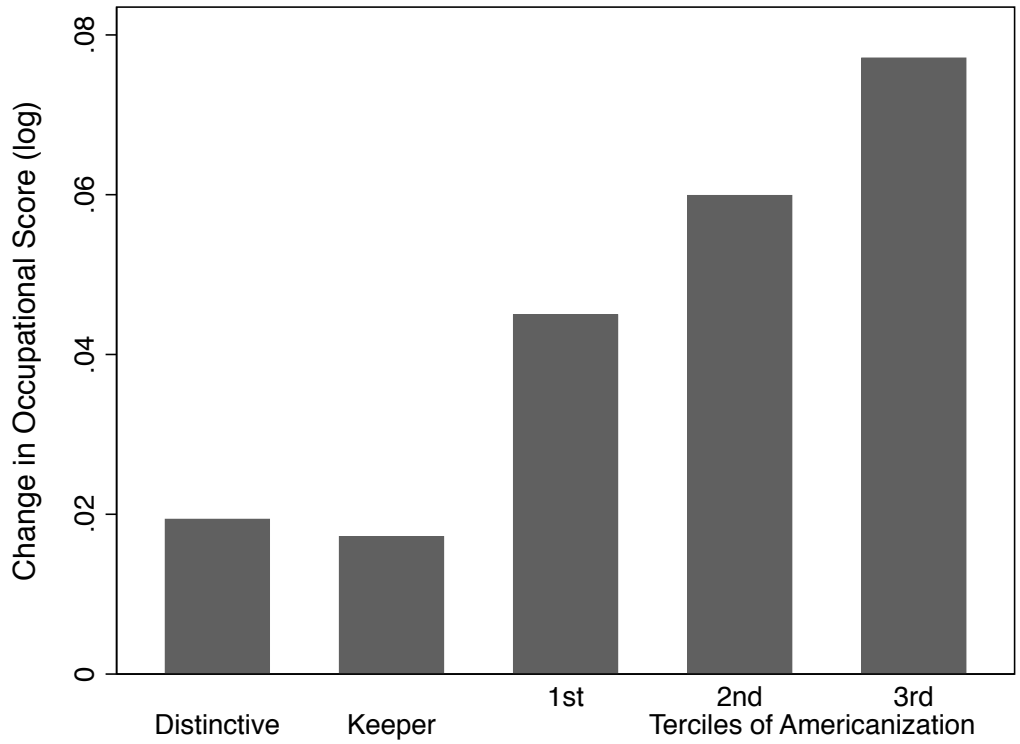
This Paper: Magnitude

A) name Americanization			B) Popular American Names		
Country of Origin	% Americanized	N	Name	% in U.S.-Born Population	% in Americanized Migrants
Italy	19.86	886	John	6.82	8.10
Russian Empire	57.41	749	William	5.74	2.49
Central Europe (excl. DE)	53.06	686	Joseph	3.91	6.93
Southern Europe (excl. IT)	37.69	130	Charles	3.69	2.57
Germany	24.71	437	George	3.64	2.73
Ireland	1.33	376	...	...	...
U.K.	5.00	160	Patrick	0.25	0.08
Northern Europe	19.00	279	...	...	...
Americas	10.89	202	Moishe	0.00	0.00
Other	38.76	178	...	...	...
Total	31.45	4083	Giulio	0.00	0.00

Panel A): Own tabulations from sample of naturalizing immigrants in 1930 New York City. Source: Ancestry.com. Name Americanization is defined as the custom of adopting a first name that was more frequent in the U.S.-born population than the migrant's name at arrival.
Panel B): Own tabulations from IPUMS Census, 1930 and from sample of naturalizing immigrants in 1930 New York City. % in U.S. Born Population indicates the percentage of the US-born male population in 1930 New York having a specific name. % in Americanized Migrants indicates, among those who Americanized their names, the percentage who chose a specific name.

This Paper: Consequences

Key pattern: name Americanization was associated with an increase in occupation-based earnings.



Much of this presentation dedicated to showing that this payoff is not (only) driven by individual self-selection.

Contributions

Provide empirical evidence on a well-known but under-researched phenomenon.

Create first historical longitudinal data where the panel component is not achieved through matching of Census enumerations (standard matching rates of 15-25%). Possibility to understand what has been missed from record linkage (Name Americanization is a source of panel attrition for these studies)

Build upon the literature on the "economics of names" (e.g., Fryer and Levitt, 2004; Bertrand and Mullainathan, 2004) by investigating name as a choice variable, which highlights the tradeoffs between economic and cultural assimilation.

Focus on an important period of American history. The end of Mass Migration had set the basis for the modern "melting pot": 30% Americans are living descendants of this flow.

.

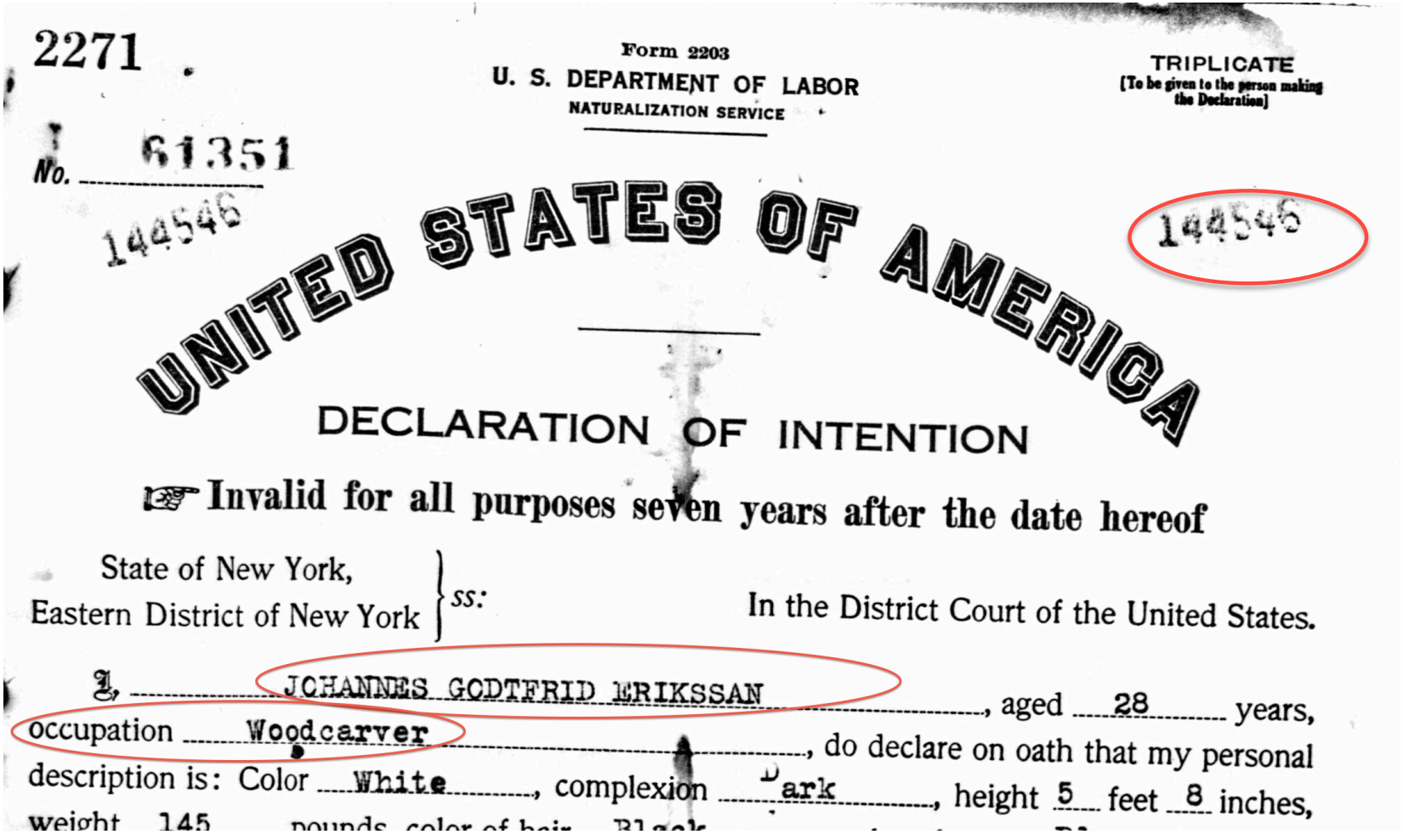
Data

We take a 20% random sample from the universe of naturalization records filed in the U.S. District Court for the Southern and Eastern District of New York in 1930. We exploit the two steps of the naturalization procedure (declaration of intention and - at least 5 years later - petition for citizenship).

No matching: we observe over time 100% of the original random sample and have information on a wide range of characteristics (internal mobility, family location, physical traits).

Cost: We only observe naturalized immigrants. But all "Free white persons, aliens of African nativity and persons of African descent of good moral character" could naturalize. External validity can be checked with Census.

