

Language and Gender Roles among Immigrants to the US: A Historical Perspective

Victor Gay[§], Daniel L. Hicks[†], and Estefania Santacreu-
Vasut[‡]

Appendix

1. Dataset

In this section, we provide details on how we constructed the variables used in the analysis. Except for the grammatical variables, the data for the U.S. censuses 1910-1990 and for the ACS 2000-2013 are from IPUMS (Ruggles et al. 2015). The dataset is available in the replication material on the authors' website.

We detail how we assign a language to individual immigrants (section 1.1), a country of birth (section 1.2), a citizenship (section 1.3), an occupation (section 1.4), a race (section 1.5) and other covariates (section 1.6).

[§] Department of Economics, The University of Chicago, victorgay@uchicago.edu.

[†] Department of Economics, University of Oklahoma, hicksd@ou.edu.

[‡] *Corresponding author*, Department of Economics, ESSEC Business School and THEMA, santacreuvvasut@essec.edu.

1.1. Language assignment

Two variables directly give the respondent's language: LANGUAGE and MTONGUE. They are available for the following different census years on IPUMS (see section 1.5).

- LANGUAGE. It reports the language that the respondent spoke at home. Available for the following years: 1910, 1920, 1930, 1980, 1990, 2000-2014.
- MTONGUE. It reports the mother tongue of the respondent. Available for the following years: 1910, 1920, 1930, 1940, 1960, 1970.

While both variables are available for 1910, 1920, and 1930, we use MTONGUE for those census years.

We generally use the general 2-digit codes on IPUMS to match a language with a language in WALS. However, this general 2-digit code is sometimes not precise enough for us to merge the IPUMS data with languages defined in WALS. As a result, we use the detailed 4-digit codes for some languages. We detail below which specific 4-digit codes were used in lieu of the general 2-digit codes (“no language information” means that the language is not registered on WALS):

- French (11): French (1100), Walloon French (1110 — merged with French), Provençal (1120), Patois (1130 — merged with French), and Haitian Creole (1140). No language information on Cajun (1150).
- Other Balto-Slavic (26): Bulgarian (2610), Sorbian (2620), Macedonian (2630).
- Other Persian dialects (30): Pashto (3010), Kurdish (3020), Baluchi (3030), Tajik (3040), Ossetic (3050).
- Hindi and related (31): Hindi (3102), Urdu (3103), Bengali (3112), Penjabi (3113), Marathi (3114), Gujarati (3115), Magahi (3116), Bagri (3117), Oriya (3118), Assamese (3119), Kashmiri (3120), Sindhi (3121), Dhivehi (3122), Sinhala (3123), Kannada (3130). No language information on

Sanskrit (3111).

- Other Altaic (37): Chuvash (3701), Karakalpak (3702), Kazakh (3703), Kirghiz (3704), Tatar (3705), Uzbek (3706), Azerbaijani (3707), Turkmen (3708), Khalkha (3710). No language information on Tungus (3711).
- Dravidian (40): Brahui (4001), Gondi (4002), Telugu (4003), Malayalam (4004), Tamil (4005), Bhili (4010), Nepali (4011).
- Chinese (43): Cantonese (4302), Mandarin (4303), Hakka (4311), Fuzhou (4314), Wu (Changzhou) (4315). No language information on Kan (4312), Hsiang (4313).
- Thai, Siamese, Lao (47): Thai (4710), Lao (4720).
- Other East/Southeast Asian (51): Ainu (5110), Khmer (5120), Yukaghir (5140), Muong (5150).
- Other Malayan (53): Taiwanese (5310), Javanese (5320), Malagasy (5330), Sundanese (5340).
- Micronesian, Polynesian (55): Carolinian (5502), Chamorro (5503), Kiribati (5504), Kosraean (5505), Marshallese (5506), Mokilese (5507), Nauruan (5509), Pohnpeian (5510), Chuukese (5511), Ulithian (5512), Woleaian (5513), Yapese (5514), Samoan (5522), Tongan (5523), Tokelauan (5525), Fijian (5526), Marquesan (5527), Maori (5529), Nukuoro (5530). No language information on Rarotongan (5528).
- Hamitic (61): Berber (6110), Hausa (6120), Beja (6130).
- Nilotic (63): Nubian (6302), Fur (6304), Swahili (6308), Koranko (6309), Fula (6310), Gurung (6311), Bété (6312), Efik (6313), Sango (6314).
- Other Indian languages (91): Yurok (9101), Makah (9112), Kutenai (9120), Haida (9130), Yakut (9131), Yuchi (9150).
- Mayan languages (9210): Purépecha (9220), Mapudungun (9230), Oto (9240), Quechua (9250), Arawak (9270), Muisca (9280), Guaraní (9290).

