Decomposing culture: An analysis of gender, language, and labor supply in the household

Appendix

Victor Gay* **Danliel L. Hicks†** **Estefania Santacreu-Vasut‡**
Amir Shoham§

The data, code, and any additional materials required to replicate all analyses in this article are available on the Harvard Dataverse Network, at: http://dx.doi.org/10.7910/DVN/XIFPS5.

*Department of Economics, The University of Chicago, 5757 S. University Ave, Chicago, IL, 60637, U.S.A.
†Department of Economics, University of Oklahoma, 308 Cate Center Drive, Norman, OK, 73019, U.S.A.
‡*Corresponding Author:* Department of Economics, ESSEC Business School and THEMA, 3 Avenue Bernard Hirsch, 95021 Cergy-Pontoise, France (santacreuvasut@essec.edu)
§Temple University and COMAS, 1801 N. Broad St, Philadelphia, PA, 19122, U.S.A.

A Data appendix

A.1 Samples

A.1.1 Regression sample

To build the regression sample, we use the following restrictions on the variables from the 1% ACS samples between 2005 and 2015 (Ruggles et al. 2015):

- **Immigrants to the U.S.**

  We keep respondents born abroad, i.e., those for which `BPLD` ≥ 15000. We also drop respondents born in U.S. Pacific Trust Territories (`BPLD` between 71040 and 71049), U.S. citizens born abroad (`CITIZEN` = 1), and respondents for which the citizenship status is unavailable (`CITIZEN` = 0).

- **Married women in married-couple family households with husband present**

  We keep female respondents (`SEX` = 2) not living in group quarters (`GQ` = 1, 2, or 5), and that are part of married-couple family households (`HHTYPE` = 1) in which the husband is present (`MARST` = 1 and `MARST_SP` = 1).[1]

- **Non-English speaking respondents aged 25 to 49**

  We drop respondents who report speaking English in the home (`LANGUAGE` = 1), and keep those with `AGE` between 25 and 49. Under the above restrictions, we are left with 515,572 respondents in the *uncorrected* sample.

- **Precise country of birth**

  We drop respondents who report an imprecise country of birth. This is the case for the following `BPLD` codes: North America, ns/nec (19900), Central America (21000), Central America ns (21090), West Indies (26000), BritishWest Indies (26040), British West Indies, ns/nec (26069), Other West Indies (26070), Dutch Caribbean, ns/nec (26079), Caribbean, ns/nec (26091), West Indies, ns (26094) Latin America, ns (26092), Americas, ns (29900), South America (30000), South America, ns (30090), Northern Europe, ns (41900), Western Europe, ns (42900), Southern Europe, ns (44000), Central Europe, ns (45800), Eastern Europe, ns (45900), Europe, ns (49900), East Asia, ns (50900), Southeast Asia, ns (51900), Middle East, ns (54700), Southwest Asia, nec/ns (54800), Asia Minor, ns (54900), South Asia, nec (55000), Asia, nec/ns (59900), Africa (60000), Northern Africa (60010), North Africa, ns (60019), Western Africa, ns (60038), French West Africa, ns (60039), British Indian Ocean Territory (60040), Eastern Africa, nec/ns (60064), Southern Africa (60090), Southern Africa, ns (60096), Africa,

[1] We also drop same-sex couples (`SEX_SP` = 2).

ns/nec (60099), Oceania, ns/nec (71090), Other, nec (95000), Missing/blank (99900). This concerns 4,597 respondents, that is, about 0.89% of the uncorrected sample.[2]

Some countries of birth are detailed at a lower level than a country. In these cases, we use the following rules to assign sub-national levels to countries using the `BPLD` codes: we group Canadian provinces (15010-15083) into Canada (15000), Panama Canal Zone (21071) into Panama (21070), Austrian regions (45010-45080) into Austria (45000), Berlin districts (45301-45303), into Germany (45300), West-German regions (45311-45333) into West- Germany (45310), East-German regions (45341-45353) into East-Germany (45300), Polish regions (45510-45530) into Poland (45500), and Transylvania (45610) into Romania (45600).

- **Precise language spoken**

  We need the `LANGUAGE` code to identify a precise language in order to assign the language variables from WALS (Dryer & Haspelmath 2011) to each respondent. In general, the general language code `LANGUAGE` identifies a unique language. However, in some cases, it identifies a grouping, and only the detailed code `LANGUAGED` allows us to identify a unique language. This is the case for the following values of `LANGUAGE`. We also indicate the relevant `LANGUAGED` codes that we use instead of the general `LANGUAGE` code.

  - French (11): French (1100), Walloon French (1110—merged with French), Provençal (1120), Patois (1130—merged with French), and Haitian Creole (1140).
  - 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).
  - Other Altaic (37): Chuvash (3701), Karakalpak (3702), Kazakh (3703), Kirghiz (3704), Tatar (3705), Uzbek (3706), Azerbaijani (3707), Turkmen (3708), Khalkha (3710).
  - 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).
  - Thai, Siamese, Lao (47): Thai (4710), Lao (4720).
  - Other East/Southeast Asian (51): Ainu (5110), Khmer (5120), Yukaghir (5140), Muong (5150).

[2]Note that we do not drop respondents whose husband has a country of birth that is not precisely defined. This is the case for 1,123 respondents' husbands.

- 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).
- 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).

In some cases, even the detailed codes do not provide a unique language. We drop respondents who report speaking such languages. This is the case for the following `LANGUAGE` and `LANGUAGED` codes: Slavic unkown (27), Other Balto-Slavic (26000), India n.e.c. (3140), Pakistan n.e.c. (3150), Other Indo-European (3190), Dravidian (4000), Micronesian (5501), Melanesian (5520), Polynesian (5521) Other Pacific (5590), Nilotic (6300), Nilo-Hamitic (6301), Saharan (6303), Khoisan (6305), Sudanic (6306), Bantu (6307), Other African (6390), African, n.s. (64), American Indian (70), Other Penutian (77), Tanoan languages (90), Mayan languages (9210), American Indian, n.s. (93), Native (94), No language (95), Other or not reported (96), N/a or blank (0). This concerns 3,633 respondents, that is, about 0.70% of the uncorrected sample.

- **Linguistic structure available**

We drop respondents who report speaking a language for which we do not have the value for the `SB` grammatical variable. This is the case for the following languages: Haitian Creole (1140), Cajun (1150), Slovak (2200), Sindhi (3221), Sinhalese (3123), Nepali (4011), Korean (4900), Trukese (5511), Samoan (5522), Tongan (5523), Syriac, Aramaic, Chaldean (5810), Mande (6309), Kru (6312), Ojibwa, Chippewa (7213), Navajo (7500), Apache (7420), Algonquin (7490), Dakota, Lakota, Nakota, Sioux (8104), Keres (8300), and Cherokee (8480). This concerns 27,671 respondents who report a precise language, that is, about 5.37% of the uncorrected sample.

Once we drop the 34,954 respondents that either do not report a precise country of birth, a precise language, or a language for which the value for the `SB` grammatical variable is unavailable, we are left with 480,618 respondents in the regression sample, or about 93.22% of the uncorrected sample.

A.1.2 Alternative samples

- **Sample married before migration**

  This sample includes only respondents that got married prior to migrating to the U.S. We construct the variable `marbef` that takes on value one if the year of last marriage (`YRMARR`) is smaller than the year of immigration to the the U.S. (`YRIMMIG`). Because the `YRMARR` variable is only available between 2008 and 2015, this subsample does not include respondents for the years 2005 to 2007.

