

Supplemental Appendix

Marriage, Fertility, and Cultural Integration in Italy

Alberto Bisin and Giulia Tura

A Data and Sample Construction

We obtained restricted access to administrative Italian data at the individual level from ISTAT through its Laboratory for Elementary Data Analysis (ADELE).¹ In what follows, we start describing our data sources and variables of interest; passing then onto a discussion of the sample construction, and finally to the computation of empirical moments.

A Marriages, Fertility, Separation and Singles

Marriage. Our empirical analysis takes advantage of administrative individual-level data from the Italian Statistical Institute (ISTAT, ADELE Laboratory), covering the universe of marriages celebrated in Italy from 1995 to 2012 (over 4 million marriages). We exploit marriage records from municipal vital statistics registries to recover matching patterns by ethnic group of the spouses. Marriage records contain the universe of marriages celebrated each year in Italy from 1995 to 2012. They provide information on the main socio-demographic characteristics of the spouses. They are collected through the ISTAT model compiled by the Registrar of the City Civil State in which the marriage took place. For each marriage, the section dedicated to the wedding reports: the date of marriage, the type of ceremony (religious or civil), the municipality of the ceremony, and the choice of the property regime by the spouses (community or separation property). The information provided for each spouse includes the date of birth, municipality of birth, municipality of residence at the time of marriage, the place of future residence of the spouses, the previous marital status, the education level, the employment status, and for immigrant individuals the nationality and the country of origin. To account for the out-migration selection of families, the sample is restricted to marriages where at least one spouse is resident in Italy at the time of the marriage.

From these data, we recover the distribution of marriages, by cultural-ethnic group and education of spouses and by region; see Figure C3 for the bivariate marriage distribution.

¹Requests for accessing the data for research purposes should be addressed to ISTAT through an open application procedure. Authorized researchers can access and use the data from work stations located in secure rooms within the ISTAT offices. The output of the analysis is made available upon inspection by ADELE officers in compliance with the laws on the protection of statistical confidentiality and personal data. For further information, visit <https://www.istat.it/it/informazioni-e-servizi/per-i-ricercatori/laboratorio-adele>.

Crucially, these data account only for legal marriages and only when celebrated in Italy.²

Fertility. Fertility rates come from municipality birth registries, which contain the universe of individual birth records of residents in the municipality of enrollment, for each year from 1990 to 2012. Individual birth records include socio-demographic variables of interest such as gender, date, and province of birth, citizenship, and parental information regarding their date of birth, citizenship, and marital status. Unfortunately, these data do not provide information on the educational level of spouses.

Fertility from Census 2001 and 2011. We recover fertility by cultural-ethnic group and, importantly, spouses' education from the decennial Census of 2001 and 2011, which contain the universe of the population resident in Italy in 2001 and 2011. Individual Census data provide information about the family and the family identifier, the municipality of residence, and the main socio-demographic characteristics of individuals including gender, age, marital status, and education. For immigrant individuals, Census data provide additional information about the nationality, the country of origin, and the year of arrival in Italy. We identify marriages as heterosexual love relations between spouses who are living together and who are legally married, by exploiting information on the family id, marital status, and type of the relationship. Coherently with our main dataset, we focus on marriages formed after 1995 for Census 2001. Census 2011, instead, does not report the year of marriage. We thus select marriages of *natives* based on spouses' years of birth, i.e., we select native men born from 1955 to 1984 and native women born from 1959 to 1988. In addition, we select marriages of *immigrants* with children born after the year of arrival in Italy and we select marriages of immigrants not arriving in the same year and from the same country, as likely marriages formed outside the Italian marriage market. Even applying these selections, in this sample, fertility rates are mechanically higher compared to our main analysis. Thus, from these data, we exploit the variation in fertility by spouses' education and we calculate the fertility rate (number of children including zeros) by cultural-ethnic group and education of spouses. We then match these fertility rates with the average fertility rates derived from municipality birth registries to recover variation in fertility choices by spouses' education.

²The increase in cohabiting couples is a very recent phenomenon. In 2011, the share of cohabiting native couples was still only about 9% in Italy - compared to the averages in the EU (14.5%) and the OECD (16.7%). Also, cohabitations are less relevant for our study since they are typically less stable and hence less motivated by fertility and children socialization; see (Lundberg and Pollak, 1993; Chiappori, Salanié and Weiss, 2017). Also, marriages celebrated outside of the Italian marriage market are arguably less motivated by cultural integration issues.

Separation. Separation data come from the registries of civil court chancelleries and cover the universe of legal separations registered in Italy, covering the period 1995-2012.³ We focus on separation rates, which better represent marital dissolution decisions in the Italian context compared to divorces, for two main reasons. First, separation is the juridical act that launches the divorce proceedings. With Law 74/1987 and until 2015, a minimum period of 3 years of legal separation was required before eventually submitting a divorce request. Second, on average, only 65% of separations are followed by a divorce, which implies that divorce choices significantly underestimate marital dissolution behaviors. The data allow us to analyze various aspects of the marital dissolution phenomenon. We investigate, in particular, the custody assignment of children.⁴

Single Individuals. We derive the population vectors by ethnic group, education, gender, and marital status from individual Italian Census data of 2001 and 2011. We select adult individuals (more than 18 years old). Census data classify the marital status of an individual as: never married, at present married, separated *de facto*, legally separated, divorced, or widowed. We consider an individual available in the case that she/he is never married, legally separated, divorced, or widowed. We also discard institutional households, corresponding to correctional institutions, but also military and mental care facilities. We take into account potential measurement error concerns due to truncation of unmatched population vectors, we follow Chiappori, Salanié and Weiss (2017). Specifically, to account for the possibility that single individuals might marry in the near future, we restrict the set of single individuals to single men and women after their marriageable age, defined as the 90th percentile of the age at first marriage distribution for men and women, respectively. In our data, single rates increase quite symmetrically for all ethnic groups, from 2001 to 2011, the overall Spearman rank correlation test is as high as 0.88, and equal to 0.57 and 0.98 for available adult men and adult women, in turn, suggesting that the ethnic-group rank order remains stable over the period, especially for women. The distribution of single male and female adult individuals is reported in Table C9.