How our data look like



2271

Form 2203
U. S. DEPARTMENT OF LABOR
NATURALIZATION SERVICE

TRIPLICATE
(To be given to the person making the Declaration)

No. 61351
144546

144546

UNITED STATES OF AMERICA

DECLARATION OF INTENTION

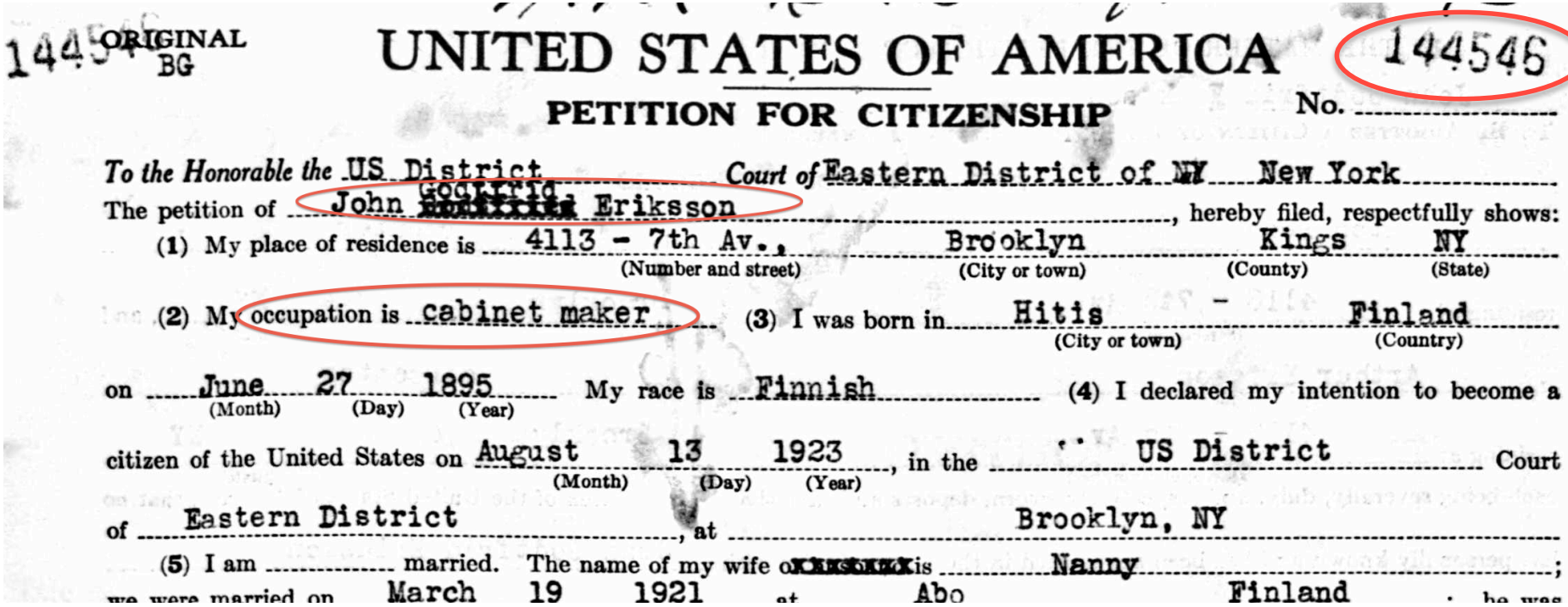
☞ Invalid for all purposes seven years after the date hereof

State of New York,
Eastern District of New York } ss:

In the District Court of the United States.

I, JOHANNES GODTFRID ERIKSSAN, aged 28 years,
occupation Woodcarver, do declare on oath that my personal
description is: Color White, complexion Dark, height 5 feet 8 inches,
weight 145 pounds, color of hair Black

How our data look like



1445 ORIGINAL BG

UNITED STATES OF AMERICA 144546

PETITION FOR CITIZENSHIP No. _____

To the Honorable the US District *Court of* Eastern District of NY New York

The petition of John Godfrid Eriksson, hereby filed, respectfully shows:

(1) My place of residence is 4113 - 7th Av., (Number and street) Brooklyn (City or town) Kings (County) NY (State)

(2) My occupation is cabinet maker (3) I was born in Hitis (City or town) Finland (Country)

on June (Month) 27 (Day) 1895 (Year) My race is Finnish (4) I declared my intention to become a

citizen of the United States on August (Month) 13 (Day) 1923 (Year), in the US District Court

of Eastern District, at Brooklyn, NY

(5) I am _____ married. The name of my wife o[redacted] is Nanny;

we were married on March 19 1921 at Abo Finland . he was

Definitions

Before we proceed:

- How do we define earnings? We rely on income scores indicating the median yearly income in hundreds of 1950 dollars of a person in a given occupation.
- How to measure name Americanization? We assign to each migrant i at time t the frequency that his name has in the American-born population living in the state of NY. We normalize the index to vary between 0 and 1.

Definitions: How to measure name Americanization?

$$A_{it} = \frac{\sum_k I(Name_k = i)}{\max_j \sum_k I(Name_k = j)} \qquad \text{for } k \text{ in US-born, NY}$$

Examples:

- John: $A_{it} = 1$ (most common name among the natives).
- Giorgio: $A_{it} = 0$ (no natives called Giorgio in the native born population).
- Giorgio $\rightarrow$ John $A_{i0} = 0$ and $A_{i1} = 1$
- John $\rightarrow$ John $A_{i0} = 1$ and $A_{i1} = 1$
- Giorgio $\rightarrow$ Salvatore $A_{i0} = 0$ and $A_{i1} = 0$

Some Examples from the Data

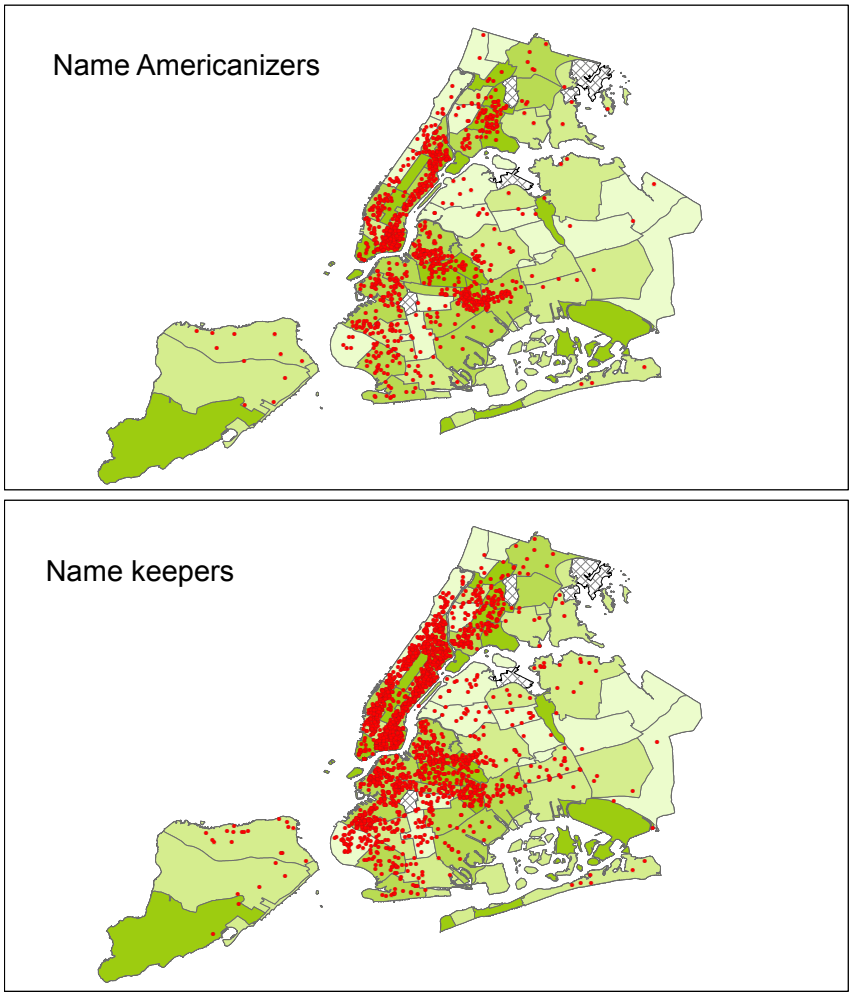
	Name at Arrival	Normalized Frequency of Name at Arrival	Name at Petition	Normalized Frequency of Name at Petition
Into popular names:	Ivan	0.0038334	John	1
	Peter	0.1058411	John	1
	Domenico	0.0001949	Charles	0.5401209
	Dimitrios	0	James	0.4496784
Into popular (but less common) names:	Ciro	0.0007147	Louis	0.1765317
	Moische	0.0552271	Morris	0.0562017
Into distinctive names:	Alexander	0.0395686	Jakov	0
	Aaron	0.0135794	Elias	0.0041583

Characteristics at Declaration

Variable	All	Distinctive	Keepers	Americanize	1st	2nd	3rd
Log Occupational Score	3.1771	3.1662	3.1722	3.1886	3.1752	3.2195	3.1723
	(0.4360)	(0.4237)	(0.4026)	(0.4989)	(0.5099)	(0.4945)	(0.4915)
Age	30.6166	29.7679	30.9001	30.1486	30.3395	29.4988	30.5919
	(8.6947)	(8.2484)	(8.6202)	(8.8815)	(9.2792)	(8.5533)	(8.7626)
Years Since Migration	7.0439	6.9415	5.9520	9.2897	8.8219	9.5502	9.5140
	(7.2867)	(6.6551)	(6.8589)	(7.6963)	(7.7498)	(7.7777)	(7.5560)
Married	0.4747	0.4503	0.4684	0.4907	0.4749	0.4809	0.5164
	(0.4994)	(0.4990)	(0.4991)	(0.5001)	(0.4999)	(0.5002)	(0.5003)
Has U.S.-Born Spouse	0.0509	0.0409	0.0441	0.0662	0.0320	0.0933	0.0748
	(0.2199)	(0.1987)	(0.2054)	(0.2487)	(0.1761)	(0.2912)	(0.2633)
Number of Children	0.9336	0.8830	0.9304	0.9470	1.0023	0.8541	0.9813
	(1.5351)	(1.6264)	(1.5462)	(1.5004)	(1.5757)	(1.3800)	(1.5333)
Has U.S.-Born Child(ren)	0.2143	0.1930	0.1925	0.2617	0.2329	0.2679	0.2850
	(0.4104)	(0.3958)	(0.3944)	(0.4397)	(0.4231)	(0.4434)	(0.4520)
Moved into N.Y.	0.1129	0.0936	0.1290	0.0826	0.0982	0.0670	0.0818
	(0.3165)	(0.2921)	(0.3353)	(0.2753)	(0.2979)	(0.2503)	(0.2743)
Arrival Cohort	1917	1918	1918	1916	1917	1916	1916
	(42.9888)	(6.1772)	(53.3214)	(7.1538)	(7.1106)	(7.3143)	(7.0373)
N	4083	171	2628	1284	438	418	428

Standard deviations in parentheses. All characteristics are measured at the time of Declaration in the first two column and as difference between petition and declaration in the last three columns. Keepers are migrants whose Americanization index has not changed. YSM = years since migration.