This leaves us with 93 distinct languages, with 51 languages containing more than 4,900 observations (i.e. 0.1% of the dataset). In Table 1 below, we show the list of these 51 languages, together with the continent

from which they originate, their linguistic family, the number of individuals in our sample speaking this language in the home, and the value for the SBI variable.

Table 1: List of languages

Language	Continent	Family	Observations	Share %	SBI
Amharic	Africa	Afro-Asiatic	14,865	0.30	1
Arabic	Africa	Afro-Asiatic	76,187	1.55	1
Beja	Africa	Afro-Asiatic	5,466	0.11	1
Swahili	Africa	Niger-Congo	5,803	0.12	0
Hebrew	Asia	Afro-Asiatic	20,389	0.42	1
Turkish	Asia	Altaic	13,325	0.27	0
Khmer	Asia	Austro-Asiatic	21,888	0.45	0
Vietnamese	Asia	Austro-Asiatic	146,752	2.99	0
Indonesian	Asia	Austronesian	13,109	0.27	0
Kannada	Asia	Dravidian	4,982	0.10	1
Tamil	Asia	Dravidian	18,385	0.37	1
Armenian	Asia	Indo-European	28,095	0.57	0
Bengali	Asia	Indo-European	21,340	0.43	0
Hindi	Asia	Indo-European	70,874	1.44	1
Latvian	Asia	Indo-European	18,396	0.37	1
Marathi	Asia	Indo-European	7,370	0.15	1
Nepali	Asia	Indo-European	4,991	0.10	1
Panjabi	Asia	Indo-European	22,137	0.45	1
Persian	Asia	Indo-European	45,462	0.93	0
Russian	Asia	Indo-European	106,329	2.17	1
Ukrainian	Asia	Indo-European	19,286	0.39	1
Japanese	Asia	Japanese	59,686	1.22	0
Korean	Asia	Korean	130,761	2.66	0
Burmese	Asia	Sino-Tibetan	4,910	0.10	0
Cantonese	Asia	Sino-Tibetan	132,067	2.69	0
Mandarin	Asia	Sino-Tibetan	177,658	3.62	0
Tibetan	Asia	Sino-Tibetan	14,629	0.30	0
Lao	Asia	Tai-Kadai	14,123	0.29	0
Thai	Asia	Tai-Kadai	24,393	0.50	0

Finnish	Asia	Uralic	11,326	0.23	0
Albanian	Europe	Indo-European	10,567	0.22	1
Bulgarian	Europe	Indo-European	6,564	0.13	1
Czech	Europe	Indo-European	18,873	0.38	1
Danish	Europe	Indo-European	15,327	0.31	0
Dutch	Europe	Indo-European	29,684	0.60	1
French	Europe	Indo-European	112,487	2.29	1
Gaelic	Europe	Indo-European	16,327	0.33	1
German	Europe	Indo-European	228,832	4.66	1
Greek	Europe	Indo-European	44,318	0.90	1
Italian	Europe	Indo-European	192,055	3.91	1
Norwegian	Europe	Indo-European	24,938	0.51	1
Polish	Europe	Indo-European	133,445	2.72	1
Portuguese	Europe	Indo-European	83,362	1.70	1
Romanian	Europe	Indo-European	21,359	0.44	1
Serbian-Croatian	Europe	Indo-European	37,424	0.76	1
Slovak	Europe	Indo-European	19,480	0.40	1
Slovene	Europe	Indo-European	5,868	0.12	1
Spanish	Europe	Indo-European	2,333,144	47.54	1
Swedish	Europe	Indo-European	45,711	0.93	0
Hungarian	Europe	Uralic	32,075	0.65	0
Tagalog	Oceania	Austronesian	231,419	4.72	1
Other			9,091	0.19	
Total			4,907,334	100.00	

1.2. Country of birth

The country of birth is given by the IPUMS variable BPLD for the detailed coding. We exclude respondents whose country of birth is grouped at a wide level of aggregation e.g. “Northern Europe”. It is the case for the following grouped countries, where we indicate the IMPUS country code and the number of observations dropped after having dropped the 1950 census for which no language information is available (89,499 observations), US citizens not born in mainland US (362,775 observations), and observations

for which no language information is available (94,235 observations):