Socialization. Socialization data come from the *Condition and Social Integration of Foreign*

³For our investigation period, registries of civil court chancelleries constitute the unique source for separations data, while starting from December 2014 (in the application of Law n. 162/2014) consensual separation proceedings can be submitted to the civic registrar. This rules out potential sample selection concerns.

⁴In our model, we introduce an asymmetry between spouses in the probability of child custody assignment upon dissolution, independently from the ethnic groups h, j . From separation proceedings data, we calculate that the mother is given *effective* custody of children in 88% of the cases. We uncover some significant but small differences in custody assignment conditional on mother and father migrant status. We abstract from incorporating them in the model for the sake of simplicity.

Nationals Survey, conducted in 2011 and 2012 in all Italian regions on a sample of 9,600 households. The survey targeted foreign residents in Italy and it was conducted at the household level to provide socio-demographic information about all household members, for a total sample of 25,356 respondents.⁵ The final questionnaire was translated in 10 different languages to overcome potential language barriers and to reduce attrition. The actual survey was conducted through direct interviews supported by the CAPI (Computer Assisted Personal Interview) system to ease the development of the whole questionnaire.⁶ In each selected household, all members were interviewed, both foreign-born and natives.

We proxy the cultural-ethnic transmission with language socialization. In particular, the socialization measure we construct for our analysis is based on the *language spoken at home* by children and young adults (less than 25 years old), living with their parents at the time of the interview: an individual is socialized to the Italian language if he/she declares to speak Italian within the family; otherwise, we assume he is socialized to his mother language, defined as idiom acquired during the preschool period of childhood.⁷

We exclude from our analysis, single respondents and households without children. For our analysis, we consider children and young adults less than 25 years old, living with their parents at the time of the interview. The final sample consists of 8,007 individuals belonging to about 5,000 households, 86.7% of these households are married while the remaining are either separated or divorced. We consider the sample representative for the study of immigrant linguistic integration by ethnic group in each region of residence. We construct our measure of socialization based on the language spoken at home. The survey also provides questions to evaluate the level of Italian language proficiency and we check individual self-declared responses on language spoken.

B Dataset Construction and Empirical Moments

The empirical estimation is based on a unique dataset that links household information across different sources. We matched marriage, birth, and separation records on the exact date of marriage and spouses' date and place of birth (Italian province for natives and country of origin for foreigners), which are reported in all registries. In the birth records matching, the

⁵The survey follows a pivotal survey conducted in 5 sampled regions on a sample of 250 families with at least one foreign member. The pivotal survey was particularly useful in the definition and evaluation of the questionnaire, which also requires the participation of sociologists and cultural mediators.

⁶Examples of the questionnaire and invitation letter are available at <http://www.istat.it/it/archivio>.

⁷The three questions we exploit are framed in the survey in the following way. Language spoken at home: *In Italy, in your family, do you speak more often Italian or another language?* Mother tongue (main): *What language did you speak when you were young, before going to school?* Mother tongue (secondary): *In addition to this, did you also speak another language when you were young and which one?*

combination of these characteristics allows for an exact one-to-one matching for 98.8% of marriages, while in the separation matching, we match exactly 99.5% of marriages, and we discard the remaining fraction. Such low percentages suggest that marriages can be uniquely identified through the set of time-invariant characteristics listed above. Unfortunately, during the revision process, ISTAT changed data availability rules for researchers and we are no longer able to match marriage and fertility data. We then matched fertility moments by cultural-ethnic group and education of spouses, and region as described above. The final sample of 4.2 million marriages corresponds to 94% of the universe of marriages celebrated in Italy during the time interval 1995-2012. In the final dataset, the fertility rate corresponds to 69.6% with an average of 1.54 children per family. Of all marriages, 7% end up in separation in the first years of the marital union.

From this final sample, we recover the following empirical moments. The marital utility net of the outside options of singlehood $\hat{U}_{hj}$ for the household of type hj is identified from equation (2), exploiting the number of hj marriages formed in each region r , μ_{hj} , and the number of unmatched men of type h and women of type j for each region r , μ_{h0}, μ_{0j} . Individuals are heterogeneous both in their cultural-ethnic traits and their education level.

Fertility rates $\hat{N}_{hj}$ and separations rates $\hat{\pi}_{hj}$ for each household type hj are computed as follows (for all regions r):

$$\hat{N}_{hj} = \frac{1}{\mu_{hj}} \sum_{m=1}^{\mu_{hj}} N_{hj},$$

$$\hat{\pi}_{hj} = \frac{1}{\mu_{hj}} \sum_{b=1}^{\mu_{hj}} D_{hj},$$

with N_{hj} the number of children born from within a hj household, and D_{hj} is a dummy equal to one if the hj marriage end up in separation during the investigation period.