Local labor markets and name Americanization at Declaration



Change in Characteristics over Time

Variable	All	Distinctive	Keepers	Americanize	1st	2nd	3rd
Log Occupational Score	0.0309	0.0194	0.0172	0.0605	0.0450	0.0599	0.0771
	(0.4777)	(0.4393)	(0.4392)	(0.5516)	(0.6524)	(0.4906)	(0.4929)
Age	5.3150	5.3772	5.4019	5.1284	5.2424	4.9903	5.1473
	(2.5585)	(2.0873)	(2.7462)	(2.1807)	(2.3376)	(2.2119)	(1.9683)
Years Since Migration	4.7017	4.8830	4.7610	4.5561	4.6438	4.3684	4.6495
	(1.7162)	(1.8209)	(1.6890)	(1.7486)	(1.8016)	(1.7113)	(1.7188)
Married	0.2194	0.2749	0.2234	0.2040	0.2100	0.2201	0.1822
	(0.4139)	(0.4478)	(0.4166)	(0.4032)	(0.4078)	(0.4148)	(0.3865)
Has U.S.-Born Spouse	0.0561	0.0936	0.0487	0.0662	0.0776	0.0598	0.0607
	(0.2301)	(0.2921)	(0.2153)	(0.2487)	(0.2679)	(0.2374)	(0.2391)
Number of Children	0.2976	0.2982	0.3021	0.2882	0.2877	0.2751	0.3014
	(0.6149)	(0.5930)	(0.6310)	(0.5839)	(0.6046)	(0.5168)	(0.6237)
Has U.S.-Born Child(ren)	0.1548	0.1988	0.1488	0.1612	0.1667	0.1794	0.1379
	(0.3764)	(0.4003)	(0.3716)	(0.3824)	(0.3792)	(0.3842)	(0.3837)
Moved into N.Y.	-0.1007	-0.0877	-0.1145	-0.074	-0.0936	-0.0622	-0.0654
	(0.3334)	(0.3037)	(0.3558)	(0.2847)	(0.3069)	(0.2515)	(0.2910)
N	4083	171	2628	1284	438	418	428

Standard deviations in parentheses. Distinctive refers to migrants for which $\Delta Ai < 0$; Keepers are migrants for which $\Delta Ai = 0$; Americanize refers to migrants for which $\Delta Ai > 0$; quartiles refer to migrants who americanize.

Empirical Model

Log-Occupational Scores y_{it} of individual i at time t are a function of name Americanization A_{it}, i.e. how common a migrant's name is among U.S.-born individuals:

$$y_{it} = \beta_0 + \beta_1 A_{it} + \tau_i + \epsilon_{it}.$$

Empirical Model

Log-Occupational Scores y_{it} of individual i at time t are a function of name Americanization A_{it}, i.e. how common a migrant's name is among U.S.-born individuals:

$$y_{it} = \beta_0 + \beta_1 A_{it} + \beta_2 T_{it} + \boldsymbol{x}'_{it}\gamma + \tau_i + \epsilon_{it}.$$

$\boldsymbol{x}'_{it}$: time varying socio-economic variables, such as marital status, spouse being U.S.-born, number of children, having U.S.-born children, whether the individual has moved to the state of New York from other states over time.

Empirical Model

Log-Occupational Scores y_{it} of individual i at time t are a function of name Americanization A_{it}, i.e. how common a migrant's name is among U.S.-born individuals:

$$y_{it} = \beta_0 + \beta_1 A_{it} + \beta_2 T_{it} + \boldsymbol{x}'_{it}\gamma + \beta_3 (T_{it} * COB_i) + \\ + \qquad\qquad + \tau_i + \epsilon_{it}.$$

$\boldsymbol{x}'_{it}$: time varying socio-economic variables, such as marital status, spouse being U.S.-born, number of children, having U.S.-born children, whether the individual has moved to the state of New York from other states over time.

$(T_{it} * COB_i)$: nationality-specific time trends

Empirical Model

Log-Occupational Scores y_{it} of individual i at time t are a function of name Americanization A_{it}, i.e. how common a migrant's name is among U.S.-born individuals:

$$y_{it} = \beta_0 + \beta_1 A_{it} + \beta_2 T_{it} + \boldsymbol{x}'_{it}\gamma + \beta_3 (T_{it} * COB_i) + \\ + \beta_4 (T_{it} * LabMkt_i) + \qquad\qquad + \tau_i + \epsilon_{it}.$$

$\boldsymbol{x}'_{it}$: time varying socio-economic variables, such as marital status, spouse being U.S.-born, number of children, having U.S.-born children, whether the individual has moved to the state of New York from other states over time.

$(T_{it} * COB_i)$: nationality-specific time trends

$(T_{it} * LabMkt_i)$: local labor market-specific time trends.

$(T_{it} * Cohort_i)$: arrival cohort-specific time trends.

Empirical Model

Log-Occupational Scores y_{it} of individual i at time t are a function of name Americanization A_{it}, i.e. how common a migrant's name is among U.S.-born individuals:

$$y_{it} = \beta_0 + \beta_1 A_{it} + \beta_2 T_{it} + \boldsymbol{x}'_{it}\gamma + \beta_3(T_{it} * COB_i) + \\ + \beta_4(T_{it} * LabMkt_i) + \beta_5(T_{it} * Cohort_i) + \tau_i + \epsilon_{it}.$$

$\boldsymbol{x}'_{it}$: time varying socio-economic variables, such as marital status, spouse being U.S.-born, number of children, having U.S.-born children, whether the individual has moved to the state of New York from other states over time.

$(T_{it} * COB_i)$: nationality-specific time trends

$(T_{it} * LabMkt_i)$: local labor market-specific time trends.

$(T_{it} * Cohort_i)$: arrival cohort-specific time trends.

τ_i: time-invariant characteristics (completed education, ability, personal traits, ...)

OLS and First Differences results

	OLS			First Differences		
	I	II	III	IV	V	VI
A_{it}	0.0357**	0.0292	0.0328*	0.0952**	0.1061***	0.1143***
	(0.0170)	(0.0178)	(0.0177)	(0.0379)	(0.0381)	(0.0391)
Married	0.0415***	0.0276**	0.0258**	–0.0130	–0.0114	–0.0183
	(0.0119)	(0.0123)	(0.0122)	(0.0200)	(0.0207)	(0.0209)
Has U.S.-born Spouse	0.0404***	0.0550***	0.0466***	0.0592	0.0588	0.0688*
	(0.0154)	(0.0156)	(0.0156)	(0.0400)	(0.0402)	(0.0409)
Number of Children	–0.0124***	–0.0075*	–0.0084**	–0.0110	–0.0125	–0.0101
	(0.0039)	(0.0041)	(0.0041)	(0.0098)	(0.0104)	(0.0102)
Has U.S.-born Child(ren)	0.0344***	0.0358***	0.0348***	0.0318*	0.0365*	0.0365*
	(0.0122)	(0.0123)	(0.0124)	(0.0193)	(0.0194)	(0.0201)
Move into N.Y.	0.0224	0.0234	0.0081	–0.0008	0.0012	0.0031
	(0.0166)	(0.0167)	(0.0218)	(0.0248)	(0.0250)	(0.0957)
Arrival Cohort	–0.1285**	–0.0577	–0.0795	0.0069	0.0055	0.0051
	(0.0636)	(0.0685)	(0.0712)	(0.0046)	(0.0047)	(0.0054)
$\Delta Trend_{it}$	Yes	Yes	Yes	Yes	Yes	Yes
Lab. Mkt. Ind.	No	No	Yes	No	No	Yes
R^2	0.01	0.02	0.05	0.01	0.01	0.03
N	8166	8166	8166	4083	4083	4083
Pred. Occ. Score whole sample	3.193	3.193	3.193	0.028	0.029	0.028
Pred. Occ. Score americanizers				0.045	0.046	0.046

Robust standard errors in parenthesis. A_{it} = Americanization index, which varies between 0 and 1, with 0 representing names with the lowest frequency and 1 representing names with the highest frequency. Labor market indicators refer to dummy variables for each of the NYC community districts. All regressions include a time trend. In first difference models, arrival cohort, country of birth and labor market indicators should be interpreted as interactions between these variables and the time trend.

Is selection driving the results?

How to get around the possibility that those with (time-varying) ambition/ability were more likely to americanize their names?

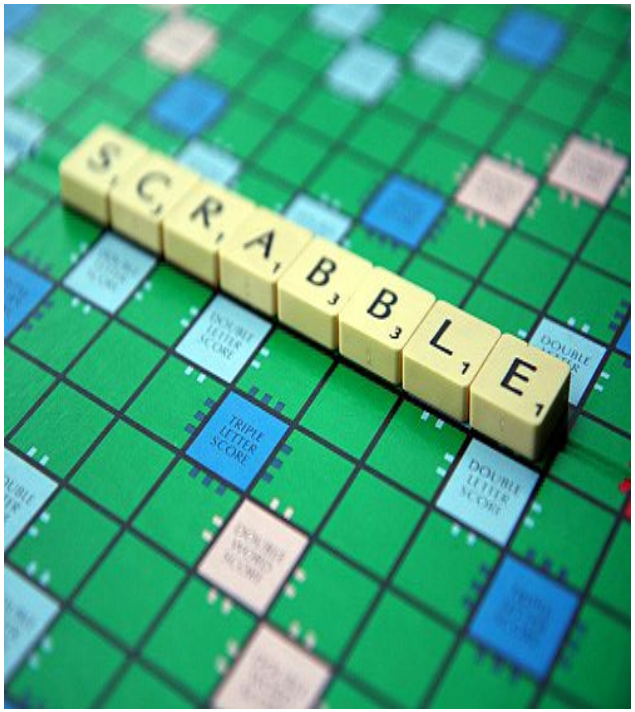
Two strategies:

- **Name Changers only**. Abstract from selection into name Americanization and keep only the name Americanizers. 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.

- **Instrumental variable**.

Instrumental Variable (IV)

Use Scrabble scores: worked out by Alfred Moscher Butts by performing a frequency analysis of letters appearing in the front page of several newspapers in the 1930s.