- **Sample of indigenous languages**

  The sample of indigenous languages drops respondents who report a language that is not indigenous to their country of birth. We define a language as not indigenous to a country if it is not listed as a language spoken in a country in the *Encyclopedia Britannica Book of the Year* (2010, pp. 766-770).

- **Sample including all households**

  The sample including all households is similar to the regression sample described in section A.1.1, except that we keep all household types (`HHTYPE`), regardless of the presence of a husband.

- **Sample excluding language quality flags**

  The sample excluding language quality flags only keeps respondents for which the `LANGUAGE` variable was not allocated (`QLANGUAG` = 4).

A.2 Variables

A.2.1 Outcome variables

We construct labor market outcomes from the variables in the ACS 1% samples 2005-2015 (Ruggles et al. 2015). Throughout the paper, we use six outcome variables. Our variables are in lowercase, while the original ACS variables are in uppercase.

- **Labor participant** (`lfp`)

  `lfp` is an indicator variable equal to one if the respondent was in the labor force the week before the census (`LABFORCE` = 2).

- **Employed** (`emp`)

  `emp` is an indicator variable equal to one if the respondent was employed the week before the census (`EMPSTAT` = 1).

- **Yearly weeks worked** (`wks` and `wks0`)

  We create measures for yearly weeks worked by using the `WKSWORK2` variable. It indicates the number of weeks that the respondent worked during the previous year, by intervals. We use the midpoints of

these intervals to build our measures. The first measure, `wks`, is constructed as follows. It takes on a missing value if the respondent did not work at least one week during the previous year (`WKSWORK2` = 0), and takes on other values according to the following rules:

- `wks` = 7 if `WKSWORK2` = 1 (1-13 weeks).
- `wks` = 20 if `WKSWORK2` = 2 (14-26 weeks).
- `wks` = 33 if `WKSWORK2` = 3 (27-39 weeks).
- `wks` = 43.5 if `WKSWORK2` = 4 (40-47 weeks).
- `wks` = 48.5 if `WKSWORK2` = 5 (48-49 weeks).
- `wks` = 51 if `WKSWORK2` = 6 (50-52 weeks).

The variable `wks0` is similarly defined, except that it takes on value zero if the respondent did not work at least one week during the previous year (`WKSWORK2` = 0).

- **Weekly hours worked** (`hrs` and `hrs0`)

  We create measures for weekly hours worked by using the `UHRSWORK` variable. It indicates the number of hours per week that the respondent usually worked, if the respondent worked during the previous year. It is top-coded at 99 hours (this is the case for 266 respondents in the regression sample). The first measure, `hrs`, takes on a missing value if the respondent did not work during the previous year, while the second measure, `hrs0`, takes on value zero if the respondent did not work during the previous year.

A.2.2 Respondent control variables

We construct respondent characteristics from the variables in the ACS 1% samples 2005-2015 (Ruggles et al. 2015). They are used as control variables throughout the empirical analysis (except the non-labor income variable `nlabinc`). Our variables are in lowercase, while the original ACS variables are in uppercase.

- **Age** (`age`, `age_sq`, `age_2529`, `age_3034`, `age_3539`, `age_4044`, `age_4549`)

  Respondents' age is from the `AGE` variable. Throughout the empirical analysis, we further control for age squared (`age_sq`), as well as for indicators for the following age groups: 30-34 (`age_3034`), 35-39 (`age_3539`), 40-44 (`age_4044`), 45-49 (`age_4549`). The age group 25-29 (`age_2529`) is the excluded category.

- **Race and ethnicity** (`race` and `hispan`)

  We create race and ethnicity variables, `race` and `hispan`, from the `RACE` and `HISPAN` variables. Our `race` variable contains four categories: `White` (`RACE` = 1), `Black` (`RACE` = 2), `Asian` (`RACE` = 4, 5, or 6), and `Other` (`RACE` = 3, 7, 8, or 9). In the regressions, `Asian` is the excluded category. We additionally

control throughout for an indicator (`hispan`) for the respondent being of Hispanic origin (`HISPAN` > 0).

- **Education** (`educyrs` and `student`)

  We measure education with the respondents' educational attainment (`EDUC`). We translate the `EDUC` codes into years of education as follows. Note that we don't use the detailed codes `EDUCD` because they are not comparable across years (e.g. grades 7 and 8 are distinct from 2008 to 2015, but grouped from 2005 to 2007).

  - `educyrs` = 0 if `EDUC` = 0. No respondent has a value `EDUCD` = 1, i.e., there is no missing observation in the regression sample.
  - `educyrs` = 4 if `EDUC` = 1. This corresponds to nursery school through grade 4.
  - `educyrs` = 8 if `EDUC` = 2. This corresponds to grades 5 through 8.
  - Between grade 9 (`educyrs` = 9) and 4 years of college (`educyrs` = 16), each `EDUC` code provides the exact educational attainment.
  - `educyrs` = 17 if `EDUC` = 11. This corresponds to 5+ years of college.

  We also create an indicator variable, `student`, equal to one if the respondent is attending school. This information is given by the variable `SCHOOL`.

  Note that we also provide categories of educational attainment in the summary statistics tables. They are defined as follows:

  - `No scholling` if `EDUC` is 0 or 1.
  - `Elementary` if `EDUC` is 2.
  - `High School` if `EDUC` is between 3 and 6.
  - `College` if `EDUC` is between 7 and 11.

- **Years since immigration** (`yrsusa`)

  We measure the number of years since immigration by the number of years since the respondent has been living in the U.S., which is given by the variable `YRSUSA1`.

- **Age at immigration** (`ageimmig`)

  We measure the age at immigration as `ageimmig` = `YEAR` - `YRSUSA1` - `BIRTHYR`, where `YEAR` is the ACS year, `YRSUSA1` is the number of years since the respondent has been living in the U.S., and `BIRTHYR` is the respondent's year of birth.

- **Decade of immigration** (`dcimmig`)

  We compute the decade of migration by using the first two digits of `YRIMMIG`, which indicates the year in which the respondent entered the U.S. In the regressions, the decade 1950 is the excluded category.

- **English proficiency** (`englvl`)

  We build a continuous measure of English proficiency using the `SPEAKENG` variable. It provides 5 levels of English proficiency. We code our measure between 0 and 4 as follows:

  - `englvl` = 0 if `SPEAKENG` = 1 (Does not speak English).
  - `englvl` = 1 if `SPEAKENG` = 6 (Yes, but not well).
  - `englvl` = 2 if `SPEAKENG` = 5 (Yes, speaks well).
  - `englvl` = 3 if `SPEAKENG` = 4 (Yes, speaks very well).
  - `englvl` = 4 if `SPEAKENG` = 2 (Yes, speaks only English).

- **Non-labor income** (`nlabinc`)

  We construct the non-labor income variable as the difference between the respondent's total personal income (`INCTOT`) and the respondent's wage and salary income (`INCWAG`) for the previous year. Both variables are adjusted for inflation and converted to 1999 $ using the `CPI99` variable. We assign a value of zero to the `INCWAG` variable if the respondent was not working during the previous year. Note that total personal income (`INCTOT`) is bottom coded at -$19,998 in the ACS. This is the case for 14 respondents in the regression sample. Wage and salary income (`INCWAG`) is also top coded at the 99.5th Percentile in the respondent's State.