We compute the vector of socialization frequencies $\hat{P}_{hj}^k(d)$ for all h, j and k , conditional on being married, $d = 0$ as follow (for all regions r):

$$\hat{P}_{hj}^k(d = 0) = \frac{1}{M_{hj}} \sum_{b=1}^{M_{hj}} S_{hj}^k.$$

with M_{hj} being the number of children and young adults of less than 25 years old belonging to the hj household, and speaking language S^k .

B Education in Italy and INVALSI data

From the ISTAT data on marriages, the education levels of spouses can only be constructed by classifying as highly-educated individuals with a high school degree. For second-generations onwards, however, we have obtained administrative student-level data which allows us to construct a finer classification. These data provide information on standardized reading and math proficiency tests, as well as related survey data from the National Institute for the Evaluation of the Italian Education System (INVALSI).⁸ In what follows, we start describing the Italian educational system, our data sources, and variables of interest; passing then onto a discussion of our sample construction.

A Italian educational setting

In Italy, pupils normally enter formal schooling at the age of 6, and education is compulsory for 10 years. The Italian educational system is organized into five grades of elementary school, three grades of middle school, and five grades of high school. For each school (elementary, middle, and high school), students are assigned to classes and take all their subjects within the same class and with the same peers. In elementary and middle school, the educational curriculum is the same for all pupils and the subjects studied are the same.

High school is divided into different tracks (academic, technical, and vocational) and students freely self-select into three different tracks. The three tracks have the same duration but differ widely in terms of curriculum, difficulty, and prestige. While in principle, access to university is also possible from some schools within the vocational track, in practice, academic and technical schools offer much better educational and career prospects. Following [Carlana, La Ferrara and Pinotti \(2022\)](#), we define academic and technical schools as *high-track* schools, and we refer to vocational schools as the *low-track* ones. This early stratification in high school tracks ultimately have long-term implications for access to college and occupational careers ([Brunello and Checchi, 2007](#); [Carlana, La Ferrara and Pinotti, 2022](#)).

B INVALSI tests and questionnaire

INVALSI tests. Every year, starting from 2010, INVALSI administers standardized tests in reading and math to the entire population of Italian students. Tests are administered at various points of students' careers, specifically at the end of grades 2, 5, 6, 8, and 10. The INVALSI test is identical for all students in a given grade, it is blindly marked by an

⁸Requests for accessing the data for research purposes should be addressed to INVALSI through an open application procedure. For further information, visit <https://INVALSI-serviziostatistico.cineca.it/>.

external evaluator following a precise evaluation scheme, hence students' results are objective and fully comparable across schools in Italy. We exploit standardized test scores, with zero mean and standard deviation of one. The test consists of multiple-choice and open-ended questions, where the exact structure varies by grade.⁹

INVALSI questionnaire. Besides test scores, INVALSI data provides rich information for each student, including demographic characteristics such as year and quarter of birth, gender, citizenship acquisition, grade retention, and family background characteristics such as parents' education, migration history, macro-area of origin, employment status, and some measures of socio-economic status. Crucially for us, INVALSI also collects information from a student's questionnaire, including a specific question on the main language spoken at home by students, similar to our main analysis. The questionnaire is administered only in grade 5 (the last year of elementary school) and in grade 10 (the second year of high school).

C Sample selection

In our robustness analysis in Section IV (subsection C), we focus on students enrolled in grade 5 between school years 2012-13 and 2016-17. For these pupils, the language spoken at home is likely a choice of the parents. Thanks to a unique student identifier, we are able to follow students over time and match the scores and information of students in grade 5 with their educational career choices in grade 10. Because of the data collection scheme, we are able to track only two cohorts of students, i.e., the students enrolled in grade 5 in school years 2012-2013 and 2013-14 that we observe in grade 10 in school years 2017-2018 and 2018-19. Finally, our sample includes all students with at least one immigrant parent.

D Probability of being low-educated

We exploit INVALSI data to recover the probability that children become low-educated, i.e., enroll into *low-track* vocational schools compared to high-track schools, conditional on parents' cultural-ethnic traits and educational background, by region. We identify native and immigrant parents thanks to the country of origin, and for immigrant parents, we distinguish three cultural groups: European-EU15, Other European countries, and all other immigrant minorities are grouped together. The first two cultural groups exactly coincide with our cultural-group classification. The last group, instead, aggregates immigrants coming from very different origins and potentially different educational backgrounds. Thus, we recover variation in the probability that children become low-educated across immigrant minorities

⁹The tests are designed to align with those administered by the OECD Programme for International Student Assessment (PISA).

by exploiting the ISTAT data from the *Condition and Social Integration of Foreign Nationals* Survey, described above. Table B1 reports summary statistics of the average fraction of children with low education by ethnic group and education of spouses. Descriptively, estimates suggest that children of homogamous marriages of immigrants are less likely to enroll into high-track schools, especially so for children originating from Sub-Saharan Africa, followed by North Africa-Middle East and East Asia. Moreover, as expected, students with low-educated parents are less likely to enroll into high-track schools (columns 5 and 6), compared to students with highly-educated parents (columns 1 and 2).

As a validation exercise, in Table B2, we focus on Other European immigrants, the cultural group that coincides in the two datasets, INVALSI and ISTAT survey. We show that the average fraction of Other-European children with low education - by education of spouses - is similar when computed from INVALSI data as in panel A, or from ISTAT survey data as in panel B.