Scrabble points capture the linguistic structure of a name, measuring both the length of it and how uncommon its letters are. They provide a measure encapsulating the graphemic and phonemic features of names. Measure of distance between the scrabble points of the migrant's name at arrival and the scrabble points associated with the "American norm".

$$S_{jArrival} = \frac{SP_{jArrival}}{\sum_{w \neq j} SP_w/(N-1)},$$

- Relevance: the Scrabble index predicts name Americanization.
- Exclusion: the Scrabble index does not make use of the semantic, etymology, ethnic origin, linguistic distance or pronunciation associated to names which might be correlated with higher earnings.

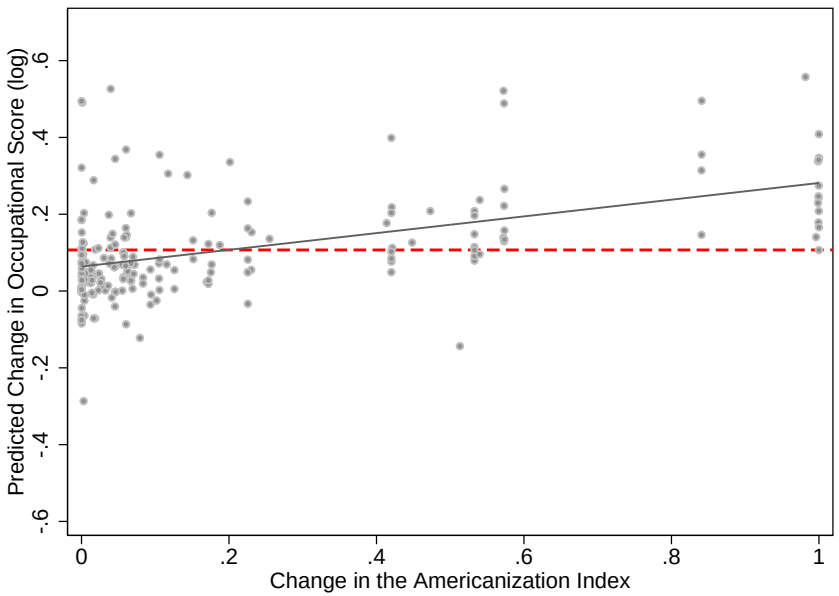
Main Results: Name Changers Only and IV

	Name changers only			Instrumental variable		
	I	II	III	IV	V	VI
A_{it}	0.2216*	0.2225*	0.2331**	0.5142**	0.4159*	0.4937**
	(0.1180)	(0.1160)	(0.1132)	(0.2249)	(0.2159)	(0.2229)
N	1738	1738	1738	4083	4083	4083
C. of birth Ind.	No	Yes	Yes	No	Yes	Yes
Lab. Mk. Ind.	No	No	Yes	No	No	Yes
				First stage		
$S_{jArrival}$				0.0586***	0.0629***	0.0611***
				(0.0060)	(0.0063)	(0.0064)
F 1^{st} stage				94.385	99.274	90.087
Partial R^2				0.022	0.024	0.023
Wooldridge test p-value				0.059	0.148	0.084
N				4083	4083	4083
Pred. Occ. Score whole sample	0.049	0.051	0.047	0.032	0.032	0.032
Pred. Occ Score americanizers	0.101	0.103	0.107	0.123	0.099	0.108

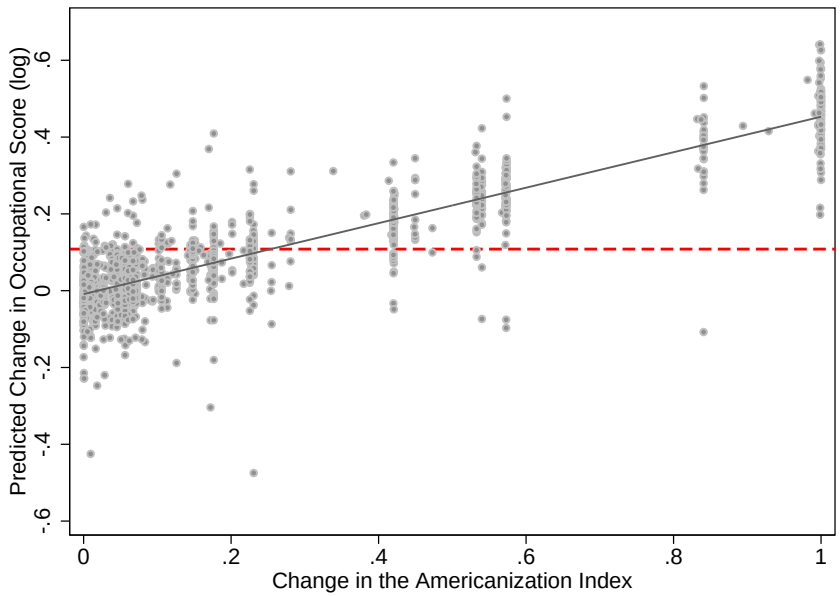
Robust standard errors in parenthesis.
A_{it} = Americanization index, which varies between 0 (names with the lowest frequency) and 1 (names with the highest frequency). $S_{jArrival}$ refers to the Scrabble index. See text for explanation.
Wooldrdige test refers to a robust score test of endogeneity (Wooldridge, 1995).
Arrival cohort, country of birth and labor market should be interpreted as interactions with the time trend.

How much did name Americanization pay?

ATT - Name Changers



ATT - IV



How much did name Americanization pay?

Name at Arrival	Name at Petition	Change in A_i	Change in Earnings
Average among name Americanizers (NC and IV)		0.2509	0.1083
		(0.3044)	(0.1652)
Average in the whole sample (NC and IV)		0.0766	0.0319
		(0.2096)	(0.1241)
Giovanni	John	0.9998	0.4595
Jan	John	0.9994	0.4392
Johann	John	1.0000	0.4591
Wilhelm	William	0.8408	0.3234
Giuseppe	Joseph	0.5722	0.2593
Francesco	Frank	0.4201	0.1719
Franz	Frank	0.4202	0.1615
Heinrich	Henry	0.2255	0.0824
Leib	Louis	0.1765	0.0843
Moische	Morris	0.0562	0.0373

Who gained the most?

	Old Migrants				New Migrants			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0025	0.0918*	0.1102	–0.1291	0.0562**	0.1254***	0.2533	0.6476**
	(0.0267)	(0.0556)	(0.0979)	(0.3513)	(0.0244)	(0.0469)	(0.1602)	(0.2802)
N	2504	1252	247	1252	5662	2831	1491	2831
	Tall				Short			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	–0.0179	0.0987	0.2490	0.3415	0.0520**	0.1141***	0.1866*	0.5704**
	(0.0376)	(0.0951)	(0.2788)	(0.4208)	(0.0205)	(0.0430)	(0.1027)	(0.2481)
N	2050	924	331	924	6116	3058	1368	3058
	High Exposure				Low Exposure			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	–0.0367	0.1872	–0.1715	–0.1125	0.0495**	0.0667*	0.1853**	0.6086**
	(0.0660)	(0.1379)	(0.3579)	(0.7241)	(0.0213)	(0.0365)	(0.0833)	(0.2616)
N	1224	612	264	612	5220	2610	1163	2610

Robust standard errors in parenthesis. All models include all the covariates.

A_{it} = Americanization index, which varies between 0 (names with the lowest frequency) and 1 (names with the highest frequency). See text for explanation.

Old Migrants refers to migrants from Germany, Ireland, U.K. and Northern Europe. New Migrants refers to migrants from Italy, Russian Empire, Central and Southern Europe, Americas and Other.

Tall refers to migrants with a height above the 3rd quartile of the height distribution. Height is reported only by 3,982 migrants.

Exposure is the ratio between the number of migrants from each country of birth and the population in the tract. High (Low) Exposure indicates whether the migrant in 1920 was living in a tract with exposure above (below) the 95% level of New York City. Exposure is available only for 3,222 migrants due to the size of the 1920 Census and since 542 migrants lived outside New York City at the time of declaration.

First Stage F-tests. Old Migrants: 20.645; New Migrants: 67.465; Tall: 22.213; Short: 64.971; High Exposure: 3.198; Low Exposure: 67.895;

Instrument Validity

Challenges to the instrument:

- Is the linguistic structure of the name directly associated with labor market outcomes, perhaps due to preferences of the employers or customers? We show that there is no association between labor market outcomes and the Scrabble index for groups of individuals who (arguably) have not Americanized their names: the US-born population.