A.2.3 Household control variables

We construct household characteristics from the variables in the ACS 1% samples 2005-2015 (Ruggles et al. 2015). They are used as control variables throughout the empirical analysis (except time since married, `timemarr`, and county of residence, `county`). Our variables are in lowercase, while the original ACS variables are in uppercase. These variables are common to the respondent and her husband.

- **Children aged < 5** (`nchild_l5`)

  The number of own children aged less than 5 years old in the household is given by the `NCHLT5` variable.

- **Household size** (`hhsize`)

  The number of persons in the household is given by the `NUMPREC` variable.

- **State of residence** (`state`)

  The State of residence is given by the `STATEICP` variable.

- **Time since married** (timemarr)

  The number of years since married is given by the YRMARR variable. It is only available for the years 2008-2015.

- **County of residence** (county)

  The county of residence variable, county, is constructed from the COUNTY together with the STATEICP variables. However, not all counties are identifiable. For 2005-2011, 384 counties are identifiable, while for 2012-2015, 434 counties are identifiable (see the excel spreadsheet provided by Ruggles et al. 2015). About 79.6% of the respondents in the regression sample reside in identifiable counties.

A.2.4 Husband control variables

We construct the husbands characteristics from the variables in the ACS 1% samples 2005-2015 (Ruggles et al. 2015). They are used as control variables in the household-level analyses. Our variables are in lowercase, while the original ACS variables are in uppercase. In general, the husband variables are similarly defined as the respondent variables, with the extension _sp added to the variable name. We only detail the husband control variables that differ from the respondent control variables.

- **Years since immigration** (yrsusa_sp)

  It is similarly defined as yrsusa, except we assign the value of age_sp to yrsusa_sp if the respondent's husband was born in the U.S.

- **Age at immigration** (ageimmig_sp)

  It is similarly defined as ageimmig, except we assign a value of zero to ageimmig_sp if the respondent's husband was born in the U.S.

A.2.5 Bargaining power measures

- **Age gap** (age_gap)

  The age gap between spouses is defined as the age difference between the husband and the wife: age_gap = age_sp - age.

- **Non-labor income gap** (nlabinc_gap)

  The non-labor income gap between spouses is defined as the difference between the husband's non-labor income and the wife's non-labor income: nlabinc_gap = nlabinc_sp - nlabinc. We assign a value zero to the 16 cases in which one of the non-labor income measures is missing.

A.2.6 Country of birth characteristics

- **Labor force participation**

  - **Country of birth female labor force participation** (`ratio_lfp`)

    We assign the value of female labor participation of a respondent's country of birth at the time of her decade of migration to the U.S. More precisely, we gather data for each country between 1950 and 2015 from the International Labor Organization. Then, we compute decade averages for each country—many series have missing years. Finally, for the country-decades missing values, we impute a weighted decade average of neighboring countries, where the set of neighboring countries for each country is from the `contig` variable from the dyadic `GEODIST` dataset (Mayer & Zignago 2011). To avoid measurement issues across countries, we compute the ratio of female to male participation rates in each country.

  - **Past migrants female labor force participation** (`lfp_cob`)

    We compute the female labor force participation of a respondent's ancestor migrants by computing the average female labor participation rate (`LABFORCE`) of immigrant women to the U.S. aged 25 to 49 living in married-couple family households (`HHTYPE` = 1) that are from the respondent's country of birth at the time of the census prior to the respondent's decade of migration. We use the following samples from Ruggles et al. (2015): the 1940 100% Population Database, the 1950 1% sample, the 1960 1% sample, the 1970 1% state fm1 sample, the 1980 5% sample, the 1990 5% sample, the 2000 5% sample, and the 2014 5-years ACS 5% sample for the 2010 decade.

- **Education**

  - **Country of birth female education** (`yrs_sch_ratio`)

    We assign the value of female years of schooling of a respondent's country of birth at the time of her year of migration to the U.S. More precisely, we gather data for female and male years of schooling aged 15 and over for each country between 1950 and 2010 from the Barro-Lee Educational Attainment Dataset (Barro & Lee 2013), which provides five years averages. The original data is available here and here. We impute a weighted average of neighboring countries for the few missing countries, where the set of neighboring countries for each country is from the `contig` variable from the dyadic `GEODIST` dataset (Mayer & Zignago 2011). Finally, to avoid measurement issues across countries, we compute the ratio of female to male years of schooling in each country.

  - **Past migrants female education** (`educyrs_cob`)

    We compute the female years of schooling of a respondent's ancestor migrants by computing the average female educational attainment (`EDUC`) of immigrant women to the U.S. aged 25 to 49 living in married-couple family households (`HHTYPE` = 1) that are from the respondent's country

of birth at the time of the census prior to the respondent's decade of migration. We use the same code conversion convention as with the `educyrs` variable. We use the following samples from Ruggles et al. (2015): the 1940 100% Population Database, the 1950 1% sample, the 1960 1% sample, the 1970 1% state fm1 sample, the 1980 5% sample, the 1990 5% sample, the 2000 5% sample, and the 2014 5-years ACS 5% sample for the 2010 decade.

- **Fertility** (`tfr`)

  We assign the value of total fertility rate of a respondent's country of birth at the time of her decade of migration to the U.S. More precisely, we gather data for each country between 1950 and 2015 from the Department of Economic and Social Affairs of the United Nations (DESA 2015), which provides five years averages. The original data is available here. We impute a weighted average of neighboring countries for the few missing countries, where the set of neighboring countries for each country is from the `contig` variable from the dyadic `GEODIST` dataset (Mayer & Zignago 2011).

- **Net migration rate** (`mig`)

  We assign the value of net migration rate of a respondent's country of birth at the time of her decade of migration to the U.S. More precisely, we gather data for each country between 1950 and 2015 from the Department of Economic and Social Affairs of the United Nations (DESA 2015), which provides five years averages. The original data is available here. We impute a weighted average of neighboring countries for the few missing countries, where the set of neighboring countries for each country is from the `contig` variable from the dyadic `GEODIST` dataset (Mayer & Zignago 2011).

- **Geography**

  - **Latitude** (`lat`)

    The latitude (in degrees) of the respondent's country of birth is the `lat` variable at the level of the country's capital from the country-specific `GEODIST` dataset (Mayer & Zignago 2011). The original data is available here.

  - **Longitude** (`lon`)

    The longitude (in degrees) of the respondent's country of birth is the `lon` variable at the level of the country's capital from the country-specific `GEODIST` dataset (Mayer & Zignago 2011). The original data is available here.

  - **Continent** (`continent`)

    The `continent` variable takes on five values: `Africa`, `America`, `Asia`, `Europe`, and `Pacific`. It is the `continent` variable from the country-specific `GEODIST` dataset (Mayer & Zignago 2011). The original data is available here. When the `continent` variable is used in the regressions, `Africa` is the excluded category.

– **Bilateral distance** (`distcap`)

The `distcap` variable is the bilateral distance in kilometers between the respondent's country of birth capital and Washington, D.C. It is the `distcap` variable from the dyadic `GEODIST` dataset (Mayer & Zignago 2011). The original data is available here.