Table B1: Fraction of “Low educated” Children by Ethnic Group and Education of Spouses

	(1)	(2)	(3)	(4)	(5)	(6)
	High-high		High and Low		Low-low	
	Homog	Heterog	Homog	Heterog	Homog	Heterog
Italian	0.29	0.39	0.50	0.56	0.65	0.66
Europe-EU15	0.55	0.33	0.64	0.52	0.71	0.64
Other Europe	0.56	0.34	0.66	0.53	0.76	0.63
North Africa-Middle East	0.77	0.33	0.82	0.52	0.86	0.63
Sub-Saharan Africa	0.91	0.58	0.81	0.68	0.83	0.76
East Asia	0.77	0.59	0.85	0.67	0.83	0.76
Latin America	0.72	0.58	0.83	0.68	0.88	0.77

Notes: This Table reports the average fraction of children with low education by ethnic group of spouses. The outcome variable is an indicator of whether the child enrolls in vocational schools compared to academic and technical schools. Estimates are reported separately for homogamous (odd columns) and heterogamous (even columns) households by spouses’ education, distinguishing households with both highly-educated spouses (columns 1 and 2), households with one highly-educated spouse (columns 3 and 4), and households of both spouses with low education (columns 5 and 6). Sources: ISTAT survey data (estimates in bold); INVALSI data (other estimates for homogamous families).

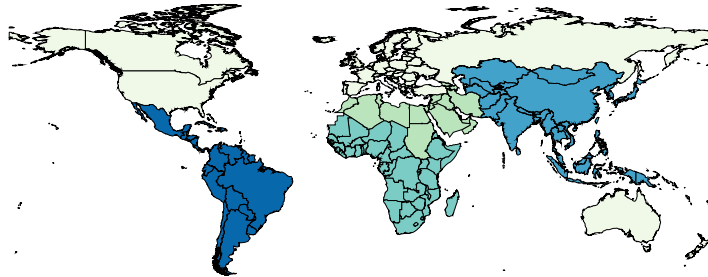
Table B2: Validation: Fraction of “Low educated” Children for Other European

	(1) Mean	(2) High Educated	(3) High-Low Educated	(4) Low Educated
Panel A. INVALSI				
Prob of low education	0.690 [0.648,0.731]	0.593 [0.543,0.643]	0.684 [0.650,0.718]	0.778 [0.740,0.817]
Panel B. ISTAT Survey 2011				
Prob of low education	0.752 [0.542,0.962]	0.592 [–]	0.812 [0.788,0.835]	0.831 [–]

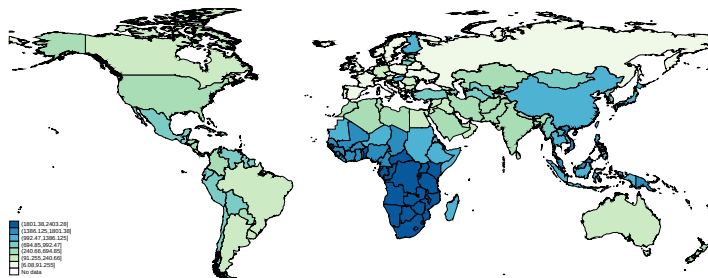
Notes: This Table reports the average fraction of children with low education for culturally homogenous Other European households, computed on INVALSI data (panel A) and ISTAT survey data (panel B). The outcome variable is an indicator of whether the child enrolls in vocational schools compared to academic and technical schools. Estimates are reported separately for households by spouses’ education, distinguishing households with both highly-educated spouses (column 2), households with one highly-educated spouse (column 3), and households of both spouses with low education (column 4).

C Additional Figures and Tables

A. Our Cultural-Ethnic Group Classification



B. Genetic Distance to Italy



C. Linguistic Distance to Italy

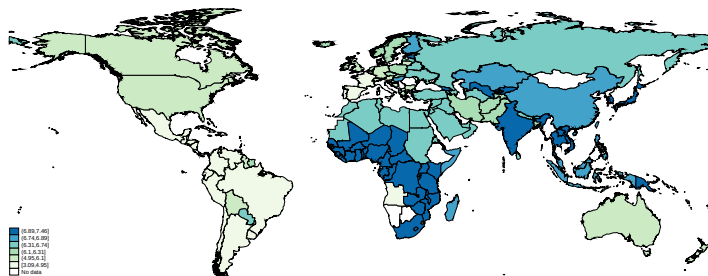


Figure C1: Ethnic-Group Classification and Cultural Distance wrt Italy

Notes: This Figure shows our classification of countries in cultural-ethnic groups (panel A) and plots the cultural distance of each country towards Italy as proxied by genetic (panel B) and ethnolinguistic distance (panel C). Data for genetic and ethnolinguistic distance are available thanks to Spolaore and Wacziarg (2016). The classification of immigrants' countries of origin by cultural-ethnic group is reported in Table C1. In particular, our classification parallels the heterogeneity in genetic distance within Africa, between the Arabic countries in the North and Sub-Saharan countries, as well as the within Asia divide between Middle-East and East Asia countries.