- Atlantic Islands (16000; 208).
- North America, ns/nec (19900; 124).
- Central America (21000; 155).
- Central America ns (21090; 576).
- Antilles, ns (26090; 4).
- Latin America, ns (26092; 16).
- Americas, ns (29900; 1,643).
- South America (30000; 392).
- South America, ns (30090; 3,366).
- Northern Europe, ns (41900; 8).
- Western Europe, ns (42900; 173).
- Southern Europe, ns (44000; 2).
- Central Europe, ns (45800; 357).
- Eastern Europe, ns (45900; 10).
- Baltic States, ns (46300; 8).
- Europe, ns (49900; 3,207).
- East Asia, ns (50900; 926).
- Southeast Asia, ns (51900; 37).
- Persian Gulf States, ns (54600; 54).
- Middle East, ns (54700; 40).
- Southwest Asia, nec/ns (54800; 413).
- Asia Minor, ns (54900; 99).
- South Asia, nec (55000; 1).
- Asia, nec/ns (59900; 7,150).
- Africa (60000; 30).
- Northern Africa (60010; 147).
- North Africa, ns (60019; 1,092).
- Western Africa, ns (60038; 3,927).
- British Indian Ocean Territory (60040; 7).
- Europa (60059; 1).
- Eastern Africa, nec/ns (60064; 1,903).
- Central Africa (60070; 217).
- Central Africa, ns (60080; 8).
- Southern Africa (60090; 24).

- Southern Africa, ns (60096; 83).
- Africa, ns/nec (60099; 14,170).
- Pacific Islands (71000; 38).
- Polynesia, ns (71029; 4)
- U.S. Pacific Trust Territories (71040; 18).
- Pacific Trust Terr, ns (71049; 41).
- Oceania, ns/nec (71090; 564).
- Antarctica, ns/nec (80000; 14).
- British Antarctica Terr. (80020; 13).
- Aborad (unkown) or at sea (90000; 85).
- Aborad, ns (90010; 56,080).
- At sea (90020; 322).
- Other, nec (95000; 1,086).
- Missing/blank (99900; 12).

Also, we drop groups of Caribbean islands that are too widely defined:

- West Indies (26000; 266).
- British West Indies (26040; 686).
- British West Indies, ns/nec (26069; 205).
- Other West Indies (26070; 485).
- Caribbean, ns/nec (26091; 3,947).
- West Indies, ns (26094; 6,638).

In total, this procedure drops 111,081 observations, i.e. 1.7% of the dataset.

Some places of birth are detailed at a lower level than a country. In those cases, we group these sub-levels into an aggregate level. Some groupings may seem extraneous from a contemporary perspective: for instance, we group Slovakia and Czech Republic into Czechoslovakia. However, individuals in our dataset were born between 1845 and 1998. Hence, we believe it makes sense to make such groupings (Slovakia and Czech Republic were separated in 1993). We group the following sub-levels:

- Canadian provinces (15010–15083; 51,347) into Canada (15000).

- Panama Canal Zone (21071; 15) into Panama (21070).
- Austrian regions (45010–45080) into Austria (45000).
- Berlin districts (45301-45303; 422), West-German regions (45310-45333; 38,185), East-German regions (45340-45353; 1,614), Prussia, nec (45360; 648) into Germany (45300).
- Polish regions (45510–45530; 9,828) into Poland (45500).
- Transylvania (45610; 90) into Romania (45600).
- Bessardia (46521; 53) into Moldavia (46520).
- Anegada (26046; 22), Cooper (26047; 4), Jost Van Dyke (26048; 3), Peter (26049; 1), Tortola (26050; 25), Virgin Gorda (26051; 6), British Virgin Islands, ns/nec (26052; 37) into British Virgin Islands (26045).
- Aruba (26071; 235), Bonaire (26073; 25), Curacao (26074; 58), Dutch St. Maarten (26075; 64), Saba (26076; 5), St. Eustatius (26077; 3), Dutch Caribbean, ns/nec (26079; 38) into Netherlands Antilles (26072).
- French St. Maarten (26080; 8), Guadeloupe (26081; 60), Martinique (26082; 74), St. Barthelemy (26083; 22) into French Caribbeans (26089).
- French Guyana (30035; 3) into Guyana (30040).
- Lapland (40300; 4) into Finland (40100).
- Svalbard and Jan Meyen (40410; 2) into Norway (40400).
- Channel Islands (41010; 22), Guernsey (41011; 10), Jersey (41012; 39), Isle of Man (41020; 96), United Kingdom, ns/nec (41300; 34,696) into England (41000).
- Alsace-Lorraine (42110; 575), Alsace (42111; 123), Lorraine (42112; 15) into France (42100).
- Dodecanese Islands (43310; 2), Turkey Greece (43320; 51), Macedonia (43330; 2,926) into Greece (43300).
- Azores (43610; 7,394), Madeira Islands (43620; 176), Cape Verde Islands (43630; 64), St. Miguel (43640; 23) into Portugal (43600).
- Vatican City (43900; 2) into Italy (43400).
- Bohemia (45210; 2,063), Bohemia-Moravia (45211; 1,418), Slovakia (45212; 3,684), Czech Republic (45213; 3,213) into Czechoslovakia (45200).