- **GDP per capita** (`gdpko`)

We assign the value of the GDP per capita of a respondent's country of birth at the time of her decade of migration to the U.S. More precisely, we gather data for each country between 1950 and 2014 from the Penn World Tables (Feenstra et al. 2015). We build the `gdpko` variable by dividing the output-side real GDP at chained PPPs in 2011 US$ (`rgdpo`) by the country's population (`pop`). The original data is available here. We impute a weighted average of neighboring countries for the few missing countries, where the set of neighboring countries for each country is from the `contig` variable from the dyadic `GEODIST` dataset (Mayer & Zignago 2011).

- **Genetic distance** (`gendist_weight`)

We assign to each respondent the population weighted genetic distance between her country of birth and the U.S., as given by the `new_gendist_weighted` variable from Spolaore & Wacziarg (2016). The original data is available here.

- **Common language** (`comlang_ethno`)

The `comlang_ethno` variable is an indicator variable equal to one if a language is spoken by at least 9% of the population both in the respondent's country of birth and in the U.S. It is the `comlang_ethno` variable from the dyadic `GEODIST` dataset (Mayer & Zignago 2011). The original data is available here.

A.2.7 County-level variables

- **County-level country of birth density** (`sh_bpl_w`)

The `sh_bpl_w` variable measures the density of immigrant workers from the same country of birth as a respondent in her county of residence. For a respondent from country c residing in county e, it is computed as

$$\texttt{sh\_bpl\_w}_{ce} = \frac{\text{Immigrants from country } c \text{ in county } e}{\text{Immigrants in county } e} \times 100$$

where the sample is the pooled 2005-2015 1% ACS samples (Ruggles et al. 2015) of immigrants above age 16 of both sexes that are in the labor force.

- **County-level language density** (`sh_lang_w`)

The `sh_lang_w` variable measures the density of immigrant workers speaking the same language as a respondent in her county of residence. For a respondent speaking language l residing in county e, it is computed as

$$\texttt{sh\_lang\_w}_{le} = \frac{\text{Immigrants speaking language } l \text{ in county } e}{\text{Immigrants in county } e} \times 100$$

where the sample is the pooled 2005-2015 1% ACS samples (Ruggles et al. 2015) of non-English speaking immigrants above age 16 of both sexes that are in the labor force..

B Missing language structures

While our main source for language structures is WALS (Dryer & Haspelmath 2011), we complement our dataset from other source in collaboration with linguists, as well as with Mavisakalyan (2015). We scrupulously followed the definitions in WALS (Dryer & Haspelmath 2011) when inputting missing values. The linguistic sources used for each language, as well as the values inputted are available in the data repository. The languages for which at least one of the four linguistic variables were inputted by linguists are the following: Albanian, Bengali, Bulgarian, Cantonese, Czech, Danish, Gaelic, Guarati, Italian, Japanese, Kurdish, Lao, Malayalam, Marathi, Norwegian, Panjabi, Pashto, Polish, Portuguese, Romanian, Serbian-Croatian, Swahili, Swedish, Taiwanese, Tamil, Telugu, Ukrainian, Urdu, and Yiddish. Mavisakalyan (2015) was used to input the `GP` variable for the following languages: Fula, Macedonian, and Uzbek.

C Appendix tables

Table C.1. Language Distribution by Family in the Regression Sample

Language	Genus	Respondents	Percent
Afro-Asiatic			
Beja	Beja	716	0.15
Arabic	Semitic	9,567	1.99
Amharic	Semitic	2,253	0.47
Hebrew	Semitic	1,718	0.36
	Total	14,254	2.97
Altaic			
Khalkha	Mongolic	3	0.00
Turkish	Turkic	1,872	0.39
Uzbek	Turkic	63	0.01
	Total	1,938	0.40
Austro-Asiatic			
Khmer	Khmer	2,367	0.49
Vietnamese	Viet-Muong	18,116	3.77
	Total	20,483	4.26
Austronesian			
Chamorro	Chamorro	9	0.00
Tagalog	Greater Central Philippine	27,200	5.66
Indonesian	Malayo-Sumbawan	2,418	0.50
Hawaiian	Oceanic	7	0.00
	Total	29,634	6.17
Dravidian			
Telugu	South-Central Dravidian	6,422	1.34
Tamil	Southern Dravidian	4,794	1.00
Malayalam	Southern Dravidian	3,087	0.64
Kannada	Southern Dravidian	1,279	0.27
	Total	15,582	3.24
Niger-Congo			
Swahili	Bantoid	865	0.18
Fula	Northern Atlantic	175	0.04
	Total	1,040	0.22
Sino-Tibetan			
Tibetan	Bodic	1,724	0.36
Burmese	Burmese-Lolo	791	0.16
Mandarin	Chinese	34,878	7.26
Cantonese	Chinese	5,206	1.08
Taiwanese	Chinese	908	0.19
	Total	43,507	9.05
Zuni	Zuni	1	0.00
Indo-European			
Albanian	Albanian	1,650	0.34
Armenian	Armenian	2,386	0.50
Latvian	Baltic	524	0.11
Gaelic	Celtic	118	0.02
German	Germanic	5,738	1.19
Dutch	Germanic	1,387	0.29
Swedish	Germanic	716	0.15
Danish	Germanic	314	0.07
Norwegian	Germanic	213	0.04
Yiddish	Germanic	168	0.03
Greek	Greek	1,123	0.23
Hindi	Indic	12,233	2.55
Urdu	Indic	5,909	1.23
Gujarati	Indic	5,891	1.23
Bengali	Indic	4,516	0.94
Panjabi	Indic	3,454	0.72
Marathi	Indic	1,934	0.40
Persian	Iranian	4,508	0.94
Pashto	Iranian	278	0.06
Kurdish	Iranian	203	0.04
Spanish	Romance	243,703	50.71
Portuguese	Romance	8,459	1.76
French	Romance	6,258	1.30
Romanian	Romance	2,674	0.56
Italian	Romance	2,383	0.50
Russian	Slavic	12,011	2.50
Polish	Slavic	6,392	1.33
Serbian-Croatian	Slavic	3,289	0.68
Ukrainian	Slavic	1,672	0.35
Bulgarian	Slavic	1,159	0.24
Czech	Slavic	543	0.11
Macedonian	Slavic	265	0.06
	Total	342,071	71.17
Tai-Kadai			
Thai	Kam-Tai	2,433	0.51
Lao	Kam-Tai	1,773	0.37
	Total	4,206	0.88
Uralic			
Finnish	Finnic	249	0.05
Hungarian	Ugric	856	0.18
	Total	1,105	0.23
Japanese	Japanese	6,793	1.41
Aleut	Aleut	4	0.00