Table C1: Cultural-Ethnic Group Classification of Migrants' Countries of Origin

Cultural-Ethnic Group	(%)	Countries
Europe-EU15, i^E	4.57	Austria, Belgium, Denmark, France, Finland, Germany, Greece, Ireland, Luxembourg, Netherlands, Portugal, United Kingdom, Spain, Sweden
Other Europe, i^O	46.29	Albania, Andorra, Belarus, Bosnia and Herzegovina, Bulgaria, Cyprus, Croatia, Czech Republic, Estonia, Hungary, Iceland, Isle of Man, Liechtenstein, Latvia, Lithuania, Kosovo, Macedonia (FYROM), Malta, Poland, Republic of Moldova, Monaco, Norway, Russian Federation, San Marino, Vatican City State, Serbia and Montenegro, Romania, Switzerland, Slovakia, Slovenia, Turkey, Ukraine, Vatican City State, United States, Canada
North Africa-Middle East, i^M	17.15	Algeria, Egypt, Libyan Arab Jamahiriya, Morocco, Tunisia, Afghanistan, Saudi Arabia, Armenia, Azerbaijan, United Arab Emirates, Islamic Republic Of Iran, Iraq, Israel, Kazakhstan, Kyrgyzstan, Kuwait, Lebanon, Qatar, Syrian Arab Republic, Palestinian Territory, Turkmenistan, Uzbekistan
Sub-Saharan Africa, i^A	7.33	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, The Democratic Republic of Congo, Cote D'Ivoire, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Djibouti, Guinea, Guinea-Bissau, Equatorial Guinea, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, South Africa, Swaziland, United Republic of Tanzania, Togo, Uganda, Zambia, Zimbabwe
East Asia, i^S	16.47	Brunei Darussalam, Cambodia, China, Democratic People's Republic of Korea, Republic of Korea, Philippines, Japan, Jordan, Indonesia, Lao Ppeople's Democratic Republic, Malaysia, Mongolia, Myanmar, Singapore, Taiwan, Thailand, East Timor, Vietnam, Australia, Fiji, Kiribati, Marshall Islands, Federated States of Micronesia, Nauru, New Zealand, Palau, Papua New Guinea, Solomon Islands, Samoa, Tonga, Tuvalu, Vanuatu, Bahrain, Bangladesh, Bhutan, Georgia, India, Maldives, Nepal, Oman, Pakistan, Sri Lanka, Tajikistan, Yemen
Latin America, i^L	8.2	Antigua and Barbuda, Argentina, Bahamas, Barbados, Belize, Plurinational State of Bolivia, Brazil, Costa Rica, Cuba, Chile, Colombia, Dominica, Dominican Republic, Ecuador, El Salvador, Jamaica, Grenada, Guatemala, Guyana, Haiti, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Saint Kitts and Nevis, Saint Lucia, Saint Vincent and The Grenadines, Suriname, Trinidad and Tobago, Uruguay, Venezuela

Notes: This Table reports our classification of foreign countries by cultural-ethnic group. We note that the Other European group includes Eastern European countries which became EU members after the enlargements in 2004 and 2007. These enlargements altered the incentives of some immigrants to marry natives, hence the composition of intermarriages in the data is an average over the period; see [Adda, Pinotti and Tura \(2025\)](#).

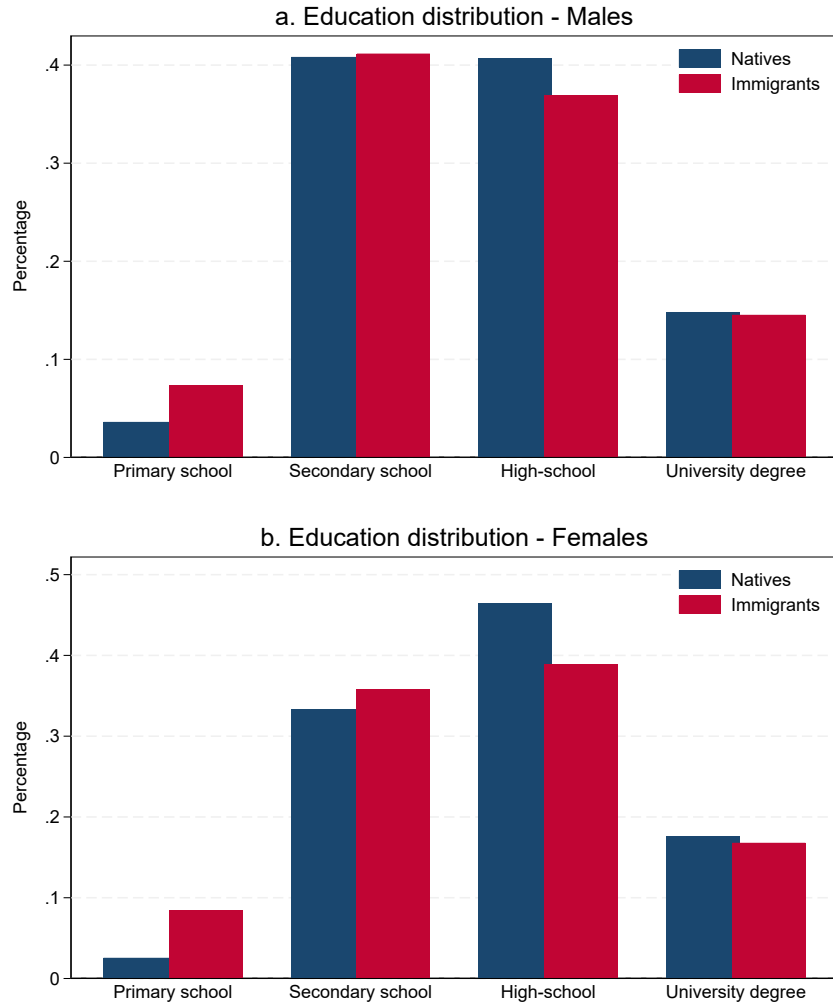


Figure C2: Education Distribution for Native and Immigrants

Notes: This Figure shows the distribution of educational attainment (in percentages) for natives and immigrants, separately for males and females in panels A and B, respectively. Educational attainment is divided into four categories: (i) primary school degree, (ii) secondary school degree, (iii) high school degree, and (iv) university degree, which includes bachelor's, master's, and PhD degrees. The figure shows that education levels are concentrated around intermediate values (lower and upper secondary), while the tails –especially the lower end (primary school only)– have limited support.