- Croatia (45710; 5,488), Montenegro (45720; 413), Serbia (45730; 1,899), Bosnia (45740; 12,727), Dalmatia (45750; 129), Slavonia (45760; 90), Carniola (45770; 19), Slovenia (45780; 143), Kosovo (45790; 320) into Yugoslavia (45700).
- Siberia (46548; 51), USSR ns/nec (46590; 9,126) into Other USSR/Russia (46500).
- North Korea (50210; 107), South Korea (50220; 14,240) into Korea (50200).
- East Indies (51210; 25), East Timor (51220; 11) into Indonesia (51200).
- Indochina, ns (51910; 2,488) into Vietnam (51800).
- Mesopotamia (53210; 16), Iraq/Saudi Arabia (53300; 11) into Iraq (53200).
- Gaza Strip (53410; 14), Palestine (53420; 954), West Bank (53430; 25), Israel (53440; 42) into Israel/Palestine (53400).
- European Turkey (54210; 168), Asian Turkey (54220; 295) into Turkey (54200).
- Yemen, PDR (South) (54500; 43) into Yemen Arab Republic (54400).
- Western Sahara (60017; 7) into Morocco (60014).
- Cocos Islands (70014; 1), Heard and McDonald Islands (80050; 7) into Australia (70010).
- Kiribati (71032; 43), Canton and Enderbury (71033; 26), Nauru (71039; 2) into Micronesia (71039).

This leaves us with 206 distinct countries of birth, with 75 countries containing more than 4,900 observations (i.e. 0.1% of the dataset). In Table 2 below, we show the list of these 75 countries, together with their continent, and the number of individuals in our sample that were born there.

Table 2: List of countries of birth

Country of birth	Continent	Observations	Share %
------------------	-----------	--------------	---------

Egypt	Africa	17,553	0.36
Ethiopia	Africa	13,669	0.28
Kenya	Africa	5,197	0.11
Morocco	Africa	6,105	0.12
Afghanistan	Asia	6,974	0.14
Bangladesh	Asia	15,887	0.32
Burma (Myanmar)	Asia	7,892	0.16
Cambodia	Asia	22,709	0.46
China	Asia	182,962	3.73
Hong Kong	Asia	36,408	0.74
India	Asia	115,785	2.36
Indonesia	Asia	11,344	0.23
Iran	Asia	46,114	0.94
Iraq	Asia	9,516	0.19
Israel/Palestine	Asia	17,788	0.36
Japan	Asia	59,156	1.21
Jordan	Asia	8,229	0.17
Korea	Asia	131,221	2.67
Laos	Asia	30,472	0.62
Lebanon	Asia	17,657	0.36
Malaysia	Asia	7,345	0.15
Pakistan	Asia	8,243	0.17
Philippines	Asia	237,261	4.83
Syria	Asia	8,622	0.18
Taiwan	Asia	50,428	1.03
Thailand	Asia	25,811	0.53
Turkey	Asia	16,317	0.33
Vietnam	Asia	169,303	3.45
Argentina	Central and South America	23,028	0.47
Bolivia	Central and South America	8,767	0.18
Brazil	Central and South America	36,400	0.74
Chile	Central and South America	12,688	0.26
Colombia	Central and South America	81,945	1.67
Costa Rica	Central and South America	11,191	0.23