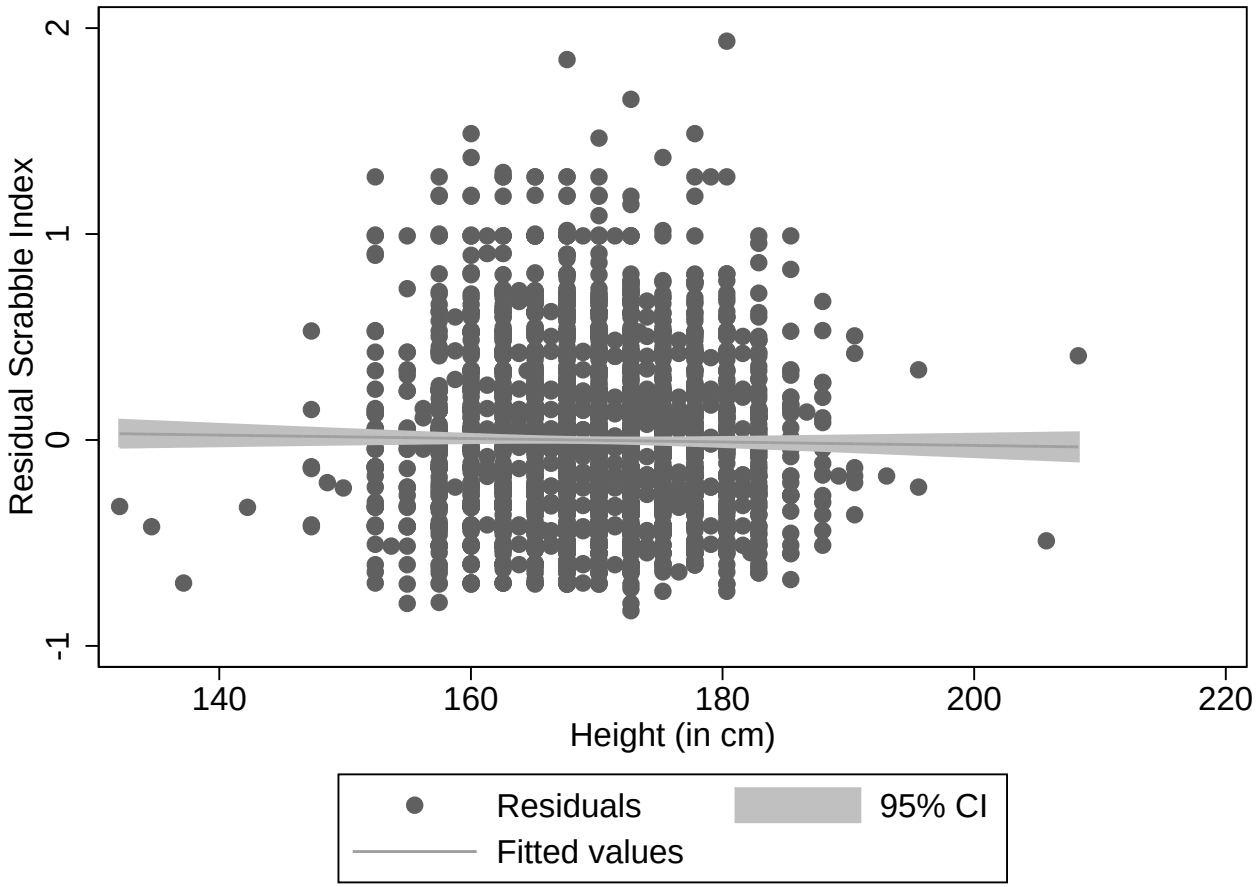
- Is the linguistic structure capturing unobserved migrants' traits that are directly correlated with wage growth? We show that the Scrabble index is uncorrelated with various measures of migrant socio-economic background (known to affect wage trajectories) within a country.

Scrabble is not associated to earnings for natives

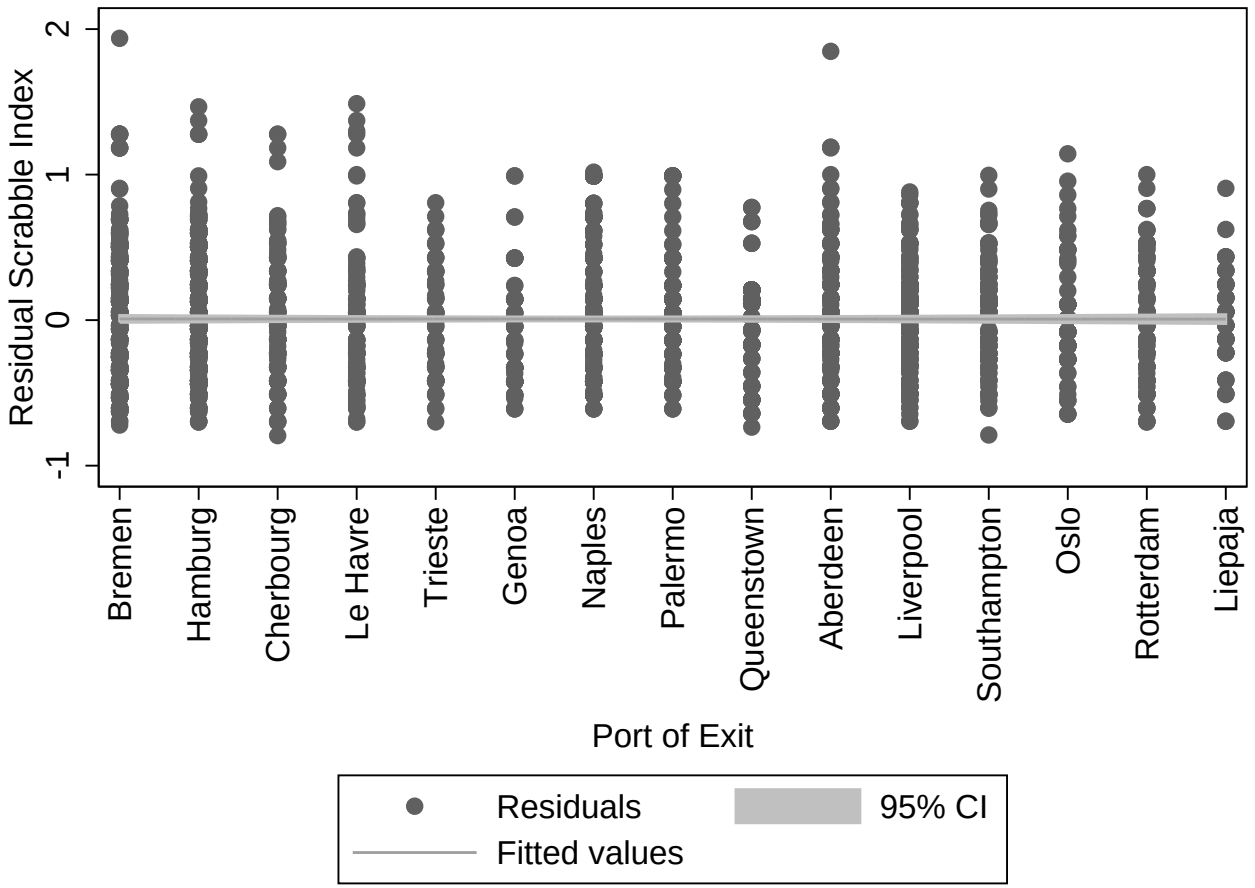
	Unconditional			
	I	II	III	IV
Americanization Index	0.0144***		0.0149***	0.0148***
	(0.0038)		(0.0040)	(0.0040)
Scrabble Index		0.0038	–0.0014	–0.0010
		(0.0037)	(0.0040)	(0.0040)
Reads and Writes				0.2386***
				(0.0167)
R^2	0.00	0.00	0.00	0.00
N	109803	109803	109803	109803
	Conditional on Characteristics			
	I	II	III	IV
Americanization Index	0.0122***		0.0121***	0.0120***
	(0.0037)		(0.0040)	(0.0040)
Scrabble Index		0.0044	0.0003	0.0006
		(0.0037)	(0.0039)	(0.0039)
Reads and Writes				0.2057***
				(0.0168)
R^2	0.03	0.03	0.03	0.03
N	109803	109803	109803	109803

Robust standard errors in parenthesis.
Source: 1930 Census.
Controls include age and indicators for state of birth and being white.

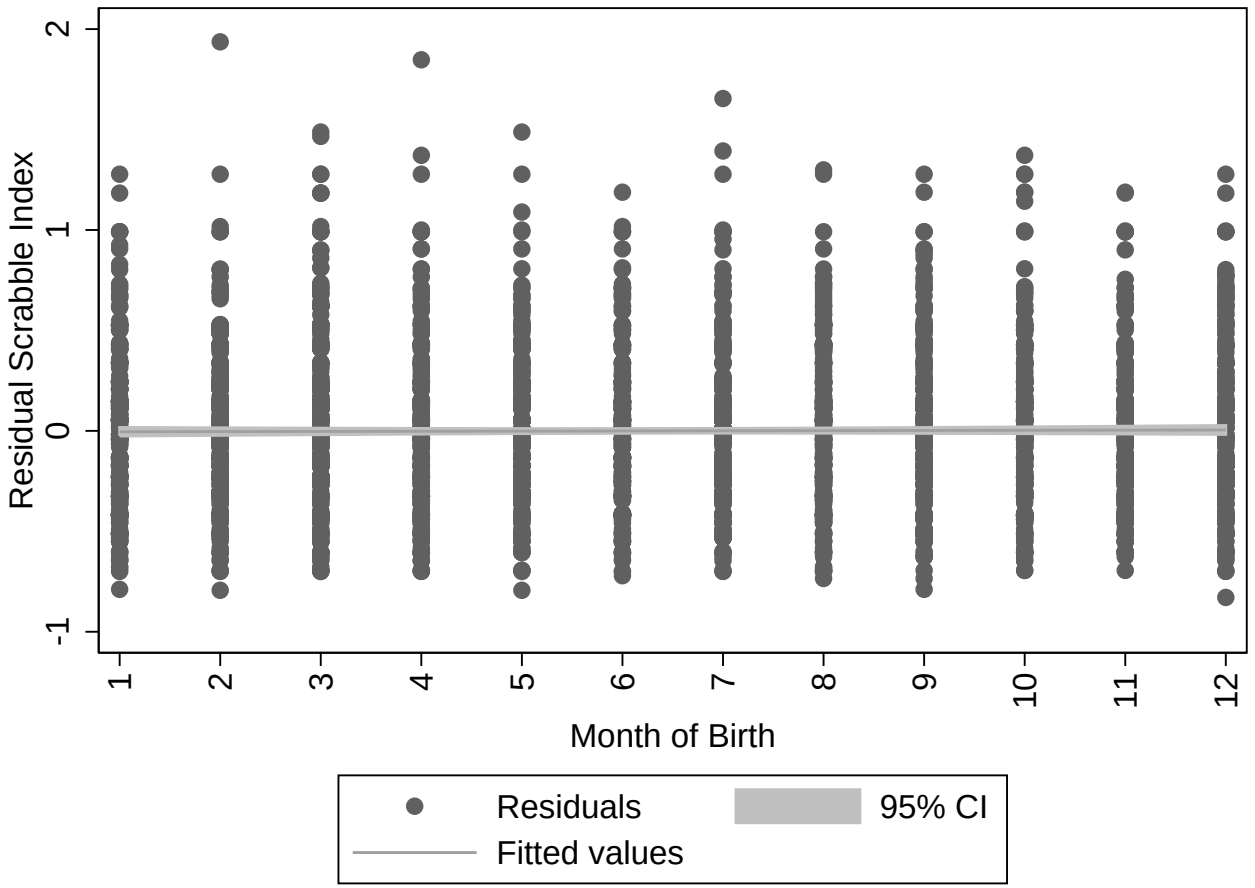
Scrabble does not correlate with socio-econ background: Height



Scrabble does not correlate with socio-econ background: Port of Emigration



Scrabble does not correlate with socio-econ background: Month of Birth



Conclusions

The rich information provided in the naturalization files allows us to study the effect of changing into a popular American name on migrants' earnings.