Table C2: Correlation of Marriage Outcomes and Cultural Distance Measures

	Panel A. Dependent variable: Gains from Marriage			
Genetic Distance	-5.343 (0.324)			
Linguistic Distance		-6.250 (0.139)		
Religious Distance			-6.659 (0.216)	
Cultural distance index, WVS				-8.378 (0.455)
Observations	2368	2368	2368	2368
R-squared	0.437	0.666	0.720	0.598
	Panel B. Dependent variable: Fertility rates			
Genetic Distance	-0.452 (0.047)			
Linguistic Distance		-0.572 (0.042)		
Religious Distance			-0.617 (0.053)	
Cultural distance index, WVS				-0.816 (0.094)
Observations	2368	2368	2368	2368
R-squared	0.285	0.489	0.541	0.483
	Panel C. Dependent variable: "Italian spoken at home"			
Genetic distance	-0.053 (0.043)			
Linguistic distance		-0.120 (0.293)		
Religious Distance			-0.031 (0.068)	
Cultural distance index, WVS				-0.036 (0.060)
Observations	596	596	596	596
R-squared	0.339	0.335	0.336	0.336

Notes: This Table shows OLS estimates of the relationship between marriage outcomes and cultural distance. The outcomes are gains from marriage, computed as in equation (2), in panel A; fertility and Italian socialization rates in panel B and C, in turn. We consider four different measures of cultural distance, i.e., distance along genetics, language, religion, and attitudes and values from the World Value Survey; we refer to Spolaore and Wacziarg (2009, 2016) for further details. All specifications control for spouses' education and include marriage market (at the regional level) fixed effects. Robust standard errors clustered at the regional level in parentheses.

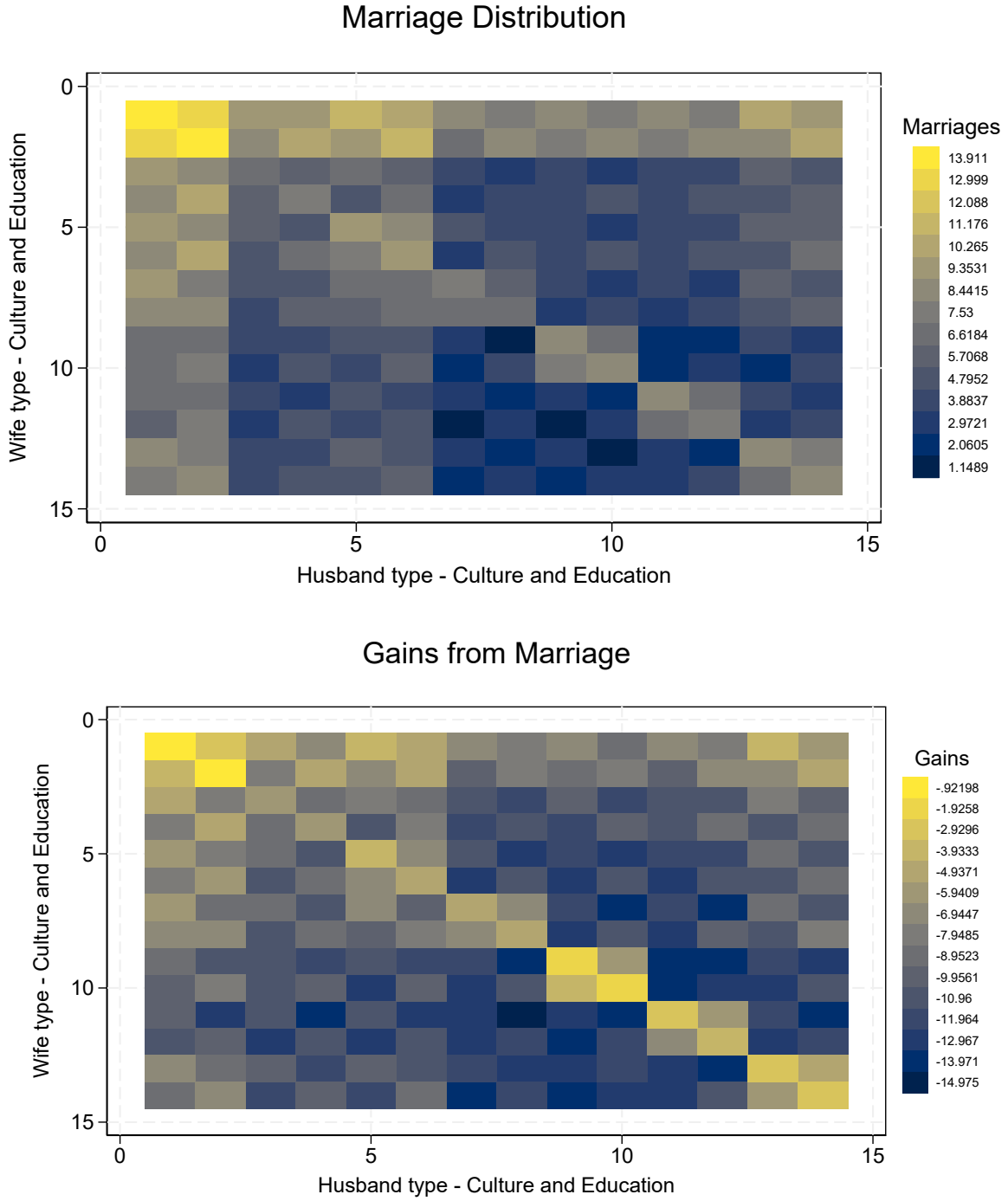


Figure C3: Marriage and Gains from Marriage Distribution (Heatmaps)

Notes: This Figure reports the heatmap of the bivariate marriage distribution (in log) and the distribution of gains from marriage by cultural-ethnic group and education of spouses. We consider seven cultural-ethnic groups: Italian, European-EU15, Other European, North African and Middle-Eastern, Sub-Saharan African, East Asian, and Latin American; and we distinguish two levels of education: highly-educated individuals and low educated individuals.