Cuba	Central and South America	146,663	2.99
Dominican Republic	Central and South America	97,294	1.98
Ecuador	Central and South America	46,143	0.94
El Salvador	Central and South America	136,417	2.78
Guatemala	Central and South America	82,094	1.67
Haiti	Central and South America	15,621	0.32
Honduras	Central and South America	46,094	0.94
Mexico	Central and South America	1,482,736	30.21
Nicaragua	Central and South America	34,504	0.70
Panama	Central and South America	13,927	0.28
Peru	Central and South America	47,062	0.96
Uruguay	Central and South America	5,668	0.12
Venezuela	Central and South America	19,980	0.41
Albania	Europe	6,883	0.14
Armenia	Europe	10,482	0.21
Austria	Europe	35,754	0.73
Belgium	Europe	8,213	0.17
Bulgaria	Europe	6,801	0.14
Czechoslovakia	Europe	37,006	0.75
Denmark	Europe	15,095	0.31
England	Europe	8,512	0.17
Finland	Europe	12,579	0.26
France	Europe	29,081	0.59
Germany	Europe	175,109	3.57
Greece	Europe	42,696	0.87
Hungary	Europe	32,627	0.66
Ireland	Europe	9,729	0.20
Italy	Europe	186,650	3.80
Lithuania	Europe	14,540	0.30
Netherlands	Europe	19,674	0.40
Norway	Europe	24,986	0.51
Poland	Europe	128,732	2.62
Portugal	Europe	45,880	0.93
Romania	Europe	25,424	0.52

Russia	Europe	85,122	1.73
Spain	Europe	16,410	0.33
Sweden	Europe	43,736	0.89
Switzerland	Europe	12,658	0.26
Ukraine	Europe	31,931	0.65
Yugoslavia	Europe	48,234	0.98
Canada	North America	53,932	1.10
Other		104,668	2.13
Total		4,907,334	100.00

1.3. Citizens

Some U.S. citizens may be born abroad if one of their parents is a U.S. citizen. We exclude U.S. citizens at birth as follows, so that we don't mistakenly consider them as migrants:

- First, we use the variable `CITIZEN`. It is available every year except 1960 and 1970 (met 2, the sample for which we know the language). From 1950, the citizenship status is ascertained for all persons who were foreign-born, regardless of their citizenship status at birth with a category “born abroad of American parents”. We exclude this category. For earlier years this variable is not available.
- For the census years for which `CITIZEN` is not available, we use the parents’ country of birth `MBPL` and `FBPL` to determine U.S. citizenship for 1910-1940, and 1960-1970. If one of them was born in the US, we exclude the respondent. Note that for 1960-1970, the State of birth of the parent is not detailed if born in the US, only “United States, ns” (099).

1.4. Occupation

We exclude women working in farming. For this, we use the following codes of occupation from the OCC variable:

- 1910-1920: 0-26, 63-69.
- 1930: 1, 2.
- 1940: 98, 99, 844, 866, 888.
- 1960: 200, 222, 901, 902, 903, 905.
- 1970: 801, 802, 806, 821, 822, 823, 824, 826.
- 1980: 473, 474, 475, 476, 477, 479, 483, 484.
- 1990: 473, 474, 475, 476, 477, 479, 480, 483, 484.
- 2000: 605.
- 2000-2004: 6050, 6020.

1.5. Race

We code race using the RACE and HISPAN variables from IPUMS. We group values from the RACE variable into White, Black and Other. We group values from HISPAN into Hispanic and Not Hispanic.

1.6. Other covariates

- Year of birth: we use the BIRTHYR variable, except for 1910, where we generate it using AGE – 1910. individuals in the dataset are born between 1845 and 1998. This implies 154 year of birth indicators in the regressions.
- Family size: we use the FAMSIZE variable. It is the number of own family members in the household.
- Number of children: we use the NCHILD variable. It is the number of own children in the household.
- Marital status: we use the MARST variable which we group into Married, Separated, Widowed and Never Married.
- Education: we use the EDUC variable which gives a dozen education

categories. For censuses of 1910, 1920 and 1930, this variable is not available. We use the `LIT` variable, which provides information on the level of literacy of the respondent, and assign `Grade 8` to literate individuals, and `No Schooling` for illiterate individuals. Then, we group these categories into four education categories: `No schooling` for individuals with less than elementary schooling, `Elementary` for individuals that attained grade 5 through grade 8, `High school` for individuals that attained grade 9 through grade 12, and `College` for individuals that went to college.