Table C.2. Summary Statistics, No Sample Selection
Replication of Table 1

	Mean	S.d.	Min.	Max.	Obs.	Difference
	A. Individual characteristics					
Age	37.7	6.6	25	49	515,572	-0.05***
Years since immigration	14.7	9.3	0	50	515,572	0.10***
Age at immigration	23.0	8.9	0	49	515,572	-0.15***
Educational Attainment						
Current student	0.07	0.25	0	1	515,572	-0.00***
Years of schooling	12.5	3.7	0	17	515,572	-0.09***
No schooling	0.05	0.21	0	1	515,572	0.00***
Elementary	0.11	0.32	0	1	515,572	0.01***
High school	0.36	0.48	0	1	515,572	0.00***
College	0.48	0.50	0	1	515,572	-0.01***
Race and ethnicity						
Asian	0.31	0.46	0	1	515,572	-0.02***
Black	0.04	0.20	0	1	515,572	-0.02***
White	0.45	0.50	0	1	515,572	0.03***
Other	0.20	0.40	0	1	515,572	0.01***
Hispanic	0.49	0.50	0	1	515,572	0.03***
Ability to speak English						
Not at all	0.11	0.31	0	1	515,572	0.01***
Not well	0.24	0.43	0	1	515,572	0.00**
Well	0.25	0.43	0	1	515,572	-0.00***
Very well	0.40	0.49	0	1	515,572	-0.00***
Labor market outcomes						
Labor participant	0.61	0.49	0	1	515,572	-0.00
Employed	0.55	0.50	0	1	515,572	-0.00
Yearly weeks worked (excl. 0)	44.2	13.2	7	51	325,148	-0.01
Yearly weeks worked (incl. 0)	27.2	23.9	0	51	515,572	-0.05
Weekly hours worked (excl. 0)	36.6	11.4	1	99	325,148	-0.03
Weekly hours worked (incl. 0)	22.6	19.9	0	99	515,572	-0.05
Labor income (thds.)	25.5	28.6	0.0	536.5	303,167	-0.11
	B. Household characteristics					
Years since married	12.3	7.5	0	39	377,361	0.05***
Number of children aged < 5	0.42	0.65	0	6	515,572	-0.00
Number of children	1.86	1.26	0	9	515,572	0.01***
Household size	4.25	1.61	2	20	515,572	0.02***
Household income (thds.)	61.3	59.9	-31.0	1,530.3	515,572	-0.21

Table C.2 notes: The summary statistics are computed using sample weights (`PERWT`) provided in the ACS (Ruggles et al. 2015), except those for household characteristics, which are computed using household weights (`HHWT`). The last column reports the difference in means between the uncorrected sample and the regression sample, along with stars indicating the significance level of a t-test of differences in means. See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

Table C.3. Summary Statistics, Husbands
Replication of Table 1

	Mean	S.d.	Min.	Max.	Obs.	SB1-SB0
	A. Individual characteristics					
Age	41.1	8.3	15	95	480,618	-1.8***
Years since immigration	14.5	11.1	0	83	480,618	-3.2***
Age at immigration	19.9	12.2	0	85	480,618	-4.6***
Educational Attainment						
Current student	0.04	0.20	0	1	480,618	-0.02***
Years of schooling	12.4	3.8	0	17	480,618	-1.6***
No schooling	0.05	0.22	0	1	480,618	0.01***
Elementary	0.12	0.33	0	1	480,618	0.10***
High school	0.36	0.48	0	1	480,618	0.10***
College	0.46	0.50	0	1	480,618	-0.21***
Race and ethnicity						
Asian	0.25	0.43	0	1	480,618	-0.68***
Black	0.03	0.16	0	1	480,618	0.01***
White	0.52	0.50	0	1	480,618	0.44***
Other	0.21	0.40	0	1	480,618	0.22***
Hispanic	0.51	0.50	0	1	480,618	0.59***
Ability to speak English						
Not at all	0.06	0.24	0	1	480,618	0.02***
Not well	0.20	0.40	0	1	480,618	0.02***
Well	0.26	0.44	0	1	480,618	-0.08***
Very well	0.48	0.50	0	1	480,618	0.04***
Labor market outcomes						
Labor participant	0.94	0.24	0	1	480,598	0.02***
Employed	0.90	0.30	0	1	480,598	0.02***
Yearly weeks worked (excl. 0)	47.9	8.8	7	51	453,262	-0.2***
Yearly weeks worked (incl. 0)	45.3	13.8	0	51	480,618	1.1***
Weekly hours worked (excl. 0)	42.9	10.3	1	99	453,262	-0.1
Weekly hours worked (incl. 0)	40.5	13.9	0	99	480,618	1.0***
Labor income (thds.)	40.6	43.7	0.0	536.5	419,762	-10.1***
	B. Household characteristics					
Years since married	12.4	7.5	0	39	352,059	-0.54***
Number of children aged < 5	0.42	0.65	0	6	480,618	0.04***
Number of children	1.87	1.26	0	9	480,618	0.30***
Household size	4.26	1.62	2	20	480,618	0.28***
Household income (thds.)	61.0	59.6	-31.0	1,530.3	480,618	-15.9***

Table C.3 notes: The summary statistics are computed using husbands sample weights (`PERWT_SP`) provided in the ACS (Ruggles et al. 2015), except those for household characteristics, which are computed using household weights (`HHWT`). The last column reports the estimate $\widehat{\beta}$ from regressions of the type $X_i = \alpha + \beta\ \texttt{SBI} + \varepsilon$, where X_i is an individual level characteristic, where `SBI` is the wife's. Robust standard errors are not reported. See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

Table C.4. Summary Statistics
Households with Female Immigrants, Married, Spouse Present, Aged 25-49
H: husband; W: wife

	Definition	Mean	S.d.	Min.	Max.	Obs.	SB1-SB0
Age gap	H age - W age	3.5	5.7	-33	64	480,618	-0.9***
American husband	H = U.S. citizen	0.18	0.38	0	1	480,618	0.03***
White husband	H = white and W $\neq$ white	0.05	0.23	0	1	480,618	-0.05***
Origin homophily	H COB = W COB	0.73	0.44	0	1	480,618	-0.00
Linguistic homophily	H language = W language	0.87	0.34	0	1	480,618	0.06***
Education gap	H education - W education	0.05	3.05	-17	17	480,618	-0.43***
Labor participation gap	H LFP - W LFP	0.34	0.54	-1	1	480,598	0.13***
Employment gap	H employment - W employment	0.35	0.59	-1	1	480,598	0.14***
Weeks gap	H weeks worked - W weeks worked	3.7	15.3	-44	44	284,275	1.4***
Hours gap	H hours worked - W hours worked	6.4	14.2	-93	97	284,275	1.6***
Labor income gap (thds.)	H labor income - W labor income	15.5	40.9	-426	521	250,164	-1.8***
Non labor participation gap (thds.)	H non labor income - W non labor income	2.9	19.3	-414	515	480,618	-0.4***

Table C.4 notes: The summary statistics are computed using household sample weights (`HHWT`) provided in the ACS (Ruggles et al. 2015). The last column reports the estimate $\widehat{\beta}$ from regressions of the type $X_i = \alpha + \beta$ `SBI` $+ \varepsilon$, where X_i is an individual level characteristic. Robust standard errors are not reported. See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

Table C.5. Summary Statistics, Country and County-Level Variables
Female Immigrants, Married, Spouse Present, Aged 25-49

Variable	Unit	Mean	S.d.	Min.	Max.	Obs.
COB LFP	Ratio female to male, %	53.25	18.20	4	118	474,003
Ancestry LFP	%	55.24	12.11	0	100	467,378
COB fertility	Number of children	3.14	1.27	1	9	479,923
Ancestry fertility	Number of children	2.08	0.57	1	4	467,378
COB education	Ratio female to male, %	86.34	15.13	7	122	480,618
Ancestry educaiton	Years of schooling	11.19	2.52	1	17	467,378
COB migration rate	%	-2.21	4.21	-63	109	479,923
COB GDP per capita	PPP 2011 US$	8,840	11,477	133	3,508,538	469,718
COB and U.S. common language	Indicator	0.71	0.45	0	1	480,618
Genetic distance COB to U.S.		0.03	0.01	0.01	0.06	480,618
Bilateral distance COB to U.S.	Km	6,792	4,315	737	16,371	480,618
COB latitude	Degrees	22.01	14.57	-44	64	480,618
COB longitude	Degrees	-16.03	88.94	-175	178	480,618
County COB density	%	22.25	26.00	0	94	382,556
County language density	%	32.72	30.26	0	96	382,556

Table C.5 notes: The summary statistics are computed using sample weights (`PERWT`) provided in the ACS (Ruggles et al. 2015). See appendix A.2.6 for more details on variables sources and definitions.