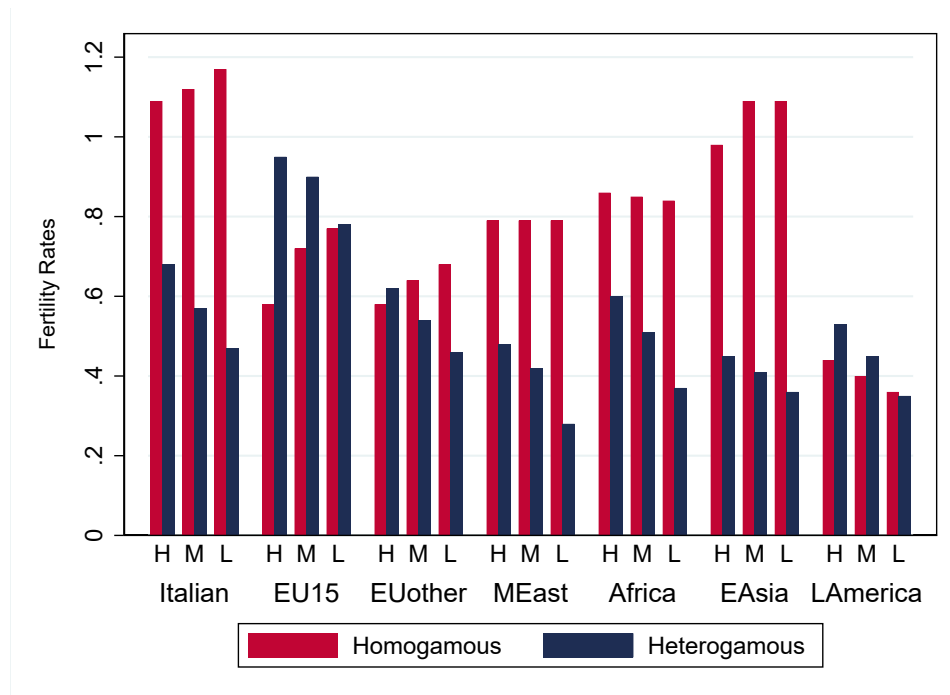


Figure C4: Fertility Rates by Ethnic Group and Education of Spouses

Notes: This Figure shows average fertility rates by ethnic group and education of spouses (H: both highly-educated spouses, M: high and low-educated spouses, L: both spouses with low education). Estimates are reported separately for homogamous (red) and heterogamous (blue) households.

Table C3: Fertility Rate by Culture and Education

Dependent variable:	(1)	(2)	(3)	(4)
	Fertility rate			
Homogamous by culture	0.280 (0.016)	0.280 (0.017)	0.280 (0.017)	0.337 (0.023)
Homogamous by education		-0.005 (0.009)		
High-high educ			0.026 (0.019)	0.050 (0.021)
High and low educ			0.019 (0.013)	0.031 (0.015)
Hom by culture $\times$ High-high educ				-0.114 (0.021)
Hom by culture $\times$ High and low educ				-0.057 (0.017)
Observations	2456	2456	2456	2456
R-squared	0.296	0.315	0.315	0.318
Marriage market (region) fixed effects	Yes	Yes	Yes	Yes
Ethnic and Education (wife) fixed effects	Yes	Yes	Yes	Yes

Notes: This Table shows the results of the regression of the fertility rate - in marriages defined by spouses' cultural ethnic-group and education, and by region - on various (combinations of) explanatory variables. In columns (1) and (2), the explanatory variables are dummies for homogamous couples by culture (*Homogamous by culture*) and education (*Homogamous by education*). Column (3) includes dummies for marriages by education, dividing marriages of high-educated (*High-high educ*) and mixed education marriages of (*High and low educ*) spouses. Column (4) includes interaction variables. All specifications include marriage market (at the regional level) fixed effects, wife educational level, and cultural-ethnic group of origin fixed effects. Standard errors clustered at the regional level are reported in parentheses.

Table C4: Separation Rate by Culture and Education

Dependent variable:	(1)	(2)	(3)	(4)	(5)
	Separation rate				
Homogamous by culture	-0.015 (0.003)		-0.015 (0.003)	-0.015 (0.003)	-0.020 (0.006)
Homogamous by education		0.013 (0.002)	0.013 (0.002)		
High educ (at least one)				-0.020 (0.004)	-0.022 (0.004)
Hom by culture $\times$ High educ (at least one)					0.007 (0.005)
Observations	2456	2456	2456	2456	2456
R-squared	0.090	0.063	0.101	0.104	0.105
Marriage market (region) fixed effects	Yes	Yes	Yes	Yes	Yes
Ethnic and Education (wife) fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: This Table shows the results of the regression of the separation rate - in marriages defined by spouses' cultural ethnic-group and education, and by region - on various (combinations of) explanatory variables. In columns (1)-(3) the explanatory variables are dummies for homogamous couples by culture (*Homogamous by culture*) and by education (*Homogamous by education*). In column (4), we include a dummy for marriages with at least one highly-educated spouse (*High educ, at least one*) and its interaction with the dummy for cultural homogamous marriages (*Hom by culture $\times$ High educ, at least one*) in column (5). All specifications include marriage market (at the regional level) fixed effects as well as spouses' cultural-ethnic group of origin fixed effects, and a dummy for wife's high education. Standard errors clustered at the regional level are reported in parentheses.