- **Currently in school:** we use the `SCHOOL` variable. We assign `Not in school` to the individuals for which this information is missing (less than 1% of the sample).
- **Ownership of dwelling:** we use the `OWNERSHP` variable. We assign `Rented` to the individuals for which this information is missing (about 2% of the sample).
- **Speaks English:** we use the `SPEAKENG` variable. This variable is missing for the censuses of 1940, 1960 and 1970, so we assign the average of the censuses of 1930, 1980 and 1980, respectively. We also assign `Does not speak English` to individuals for which this information is missing (about 0.1% of the sample). Moreover, we assign `Yes, speaks English` to individuals that speak English, regardless of whether they reported speaking “very well” or speaking “well”.
- **Decade of migration:** we use the `YRIMMIG` variable that we group into decades. This variable is missing for the censuses of 1940, 1960 and 1970, so we assign the average time since immigration from the censuses of 1930, 1980 and 1980, respectively. We denote this variable `DCIMMIG`.
- **Person weight:** we use the `PERWT` variable.
- **Survey years:** we use the U.S. censuses of 1910, 1920, 1930, 1940, 1960, 1970, 1980, 1990, and 2000. We also use the ACS 2001-2014, which we group into the following years: 2001-2005, 2006-2010, and 2011-2014. Table 3 below describes the sample distribution across censuses. The 1970 extract is the 1% state fm2 extract, and the 1980 extract is the 5% state extract. Regarding the ACS, the extracts are 0.43% for the 2001 ACS, 0.38% for the 2002 ACS,

0.42% for the 2003 ACS, 0.42% for the 2004 ACS, and 1.0% for all other ACSs.

Table 3: Sample distribution

Survey year	Extract	Observations	Share	Cum.
1910	1.0%	81,584	1.66	1.66
1920	1.0%	89,013	1.81	3.48
1930	5.0%	431,963	8.80	12.28
1940	1.0%	21,543	0.44	12.72
1960	1.0%	42,963	0.88	13.60
1970	1.0%	42,717	0.87	14.47
1980	5.0%	336,049	6.85	21.32
1990	5.0%	547,037	11.15	32.47
2000	5.0%	930,149	18.95	51.42
2001-2005	ACS	448,768	9.14	60.56
2006-2010	ACS	1,045,798	21.31	81.87
2011-2014	ACS	889,750	18.13	100.00
Total		4,907,334	100.00	

2. Immigrants origins

The following tables are the replication of Table 1 in the paper for all the census years in our sample, separately.

Table 4: Origin of female migrants in 1910 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	0.0	0.0	0.0
Asia	0.5	0.4	0.1
Central and South America	2.2	0.0	2.2
Europe	91.6	14.3	77.3
North America	5.7	0.0	5.7
Oceania	0.0	0.0	0.0
Total	32,287	4,750	27,537

Table 5: Origin of female migrants in 1920 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	0.0	0.0	0.0
Asia	1.7	1.5	0.2
Central and South America	4.2	0.0	4.1
Europe	90.3	13.5	76.9
North America	3.8	0.1	3.7
Oceania	0.0	0.0	0.0
Total	37,453	5,630	31,823

Table 6: Origin of female migrants in 1930 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	0.0	0.0	0.0
Asia	2.0	1.6	0.4
Central and South America	6.7	0.0	6.7
Europe	87.4	12.1	75.2
North America	3.9	0.0	3.8
Oceania	0.0	0.0	0.0
Total	187,757	25,870	161,887

Table 7: Origin of female migrants in 1940 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	0.0	0.0	0.0
Asia	1.8	1.3	0.5
Central and South America	5.1	0.0	5.1
Europe	88.4	13.0	75.5
North America	4.6	0.1	4.5
Oceania	0.0	0.0	0.0
Total	9,104	1,309	7,795

Table 8: Origin of female migrants in 1960 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	0.1	0.0	0.1
Asia	5.9	4.1	1.8
Central and South America	13.5	0.0	13.5
Europe	76.3	7.4	68.9
North America	4.2	0.1	4.1
Oceania	0.0	0.0	0.0
Total	21,596	2,498	19,098

Table 9: Origin of female migrants in 1970 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	0.4	0.0	0.3
Asia	10.9	7.0	3.9
Central and South America	27.8	0.1	27.7
Europe	56.1	4.2	51.9
North America	4.7	0.0	4.6
Oceania	0.1	0.0	0.1
Total	22,953	2,638	20,315