Table C.6. Gender in Language and Economic Participation, Individual Level
Replication of Columns 1 and 2 of Table 3

Dependent variable:	Labor force participation				
	(1)	(2)	(3)	(4)	(5)
Sex-based	-0.101***	-0.076***	-0.096***	-0.070***	-0.069***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.003]
β-coef.	-0.101	-0.076	-0.096	-0.070	-0.069
Years of schooling		0.019***		0.020***	0.010***
		[0.000]		[0.000]	[0.000]
β-coef.		0.072		0.074	0.037
Number of children < 5			-0.135***	-0.138***	-0.110***
			[0.001]	[0.001]	[0.001]
β-coef.			-0.087	-0.089	-0.071
Respondent char.	No	No	No	No	Yes
Household char.	No	No	No	No	Yes
Observations	480,618	480,618	480,618	480,618	480,618
R^2	0.006	0.026	0.038	0.060	0.110
Mean	0.60	0.60	0.60	0.60	0.60
`SB` residual variance	0.150	0.148	0.150	0.148	0.098

Table C.6 notes: The estimates are computed using sample weights (`PERWT`) provided in the ACS (Ruggles et al. 2015). See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

Table C.7. Gender in Language and Economic Participation, Additional Outcomes
Replication of Column 5 of Table 3, Language Indices

	Extensive margin		Intensive margins			
			Including zeros		Excluding zeros	
Dependent variable:	LFP	Employed	Weeks	Hours	Weeks	Hours
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. `Intensity_1` = `(SB + GP + GA + NG) × SB`					
`Intensity_1`	-0.009***	-0.010***	-0.43***	-0.342***	-0.20***	-0.160**
	[0.002]	[0.002]	[0.10]	[0.085]	[0.07]	[0.064]
Observations	478,883	478,883	478,883	478,883	301,386	301,386
R^2	0.132	0.135	0.153	0.138	0.056	0.034
	Panel B. `Intensity_2` = `(SB + GP + NG) × SB`					
`Intensity_2`	-0.013***	-0.015***	-0.60***	-0.508***	-0.23**	-0.214**
	[0.003]	[0.003]	[0.14]	[0.119]	[0.10]	[0.090]
Observations	478,883	478,883	478,883	478,883	301,386	301,386
R^2	0.132	0.135	0.153	0.138	0.056	0.034
	Panel C. `Intensity` = `(GP + GA + NG) × SB`					
`Intensity`	-0.010***	-0.012***	-0.51***	-0.416***	-0.26***	-0.222***
	[0.003]	[0.003]	[0.12]	[0.105]	[0.09]	[0.076]
Observations	478,883	478,883	478,883	478,883	301,386	301,386
R^2	0.132	0.135	0.153	0.138	0.056	0.034
	Panel D. `Intensity_PCA`					
`Intensity_PCA`	-0.008***	-0.009***	-0.39***	-0.314***	-0.19***	-0.150**
	[0.002]	[0.002]	[0.09]	[0.080]	[0.06]	[0.060]
Observations	478,883	478,883	478,883	478,883	301,386	301,386
R^2	0.132	0.135	0.153	0.138	0.056	0.034
	Panel E. `Top_Intensity` = 1 if `Intensity` = 3, = 0 otherwise					
`Top_Intensity`	-0.019**	-0.013	-0.44	-0.965***	0.12	-0.849***
	[0.009]	[0.009]	[0.42]	[0.353]	[0.31]	[0.275]
Observations	478,883	478,883	478,883	478,883	301,386	301,386
R^2	0.132	0.135	0.153	0.138	0.056	0.034
Respondent char.	Yes	Yes	Yes	Yes	Yes	Yes
Household char.	Yes	Yes	Yes	Yes	Yes	Yes
Respondent COB FE	Yes	Yes	Yes	Yes	Yes	Yes
Mean	0.60	0.55	27.2	22.51	44.2	36.57

Table C.7 notes: The estimates are computed using sample weights (`PERWT`) provided in the ACS (Ruggles et al. 2015). See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

In the main analysis we use the following measure of intensity of female and male distinctions in the language grammar: $\texttt{Intensity} = \texttt{SB} \times (\texttt{GP} + \texttt{GA} + \texttt{NG})$, where `Intensity` is a categorical variable that ranges from 0 to 3. The reason why we chose this specification, instead of a purely additive specification, such as in Gay et al. (2013), is that in a non-sex-based gender system, the agreements in a sentence do not relate to female / male categories. They may instead depend on whether the type of noun of animate or inanimate, human or non-human. When a grammatical gender system is organized around a construct other

than biological gender, $\texttt{SB} = 0$. In this case, $\texttt{Intensity} = 0$, since SB enters multiplicatively in the equation. In these cases, the intensity of female and male distinctions in the grammar is zero, since such distinctions are not an organizing principle of the grammatical gender system. When a grammatical gender system is sex-based, $\texttt{SB} = 1$. In this case, the intensity measure can range from 0 to 3, depending on how the rest of rules of the system force speakers to code female / male distinctions.

Appendix Table C.7 replicates the column (5) of Table 3 for all labor outcome variables using these alternative measures. Intensity_1 is defined as $\texttt{Intensity\_1} = \texttt{SB} \times (\texttt{SB} + \texttt{GP} + \texttt{GA} + \texttt{NG})$, where Intensity is a categorical variable that ranges from 0 to 4. The only difference with Intensity is that it assigns a value of 1 for a language with a sex based language but no intensity in the other individual variables. Panel A presents results using Intensity_1. Intensity_2 is defined as $\texttt{Intensity\_2} = \texttt{SB} \times (\texttt{SB} + \texttt{GP} + \texttt{NG})$, where Intensity_2 is a categorical variable that ranges from 0 to 3. The only difference with Intensity is that it excludes the variable GA since this individual variable is not available for a number of languages. Panel B presents results using Intensity_2. Intensity_PCA is the principal component of the four individual variables SB, GP, GA and NG. Panel D presents results using Intensity_PCA. Top_Intensity is a dummy variable equal to 1 if Intensity is equal to 3 (corresponding to the highest intensity) and 0 otherwise.[3] Panel E presents results using Top_Intensity. These alternative measure should not be taken as measuring absolute intensity but rather as a ranking of relative intensity across languages grammar. As Appendix Table C.7 shows, our main results are robust to these alternative specifications, providing reassuring evidence that the specification we use is not driving the results.

[3]We thanks an anonymous referee suggesting this alternative specification.