- Very common phenomenon: about 30% on average, with peaks of 55%.
- Name Americanization increased earnings. Effects are stronger for the European migrants coming from the sending regions of the second wave of Mass Migration and for migrants with lower socio-economic status.
- The direction of the bias suggests that economically unsuccessful migrants were more likely to Americanize their names.

Conclusions

When matching, less successful migrants will end up being dropped from the sample (but they will do better over time, so ambiguous prediction on this source of bias).

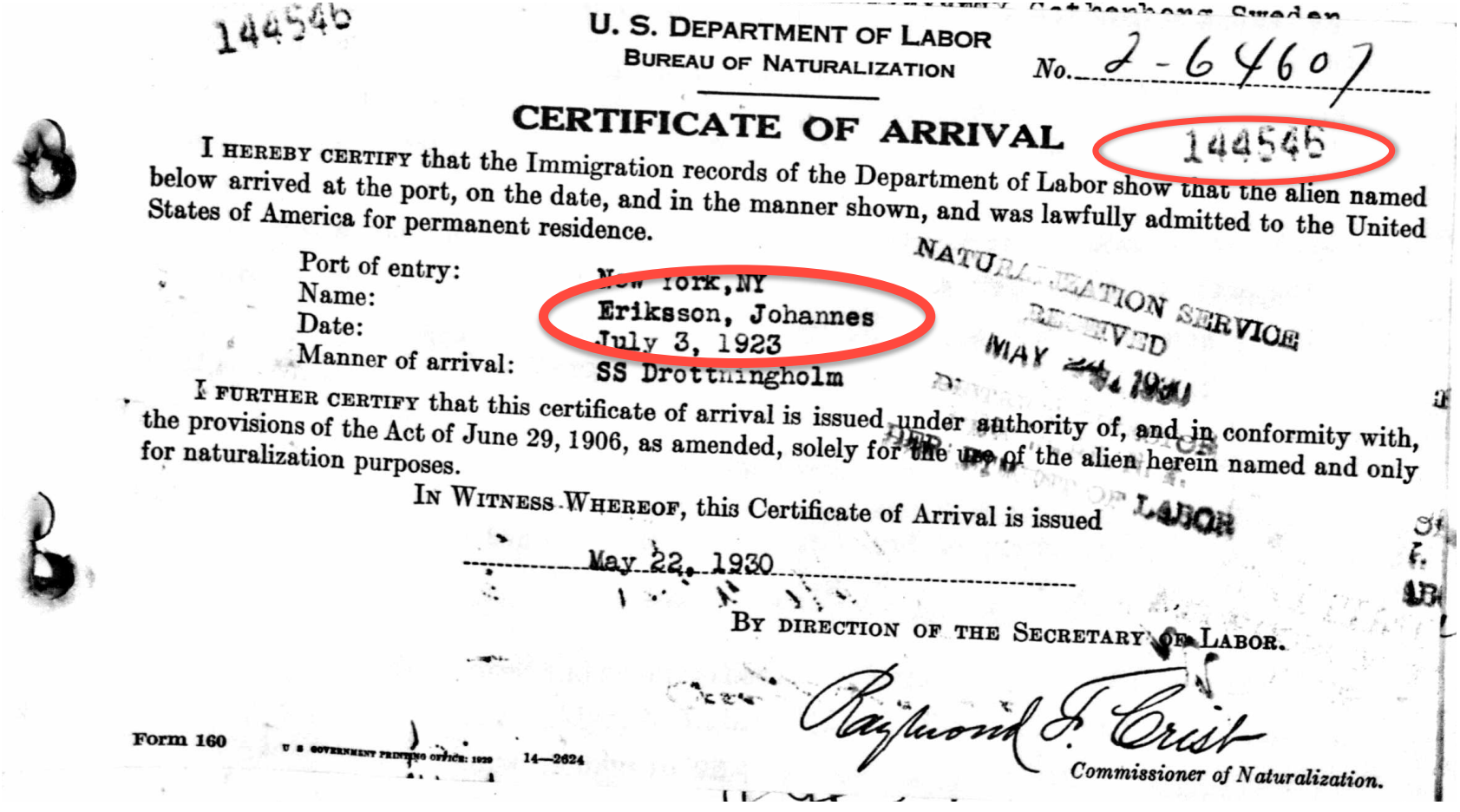
The results highlight the tradeoff between maintaining one's individual identity and labor market success, suggesting that the process of cultural assimilation at the dawn of the modern "melting pot" was instrumental for migrants' economic advancement.

Such 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 making of modern America.

THANKS!

APPENDIX

How our data look like



144546

U. S. DEPARTMENT OF LABOR
BUREAU OF NATURALIZATION

No. 2-64607

CERTIFICATE OF ARRIVAL

144545

I HEREBY CERTIFY that the Immigration records of the Department of Labor show that the alien named below arrived at the port, on the date, and in the manner shown, and was lawfully admitted to the United States of America for permanent residence.

Port of entry: New York, NY
Name: Eriksson, Johannes
Date: July 3, 1923
Manner of arrival: SS Drottningholm

NATURALIZATION SERVICE
RECEIVED
MAY 1930

I FURTHER CERTIFY that this certificate of arrival is issued under authority of, and in conformity with, the provisions of the Act of June 29, 1906, as amended, solely for the use of the alien herein named and only for naturalization purposes.

IN WITNESS WHEREOF, this Certificate of Arrival is issued

May 22, 1930

BY DIRECTION OF THE SECRETARY OF LABOR.

Raymond F. Crist
Commissioner of Naturalization.

Form 160 U S GOVERNMENT PRINTING OFFICE: 1929 14—2624

Further Results

Other checks:

- Full sample descriptives
- Timing of name change
- Reduced form
- More on channels
- Surnames
- Representativeness

Full Sample Descriptives | Types | Reduced form | Heterogeneity | Surnames | Representativeness

Descriptives, Full Sample

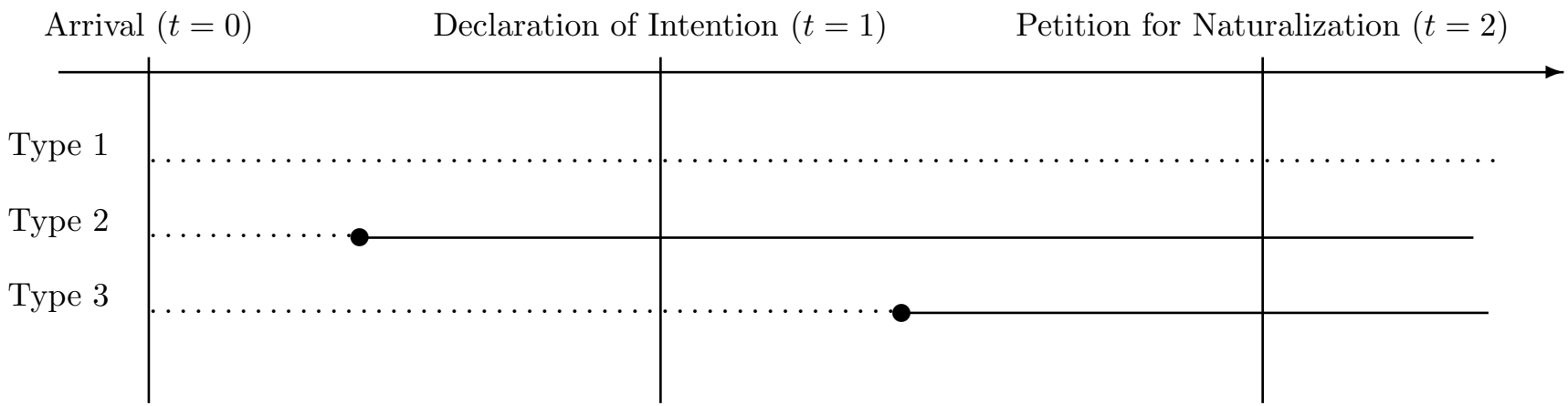
Variable	All	Distinctive	Keepers	Americanize	1st	2nd	3rd
				At Declaration			
Log Occupational Score	3.1771	3.1662	3.1722	3.1886	3.1752	3.2195	3.1723
	(0.4360)	(0.4237)	(0.4026)	(0.4989)	(0.5099)	(0.4945)	(0.4915)
Age	30.6166	29.7679	30.9001	30.1486	30.3395	29.4988	30.5919
	(8.6947)	(8.2484)	(8.6202)	(8.8815)	(9.2792)	(8.5533)	(8.7626)
Years Since Migration	7.0439	6.9415	5.9520	9.2897	8.8219	9.5502	9.5140
	(7.2867)	(6.6551)	(6.8589)	(7.6963)	(7.7498)	(7.7777)	(7.5560)
Married	0.4747	0.4503	0.4684	0.4907	0.4749	0.4809	0.5164
	(0.4994)	(0.4990)	(0.4991)	(0.5001)	(0.4999)	(0.5002)	(0.5003)
Has U.S.-Born Spouse	0.0509	0.0409	0.0441	0.0662	0.0320	0.0933	0.0748
	(0.2199)	(0.1987)	(0.2054)	(0.2487)	(0.1761)	(0.2912)	(0.2633)
Number of Children	0.9336	0.8830	0.9304	0.9470	1.0023	0.8541	0.9813
	(1.5351)	(1.6264)	(1.5462)	(1.5004)	(1.5757)	(1.3800)	(1.5333)
Has U.S.-Born Child(ren)	0.2143	0.1930	0.1925	0.2617	0.2329	0.2679	0.2850
	(0.4104)	(0.3958)	(0.3944)	(0.4397)	(0.4231)	(0.4434)	(0.4520)
Moved into N.Y.	0.1129	0.0936	0.1290	0.0826	0.0982	0.0670	0.0818
	(0.3165)	(0.2921)	(0.3353)	(0.2753)	(0.2979)	(0.2503)	(0.2743)
Arrival Cohort	1917.3	1918.2	1917.8	1916.2	1916.5	1916.1	1915.8
	(42.989)	(6.177)	(53.321)	(7.154)	(7.111)	(7.314)	(7.037)
Italy	0.2170	0.1287	0.2618	0.1371	0.0845	0.0909	0.2360
	(0.4123)	(0.3358)	(0.4397)	(0.3441)	(0.2784)	(0.2878)	(0.4251)
Russian Empire	0.1834	0.2865	0.1027	0.3349	0.4018	0.4043	0.1986
	(0.3871)	(0.4535)	(0.3037)	(0.4721)	(0.4908)	(0.4913)	(0.3994)
Central Europe (excl. DE)	0.1680	0.2924	0.1035	0.2835	0.3311	0.2536	0.2640
	(0.3739)	(0.4562)	(0.3047)	(0.4509)	(0.4711)	(0.4356)	(0.4413)
Southern Europe (excl. IT)	0.0318	0.0409	0.0282	0.0382	0.0297	0.0407	0.0444
	(0.1756)	(0.1987)	(0.1655)	(0.1917)	(0.1699)	(0.1978)	(0.2062)
Germany	0.1070	0.0819	0.1199	0.0841	0.0251	0.0981	0.1308
	(0.3092)	(0.2750)	(0.3249)	(0.2777)	(0.1567)	(0.2978)	(0.3376)
Ireland	0.0921	0.0234	0.1396	0.0039	0.0046	0.0048	0.0023
	(0.2892)	(0.1516)	(0.3467)	(0.0623)	(0.0675)	(0.0691)	(0.0483)
U.K.	0.0392	0.0234	0.0563	0.0062	0.0000	0.0120	0.0070
	(0.1941)	(0.1516)	(0.2306)	(0.0787)	(0.0001)	(0.1088)	(0.0835)
Northern Europe	0.0683	0.0351	0.0837	0.0413	0.0639	0.0311	0.0280
	(0.2523)	(0.1845)	(0.2770)	(0.1990)	(0.2449)	(0.1738)	(0.1653)
Americas	0.0495	0.0409	0.0658	0.0171	0.0183	0.0167	0.0164
	(0.2169)	(0.1987)	(0.2480)	(0.1298)	(0.1341)	(0.1285)	(0.1270)
Other	0.0436	0.0468	0.0384	0.0537	0.0411	0.0478	0.0724
	(0.2042)	(0.2118)	(0.1923)	(0.2256)	(0.1987)	(0.2137)	(0.2595)
N	4083	171	2628	1284	438	418	428