Table 10: Origin of female migrants in 1980 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	0.9	0.1	0.9
Asia	25.3	16.0	9.3
Central and South America	42.9	0.1	42.8
Europe	28.6	2.0	26.6
North America	1.9	0.1	1.9
Oceania	0.3	0.0	0.3
Total	173,283	31,650	141,633

Table 11: Origin of female migrants in 1990 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	0.9	0.1	0.9
Asia	31.9	21.3	10.7
Central and South America	48.9	0.1	48.8
Europe	17.0	1.2	15.7
North America	1.0	0.0	1.0
Oceania	0.3	0.0	0.2
Total	271,910	61,795	210,115

Table 12: Origin of female migrants in 2000 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	1.2	0.1	1.1
Asia	29.3	19.9	9.5
Central and South America	55.7	0.1	55.6
Europe	12.9	0.9	11.9
North America	0.7	0.0	0.7
Oceania	0.2	0.0	0.2
Total	452,280	95,065	357,215

Table 13: Origin of female migrants in 2001-2005 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	1.5	0.2	1.3
Asia	33.3	21.8	11.4
Central and South America	50.8	0.1	50.7
Europe	13.5	0.9	12.6
North America	0.8	0.0	0.8
Oceania	0.1	0.0	0.1
Total	229,194	52,747	176,447

Table 14: Origin of female migrants in 2006-2010 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	1.7	0.2	1.4
Asia	31.8	20.9	11.0
Central and South America	54.2	0.1	54.2
Europe	11.4	0.9	10.5
North America	0.7	0.0	0.6
Oceania	0.1	0.0	0.1
Total	530,020	116,924	413,096

Table 15: Origin of female migrants in 2011-2014 (% of total migrants)

Continent	All	SBI= 0	SBI= 1
Africa	2.1	0.3	1.8
Asia	33.0	21.5	11.5
Central and South America	53.7	0.1	53.6
Europe	10.4	0.8	9.6
North America	0.6	0.1	0.6
Oceania	0.2	0.0	0.1
Total	456,386	103,862	352,524

3. Auxiliary regressions

The following tables are the replication of Table 2 in the paper for all the census years in our sample separately. See the notes of Table 2 for more details on these regressions.

Table 16: Impact of gender in language (1910)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.012** [0.005]	0.018*** [0.005]	0.043*** [0.008]	0.048*** [0.008]
Female	-0.538*** [0.008]	-0.537*** [0.008]	-0.536*** [0.008]	-0.491*** [0.007]
Female × Sex-Based	-0.095*** [0.008]	-0.092*** [0.008]	-0.091*** [0.008]	-0.093*** [0.008]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	81,584	81,584	81,584	81,584
R ²	0.427	0.453	0.457	0.494

Table 17: Impact of gender in language (1920)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.021*** [0.005]	0.014*** [0.005]	0.027*** [0.009]	0.031*** [0.009]
Female	-0.593*** [0.007]	-0.600*** [0.007]	-0.594*** [0.007]	-0.558*** [0.007]
Female × Sex-Based	-0.069*** [0.007]	-0.065*** [0.007]	-0.071*** [0.007]	-0.073*** [0.007]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	89,013	89,013	89,013	89,013
R ²	0.457	0.470	0.475	0.502

Table 18: Impact of gender in language (1930)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	-0.012*** [0.002]	-0.030*** [0.002]	0.013*** [0.004]	0.011*** [0.004]
Female	-0.694*** [0.003]	-0.704*** [0.003]	-0.698*** [0.003]	-0.683*** [0.003]
Female × Sex-Based	-0.020*** [0.003]	-0.011*** [0.003]	-0.023*** [0.003]	-0.022*** [0.003]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	431,963	431,963	431,963	431,963
R ²	0.515	0.527	0.533	0.544

Table 19: Impact of gender in language (1940)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.010 [0.011]	-0.012 [0.011]	0.025 [0.021]	0.030 [0.020]
Female	-0.510*** [0.015]	-0.527*** [0.015]	-0.516*** [0.015]	-0.509*** [0.015]
Female × Sex-Based	-0.031* [0.017]	-0.019 [0.017]	-0.037** [0.017]	-0.034** [0.016]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	21,543	21,543	21,543	21,543
R ²	0.301	0.335	0.344	0.359

Table 20: Impact of gender in language (1960)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.001 [0.001]	0.003 [0.007]	0.013 [0.015]	0.020 [0.015]
Female	-0.477*** [0.012]	-0.475*** [0.012]	-0.470*** [0.012]	-0.465*** [0.012]
Female × Sex-Based	-0.016 [0.013]	-0.022* [0.012]	-0.030** [0.013]	-0.037*** [0.012]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	42,963	42,963	42,963	42,963
R ²	0.264	0.305	0.310	0.335