Table C.8. Gender in Language and Economic Participation, Alternative Samples
Replication of Column 5 of Table 3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. Labor force participation						
Sex-based	-0.027***	-0.029***	-0.045***	-0.073***	-0.027***	-0.028***	-0.026***
	[0.007]	[0.006]	[0.017]	[0.005]	[0.007]	[0.006]	[0.008]
Observations	480,618	652,812	411,393	555,527	317,137	723,899	465,729
R^2	0.132	0.128	0.137	0.129	0.124	0.118	0.135
Mean	0.60	0.61	0.60	0.62	0.66	0.67	0.60
	Panel B. Employed						
Sex-based	-0.028***	-0.030***	-0.044***	-0.079***	-0.029***	-0.029***	-0.027***
	[0.007]	[0.006]	[0.017]	[0.005]	[0.007]	[0.006]	[0.008]
Observations	480,618	652,812	411,393	555,527	317,137	723,899	465,729
R^2	0.135	0.131	0.140	0.131	0.126	0.117	0.138
Mean	0.55	0.56	0.55	0.57	0.61	0.61	0.55
	Panel C. Weeks worked, including zeros						
Sex-based	-1.01***	-1.28***	-1.43*	-3.88***	-1.06***	-1.24***	-0.857**
	[0.34]	[0.29]	[0.82]	[0.25]	[0.34]	[0.28]	[0.377]
Observations	480,618	652,812	411,393	555,527	317,137	723,899	465,729
R^2	0.153	0.150	0.159	0.149	0.144	0.136	0.157
Mean	27.2	27.4	27.0	27.9	30.2	30.0	27.12
	Panel D. Hours worked, including zeros						
Sex-based	-0.813***	-1.019***	-0.578	-2.969***	-0.879***	-1.014***	-0.728**
	[0.297]	[0.255]	[0.690]	[0.213]	[0.297]	[0.247]	[0.328]
Observations	480,618	652,812	411,393	555,527	317,137	723,899	465,729
R^2	0.138	0.133	0.146	0.134	0.131	0.126	0.142
Mean	22.51	22.64	22.31	23.10	24.93	24.89	22.47
	Panel E. Weeks worked, excluding zeros						
Sex-based	-0.24	-0.36*	-0.85*	-1.24***	-0.22	-0.20	-0.034
	[0.24]	[0.20]	[0.51]	[0.16]	[0.24]	[0.19]	[0.274]
Observations	302,653	413,396	258,080	357,049	216,919	491,369	292,959
R^2	0.056	0.063	0.057	0.054	0.054	0.048	0.058
Mean	44.2	44.4	44.1	44.3	44.9	44.6	44.13
	Panel F. Hours worked, excluding zeros						
Sex-based	-0.165	-0.236	0.094	-0.597***	-0.204	-0.105	-0.078
	[0.229]	[0.197]	[0.524]	[0.154]	[0.227]	[0.188]	[0.260]
Observations	302,653	413,396	258,080	357,049	216,919	491,369	292,959
R^2	0.034	0.032	0.034	0.033	0.039	0.035	0.035
Mean	36.57	36.63	36.39	36.71	37.09	36.99	36.56
Sample	Baseline	Age 15-59	Indig. lang.	Include English	Exclude Mexicans	All hh	Qual. flags
Respondent char.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Household char.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Respondent COB FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table C.8 notes: The estimates are computed using sample weights (`PERWT`) provided in the ACS (Ruggles et al. 2015). See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

Table C.9. Gender in Language and Economic Participation, Logit and Probit
Replication of Columns 1 and 5 of Table 3

	OLS		Logit		Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. Labor force participation					
Sex-based	-0.101**	-0.027***	-0.105**	-0.029***	-0.105**	-0.029***
	[0.002]	[0.007]	[0.002]	[0.008]	[0.002]	[0.008]
Observations	480,618	480,618	480,618	480,612	480,618	480,612
R^2	0.006	0.132	0.005	0.104	0.005	0.104
Mean	0.60	0.60	0.60	0.60	0.60	0.60
	Panel B. Employed					
Sex-based	-0.115**	-0.028***	-0.118**	-0.032***	-0.118**	-0.032***
	[0.002]	[0.007]	[0.002]	[0.008]	[0.002]	[0.008]
Observations	480,618	480,618	480,618	480,612	480,618	480,612
R^2	0.008	0.135	0.006	0.104	0.006	0.104
Mean	0.55	0.55	0.55	0.55	0.55	0.55
Respondent char.	No	Yes	No	Yes	No	Yes
Household char.	No	Yes	No	Yes	No	Yes
Respondent COB FE	No	Yes	No	Yes	No	Yes

Table C.9 notes: This table replicates table 3 with additional outcomes and with different estimators. The estimates are computed using sample weights (`PERWT`) provided in the ACS (Ruggles et al. 2015). See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

Table C.10. Gender in Language and Economic Participation, Individual Level
Husbands of Female Immigrants, Married, Spouse Present, 25-49 (2005-2015)

Dependent variable:	Labor force participation		
	(1)	(2)	(3)
Sex-based	0.026***	0.009***	-0.004
	[0.001]	[0.002]	[0.004]
Husband char.	No	Yes	Yes
Household char.	No	Yes	Yes
Husband COB FE	No	No	Yes
Observations	385,361	385,361	384,327
R^2	0.002	0.049	0.057
Mean	0.94	0.94	0.94

Table C.10 notes: The estimates are computed using sample weights (`PERWT_SB`) provided in the ACS (Ruggles et al. 2015). See appendix A for details on variable sources and definitions. The husband characteristics are analogous to the respondent characteristics in Table 3. The sample is the same as in Table 3 except that we drop English-speaking and native husbands.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

Table C.11. Gender in Language and Economic Participation
Replication of Column 5 of Table 3

Dependent variable	Labor force participation			
	(1)	(2)	(3)	(4)
Sex-based	-0.028***		-0.025***	-0.046***
	[0.006]		[0.006]	[0.006]
Unmarried		0.143***	0.143***	0.078***
		[0.001]	[0.001]	[0.003]
Sex-based × unmarried				0.075***
				[0.004]
Respondent char.	Yes	Yes	Yes	Yes
Household char.	Yes	Yes	Yes	Yes
Respondent COB FE	Yes	Yes	Yes	Yes
Observations	723,899	723,899	723,899	723,899
R^2	0.118	0.135	0.135	0.136
Mean	0.67	0.67	0.67	0.67

Table C.11 notes: The estimates are computed using sample weights (`PERWT`) provided in the ACS (Ruggles et al. 2015). See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