Standard deviations in parentheses. Distinctive refers to migrants for which $\Delta Ai < 0$; Keepers are migrants for which $\Delta Ai = 0$; Americanize refers to migrants for which $\Delta Ai > 0$; quartiles refer to migrants who americanize.

back to appendix back to slides

Timing of Name change

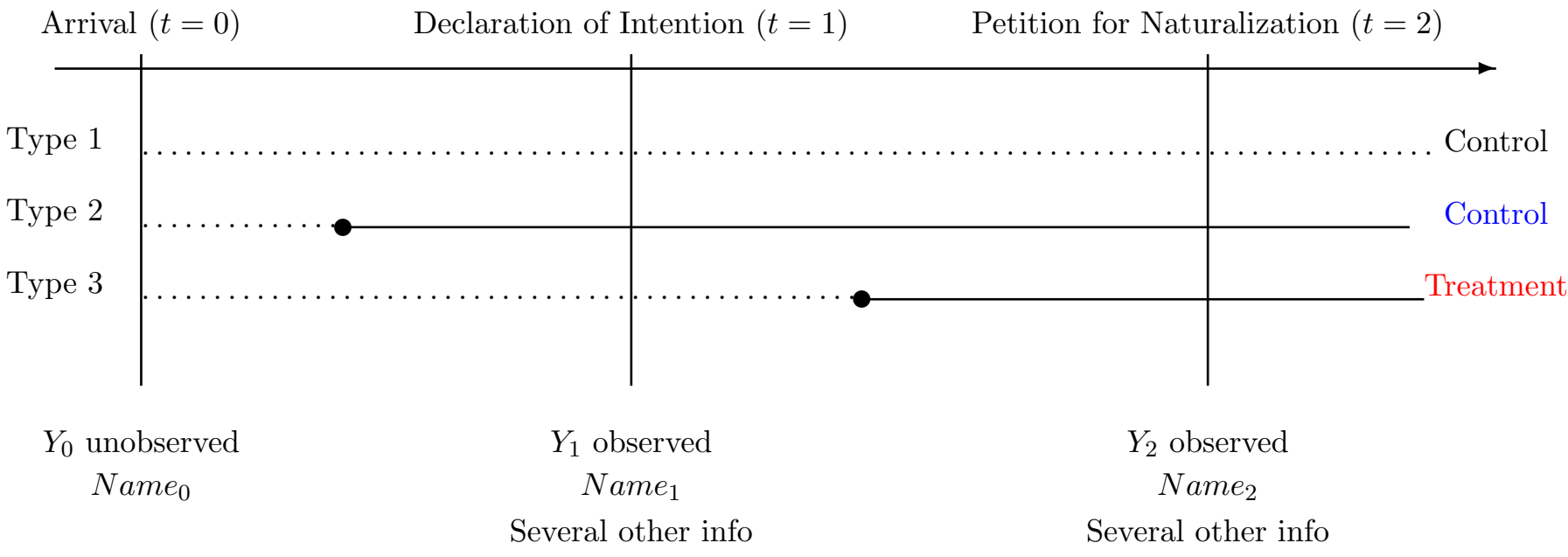
Some people change their name before declaration, but we do not observe their earnings before declaration.



back to appendix
back to slides

Timing of Name change

A true first-difference model would consider the Type 2 as non-treated:



back to appendix back to slides

Reduced form

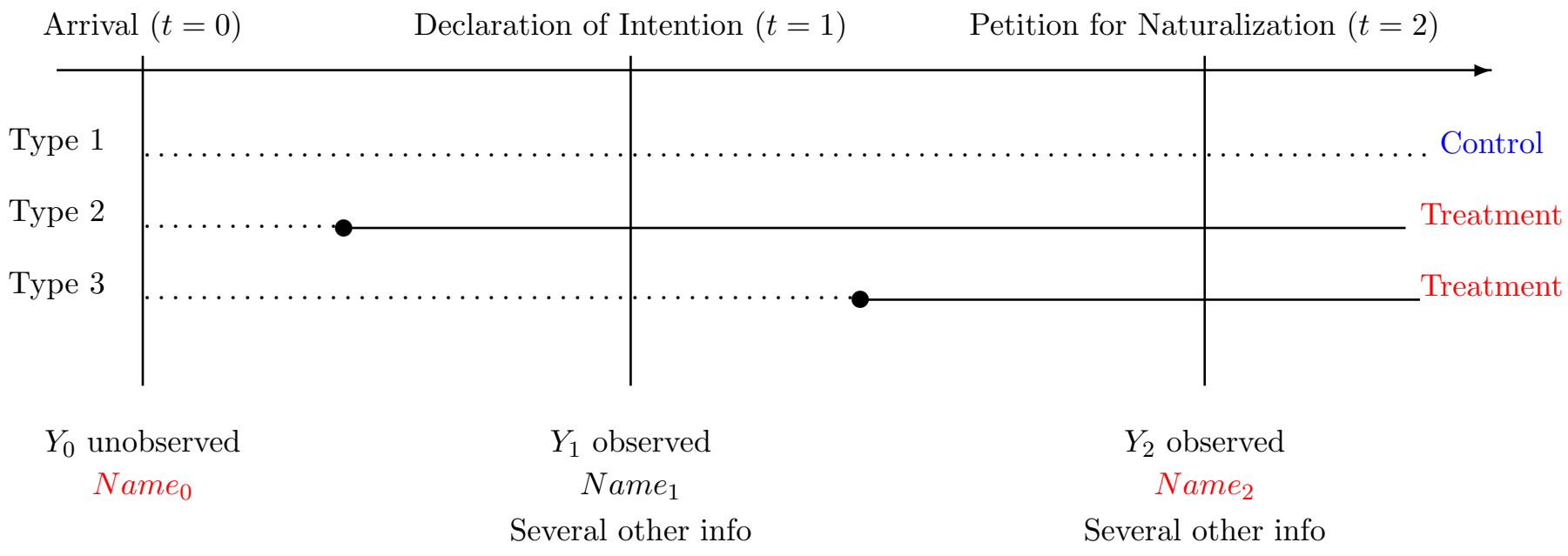
	I	II	III
Scrabble points	0.0301**	0.0262*	0.0302**
	(0.0130)	(0.0135)	(0.0136)
N	4083	4083	4083
C. of birth Ind.	No	Yes	Yes
Lab. Mk. Ind.	No	No	Yes

Robust standard errors in parenthesis.
Arrival cohort, country of birth and labor market should be interpreted as interactions with the time trend.