Table 21: Impact of gender in language (1970)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.023*** [0.009]	0.033*** [0.009]	0.010 [0.016]	0.015 [0.015]
Female	-0.361*** [0.013]	-0.369*** [0.012]	-0.355*** [0.013]	-0.350*** [0.012]
Female × Sex-Based	-0.024* [0.014]	-0.021 [0.013]	-0.036*** [0.013]	-0.044*** [0.013]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	42,717	42,717	42,717	42,717
R ²	0.167	0.216	0.225	0.254

Table 22: Impact of gender in language (1980)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.079*** [0.003]	0.123*** [0.003]	0.056*** [0.006]	0.050*** [0.006]
Female	-0.210*** [0.004]	-0.222*** [0.004]	-0.229*** [0.004]	-0.222*** [0.004]
Female × Sex-Based	-0.069*** [0.004]	-0.060*** [0.004]	-0.062*** [0.004]	-0.061*** [0.004]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	336,049	336,049	336,049	336,049
R ²	0.091	0.149	0.172	0.204

Table 23: Impact of gender in language (1990)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.063*** [0.002]	0.093*** [0.002]	0.048*** [0.006]	0.039*** [0.005]
Female	-0.198*** [0.003]	-0.200*** [0.003]	-0.195*** [0.003]	-0.182*** [0.003]
Female × Sex-Based	-0.030*** [0.003]	-0.029*** [0.003]	-0.041*** [0.003]	-0.044*** [0.003]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	547,037	547,037	547,037	547,037
R ²	0.069	0.135	0.150	0.187

Table 24: Impact of gender in language (2000)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	-0.019*** [0.002]	0.046*** [0.002]	0.004 [0.005]	0.004 [0.005]
Female	-0.171*** [0.002]	-0.173*** [0.002]	-0.170*** [0.002]	-0.156*** [0.002]
Female × Sex-Based	-0.019*** [0.003]	-0.020*** [0.003]	-0.028*** [0.003]	-0.033*** [0.003]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	930,149	930,149	930,149	930,149
R ²	0.042	0.085	0.092	0.123

Table 25: Impact of gender in language (2001-2005)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.064*** [0.003]	0.085*** [0.003]	0.061*** [0.008]	0.060*** [0.008]
Female	-0.189*** [0.004]	-0.191*** [0.004]	-0.189*** [0.003]	-0.180*** [0.004]
Female × Sex-Based	-0.077*** [0.004]	-0.074*** [0.004]	-0.080*** [0.004]	-0.078*** [0.004]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	448,768	448,768	448,768	448,768
R ²	0.082	0.135	0.140	0.170

Table 26: Impact of gender in language (2006-2010)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.037*** [0.002]	0.072*** [0.002]	0.043*** [0.005]	0.040*** [0.005]
Female	-0.176*** [0.002]	-0.178*** [0.002]	-0.175*** [0.002]	-0.168*** [0.002]
Female × Sex-Based	-0.053*** [0.003]	-0.052*** [0.002]	-0.060*** [0.002]	-0.062*** [0.002]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	1,045,798	1,045,798	1,045,798	1,045,798
R ²	0.066	0.126	0.136	0.170

Table 27: Impact of gender in language (2011-2014)

	Dependent variable: Labor Participation			
	(1)	(2)	(3)	(4)
Sex-Based	0.046*** [0.002]	0.075*** [0.002]	0.042*** [0.005]	0.039*** [0.005]
Female	-0.155*** [0.003]	-0.159*** [0.002]	-0.157*** [0.002]	-0.152*** [0.002]
Female × Sex-Based	-0.058*** [0.003]	-0.056*** [0.003]	-0.063*** [0.003]	-0.062*** [0.003]
Race and Age Controls	No	Yes	Yes	Yes
Country of Birth FE	No	No	Yes	Yes
Econ. and Demo. Controls	No	No	No	Yes
Observations	889,750	889,750	889,750	889,750
R ²	0.057	0.124	0.136	0.175

Bibliography

Ruggles, S., Genadek, K., Goeken, R. A., Grover, J., Sobek, M. (2015), *Integrated Public Use Microdata Series: Version 6.0*, University of Minnesota, Minneapolis.