Table C.12. Language and Household Bargaining
Replication of Column 3 of Table 5

	Extensive margin		Intensive margins			
			Including zeros		Excluding zeros	
Dependent variable:	LFP	Employed	Weeks	Hours	Weeks	Hours
	(1)	(2)	(3)	(4)	(5)	(6)
Sex-based	-0.027***	-0.026***	-1.095***	-0.300	-0.817**	-0.133
	[0.009]	[0.009]	[0.412]	[0.295]	[0.358]	[0.280]
β-coef.	-0.027	-0.028	-0.047	-0.020	-0.039	-0.001
Age gap	-0.004***	-0.004***	-0.249***	-0.098**	-0.259***	-0.161***
	[0.001]	[0.001]	[0.051]	[0.039]	[0.043]	[0.032]
β-coef.	-0.020	-0.021	-0.056	-0.039	-0.070	-0.076
Non-labor income gap	-0.001***	-0.001***	-0.036***	-0.012***	-0.034***	-0.015***
	[0.000]	[0.000]	[0.004]	[0.002]	[0.004]	[0.003]
β-coef.	-0.015	-0.012	-0.029	-0.017	-0.032	-0.025
Age gap × SB	0.000	-0.000	0.005	0.008	0.024	0.036**
	[0.000]	[0.000]	[0.022]	[0.015]	[0.019]	[0.015]
β-coef.	0.002	-0.001	0.001	0.003	0.006	0.017
Non-labor income gap × SB	-0.000***	-0.000***	-0.018***	0.002	-0.015***	-0.001
	[0.000]	[0.000]	[0.005]	[0.003]	[0.005]	[0.004]
β-coef.	-0.008	-0.008	-0.014	0.003	-0.014	-0.001
Respondent char.	Yes	Yes	Yes	Yes	Yes	Yes
Household char.	Yes	Yes	Yes	Yes	Yes	Yes
Husband char.	Yes	Yes	Yes	Yes	Yes	Yes
Respondent COB FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	409,017	409,017	409,017	250,431	409,017	250,431
R^2	0.145	0.146	0.164	0.063	0.150	0.038
Mean	0.59	0.54	26.40	44.07	21.82	36.44
SB residual variance	0.011	0.011	0.011	0.011	0.011	0.011

Table C.12 notes: The estimates are computed using sample weights (PERWT) provided in the ACS (Ruggles et al. 2015). See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

Table C.13. Linguistic Heterogeneity in the Household
Replication of Column 3 of Table 6

	Extensive margin		Intensive margins			
			Including zeros		Excluding zeros	
Dependent variable:	LFP	Employed	Weeks	Hours	Weeks	Hours
	(1)	(2)	(3)	(4)	(5)	(6)
Same × wife's sex-based	-0.070***	-0.075***	-3.764***	-2.983***	-0.935***	-0.406***
	[0.003]	[0.003]	[0.142]	[0.121]	[0.094]	[0.087]
Different × wife's sex-based	-0.044***	-0.051***	-2.142***	-1.149*	-1.466***	-0.265
	[0.014]	[0.015]	[0.702]	[0.608]	[0.511]	[0.470]
Different × husband's sex-based	-0.032***	-0.035***	-1.751***	-1.113**	-0.362	0.233
	[0.011]	[0.011]	[0.545]	[0.470]	[0.344]	[0.345]
Respondent char.	Yes	Yes	Yes	Yes	Yes	Yes
Household char.	Yes	Yes	Yes	Yes	Yes	Yes
Husband char.	Yes	Yes	Yes	Yes	Yes	Yes
Observations	387,037	387,037	387,037	387,037	237,307	237,307
R^2	0.122	0.124	0.140	0.126	0.056	0.031
Mean	0.59	0.54	26.40	21.84	44.11	36.49
`SB` residual variance	0.095	0.095	0.095	0.095	0.095	0.095

Table C.13 notes: The estimates are computed using sample weights (`PERWT`) provided in the ACS (Ruggles et al. 2015). See appendix A for details on variable sources and definitions.
*** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level

D Appendix figure

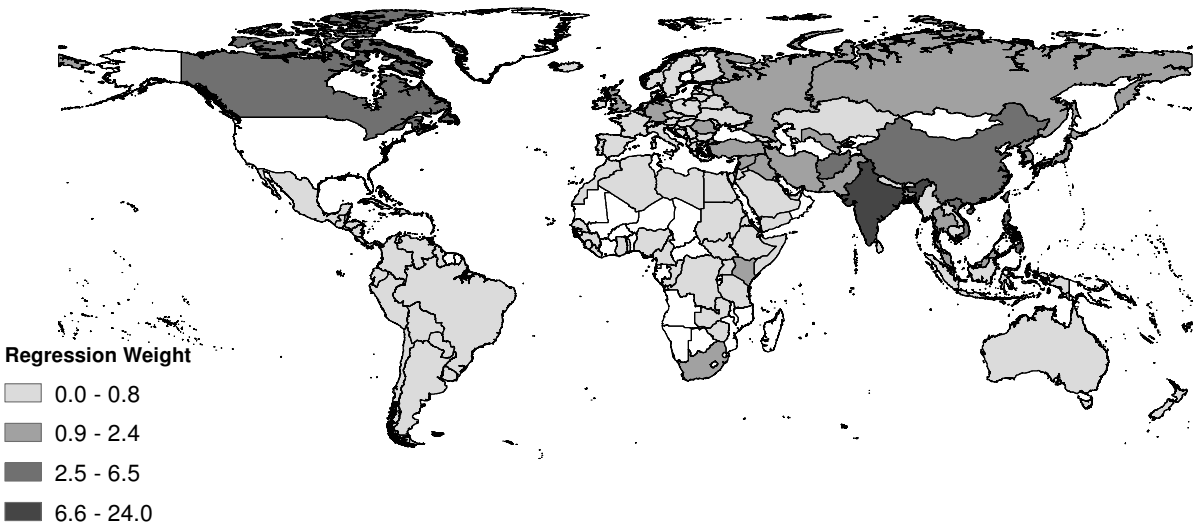

Figure D.1. Regression Weights for Table 3 column (5)

To better understand the extent to which each country of birth contributes to the identification of the coefficient in column (5) table 3, we apply Aronow & Samii's (2016) procedure to uncover the "effective sample" used in the regression. This procedure generates regression weights by computing the relative size of the residual variance of the `SB` variable for each country of birth in the sample. Appendix figure D.1 shows regression weights for each country of birth resulting from the regression of column (5) in Table 3. This specification includes respondent country of birth fixed effects. Therefore, the weights inform us of the countries where identification is coming from, and whether these are multilingual countries or not. As the figure shows, identification mostly comes from multilingual countries such as India, followed by the Philippines, Vietnam, China, Afghanistan, and Canada.

References

Aronow, P. M. & Samii, C. (2016), 'Does regression produce representative estimates of causal effects?', *American Journal of Political Science* **60**(1), 250–267.

Barro, R. J. & Lee, J. W. (2013), 'A new data set of educational attainment in the world, 1950–2010', *Journal of development economics* **104**, pp. 184–198.

DESA (2015), 'United nations, department of economic and social affairs, population division: World population prospects: The 2015 revision'.

Dryer, M. & Haspelmath, M. (2011), 'The world atlas of language structures online'.

Feenstra, R. C., Inklaar, R. & Timmer, M. P. (2015), 'The next generation of the penn world table', *The American Economic Review* **105**(10), pp. 3150–3182.

Gay, V., Santacreu-Vasut, E. & Shoham, A. (2013), The grammatical origins of gender roles, Technical report.

Inc., E. B. (2010), *2010 Britannica book of the year*, Vol. 35, Encyclopaedia britannica.

Mavisakalyan, A. (2015), 'Gender in language and gender in employment', *Oxford Development Studies* **43**(4), pp. 403–424.

Mayer, T. & Zignago, S. (2011), Notes on cepiis distances measures: The geodist database, Working Papers 2011-25, CEPII.

Ruggles, S., Genadek, K., Goeken, R., Grover, J. & Sobek, M. (2015), *Integrated public use microdata series: Version 6.0*.

Spolaore, E. & Wacziarg, R. (2016), Ancestry and development: New evidence.