▸ back to appendix ▸ back to slides

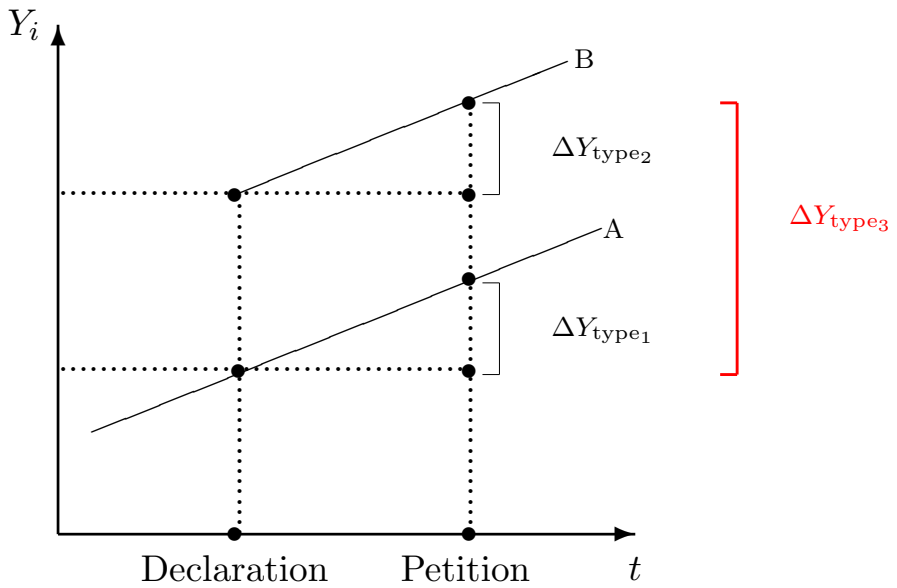
Timing of Name change

Instead, impose to consider Type 2 individuals as treated:



back to appendix back to slides

Sensitivity to Type Definition

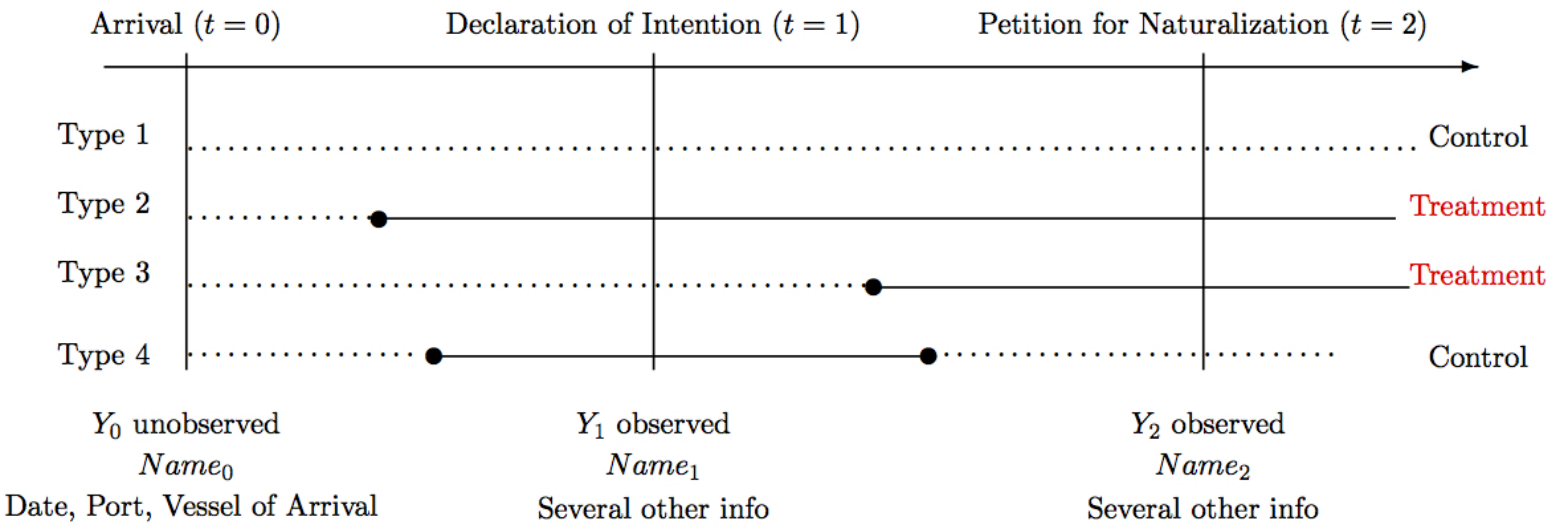


With $P = Prob(Type_2)$:

$$\hat{\beta}_1 = (P\Delta Y_{\text{Type}_2} + (1-P)\Delta Y_{\text{Type}_3}) - \Delta Y_{\text{Type}_1}$$

back to appendix back to slides

Robustness



Benchmark model: Name change if $Name_2$ differs from $Name_0$.

- Captures all permanent changes.
- Change in name is based on a variable ($Name_0$) that is predetermined with respect to any post-arrival outcome in the U.S. (no reverse causality).

▸ back to appendix ▸ back to slides

Descriptive Statistics by Scrabble points

Individual characteristics	At Declaration			Difference petition-declaration		
Variable	$\leq$ median	$>$ median	t-test	$\leq$ median	$>$ median	t-test
Log Occupational Score	3.1896	3.1645	0.0653	0.0271	0.0348	0.6044
	(0.4245)	(0.4469)		(0.4507)	(0.5036)	
Age	30.6362	30.5969	0.8861	5.3659	5.2634	0.2058
	(8.6483)	(8.7435)		(3.0312)	(1.9662)	
Years Since Migration	7.2145	6.8721	0.1335	4.6940	4.7094	0.7738
	(7.3254)	(7.2452)		(1.7390)	(1.6933)	
Married	0.4856	0.4636	0.1596	0.2250	0.2139	0.3906
	(0.4999)	(0.4988)		(0.4177)	(0.4101)	
Has U.S.-born spouse	0.0498	0.0521	0.7346	0.0561	0.0560	0.9914
	(0.2175)	(0.2223)		(0.2302)	(0.2301)	
Number of children	0.9546	0.9125	0.3807	0.3036	0.2915	0.5324
	(1.5535)	(1.5163)		(0.6262)	(0.6034)	
Has U.S.-born child(ren)	0.2235	0.2050	0.1496	0.1562	0.1534	0.8134
	(0.4167)	(0.4038)		(0.3763)	(0.3765)	
Moved into N.Y.	0.1113	0.1146	0.7407	-0.0986	-0.1028	0.6896
	(0.3145)	(0.3186)		(0.3337)	(0.3331)	
Year of arrival	1916.2	1918.4	0.1015			
	(60.2955)	(6.7683)				
Italy	0.2128	0.2212	0.5125			
	(0.4094)	(0.4152)				
Russian Empire	0.1850	0.1819	0.8006			
	(0.3884)	(0.3859)				
Central Europe (excl. DE)	0.1733	0.1627	0.3687			
	(0.3786)	(0.3692)				
Southern Europe (excl. IT)	0.0415	0.0221	0.0004			
	(0.1995)	(0.1471)				
Germany	0.1196	0.0944	0.0093			
	(0.3245)	(0.2924)				
Ireland	0.0307	0.1539	0.0000			
	(0.1727)	(0.3609)				
UK	0.0312	0.0472	0.0086			
	(0.1740)	(0.2121)				
Northern Europe	0.1044	0.0320	0.0000			
	(0.3059)	(0.1759)				
Americas	0.0571	0.0418	0.0241			
	(0.2321)	(0.2002)				
Other	0.0444	0.0428	0.7977			
	(0.2061)	(0.2024)				
N	2049	2034	4083	2049	2034	4083

More Channels

	Late Americanizers Only				Early Declarant			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0046	0.3704*	-	2.3429	0.0247	0.1140*	0.2507*	0.6512
	(0.0285)	(0.2106)	-	(2.9352)	(0.0242)	(0.0658)	(0.1487)	(0.5617)
N	4940	2470	-	2470	3668	1834	619	1834
	No Flagged Occupations				No Flagged Addresses			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0355*	0.1221**	0.2268	0.4836*	0.0343*	0.1219***	0.2495**	0.5789**
	(0.0203)	(0.0474)	(0.1383)	(0.2664)	(0.0179)	(0.0412)	(0.1248)	(0.2405)
N	6742	3371	1387	3371	7708	3854	1628	3854
	Self-Employed				Employees			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0435***	0.0525**	0.1610***	–0.1156	0.0090	0.1661**	0.3385**	1.0574**
	(0.0154)	(0.0232)	(0.0604)	(0.2018)	(0.0278)	(0.0714)	(0.1684)	(0.4216)
N	3468	1734	822	1734	4698	2349	916	2349
	Americanizes Surname				Surname Keepers			
	OLS	FD	NC	IV	OLS	FD	NC	IV
A_{it}	0.0035	0.0522	–0.1236	0.3587	0.0320*	0.1290***	0.2469**	0.5204**
	(0.1114)	(0.1126)	(0.2779)	(0.7996)	(0.0182)	(0.0429)	(0.1199)	(0.2239)
N	574	608	428	608	7592	3475	1310	3475

Robust standard errors in parenthesis.
A_{it} = Americanization index, which varies between 0 (names with the lowest frequency) and 1 (names with the highest frequency). See text for explanation.
First Stage F-tests. Late Americanizers only: 2.641; No Flagged Occupations: 65.388; No Flagged Addresses: 81.344; Americanizes Surname: 13.734; Surname Keepers: 74.755.
All models include the covariates in the last column of Table **??**.

back to presentation back to appendix

Focus on First Names

First names are a crucial marker of individual identity (Lieberson, 2000)

Link to current results:

- Fryer and Levitt (2004) provide evidence of the importance of **first** names by showing that the surge in distinctively Black names in the US since the Seventies.
- The first name is a signal for the employer about the cultural and socio-economic background. Audit studies (Emily or Greg vs Jamal and Lakisha) show that **first** names associated to a cultural minority are perceived negatively by employers (Bertrand and Mullainathan, 2004).

Surname Americanization was much less common (only 7%).

▸ back to presentation ▸ back to appendix

Representativeness

Table: Descriptives, IPUMS Comparison, NYC 1930

	Sample	Citizens, flow
Occupational Score	26.236	26.229
	(8.696)	(9.175)
Age	34.282	33.286
	(8.353)	(8.875)
Married	0.631	0.674
	(0.483)	(0.469)
Number of children	0.973	0.981
	(1.496)	(1.366)
Covariates (migrants only)		
Years since migration	7.418	8.0432
	(1.487)	(2.310)
Birthplace		
Italy	0.193	0.217
	(0.395)	(0.413)
Russian Empire	0.138	0.140
	(0.345)	(0.347)
Central Europe (excl. DE)	0.143	0.216
	(0.350)	(0.411)
Southern Europe (excl. IT)	0.025	0.028
	(0.155)	(0.165)
Germany	0.148	0.080
	(0.355)	(0.271)
Ireland	0.126	0.080
	(0.332)	(0.271)
UK	0.047	0.062
	(0.211)	(0.241)
Northern Europe	0.092	0.038
	(0.290)	(0.192)
Americas	0.052	0.098
	(0.222)	(0.297)
Other	0.036	0.042
	(0.187)	(0.271)
N	2674	5809

Standard deviations in parentheses.
Sources: Ancestry.com (Col I) and 1930 Census (Col