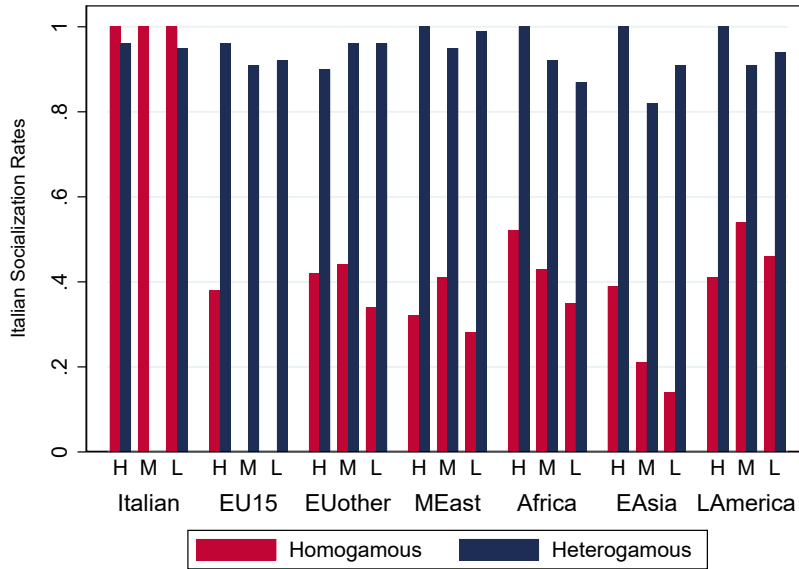


Figure C5: Italian Socialization Probabilities by Ethnic Group and Education of Spouses

Notes: This Figure shows socialization probabilities to Italian culture by ethnic group and education of spouses (H: both highly-educated spouses, M: high and low-educated spouses, L: both spouses with low education). The outcome variable is “Italian spoken at home.” Estimates are reported separately for homogamous (red) and heterogamous (blue) households.

Table C5: Fraction of “Italian spoken at home” by Culture and Education

Dependent variable:	(1)	(2)	(3)	(4)
	“Italian spoken at home”			
Homogamous by culture	-0.320 (0.023)	-0.319 (0.022)	-0.322 (0.023)	-0.409 (0.031)
Homogamous by education		-0.024 (0.022)		
High-high educ			0.044 (0.075)	-0.034 (0.074)
High and low educ			0.050 (0.058)	-0.050 (0.071)
Hom by culture $\times$ High-high educ				0.068 (0.057)
Hom by culture $\times$ High and low educ				0.141 (0.039)
Observations	615	615	615	615
R-squared	0.604	0.608	0.608	0.613
Marriage market (region) fixed effects	Yes	Yes	Yes	Yes
Ethnic and Education (wife) fixed effects	Yes	Yes	Yes	Yes

Notes: This Table shows the results of the regression of “Italian spoken at home” - in marriages defined by defined by spouses’ cultural ethnic-group and education, and by region - on various (combinations of) explanatory variables. In columns (1) and (2), the explanatory variables are dummies for homogamous couples by culture (*Homogamous by culture*) and education (*Homogamous by education*). Column (3) includes dummies for marriages by education, dividing marriages of high-educated (*High-high educ*) and mixed education marriages of (*High and low educ*) spouses. Column (4) includes interaction variables. All specifications include marriage market (at the regional level) fixed effects, wife educational level, and cultural-ethnic group of origin fixed effects. Standard errors clustered at the regional level are reported in parentheses.

Table C6: Fraction of “High educated Children” by Culture and Education

Dependent variable:	(1)	(2)	(3)	(4)	(5)
	Fraction of “High educated Children”				
Homogamous by culture	-0.081 (0.013)		-0.080 (0.013)	-0.081 (0.013)	-0.053 (0.015)
Homogamous by education		0.013 (0.006)	0.015 (0.006)		
High educ (at least one)				0.071 (0.006)	0.079 (0.006)
Hom by culture $\times$ High educ (at least one)					-0.038 (0.004)
Observations	2456	2456	2456	2456	2456
R-squared	0.508	0.301	0.618	0.640	0.641
Marriage market (region) fixed effects	Yes	Yes	Yes	Yes	Yes
Spouses’ Cultural-ethnic group fixed effects	Yes	Yes	Yes	Yes	Yes

Notes: This Table shows the results of the regression of “highly-educated children” - in marriages defined by defined by spouses’ cultural ethnic-group and education, and by region - on various (combinations of) explanatory variables. In columns (1)-(3), the explanatory variables are dummies for homogamous couples by culture (*Homogamous by culture*) and by education (*Homogamous by education*). In column (4), we includes a dummy for marriages with at least one highly educated spouse (*High educ, at least one*) and its interaction with the dummy for cultural homogamous marriages (*Hom by culture $\times$ High educ, at least one*) in column (5). All specifications include marriage market (at the regional level) fixed effects, wife educational level, and spouses’ cultural-ethnic group of origin fixed effects. Standard errors clustered at the regional level are reported in parentheses.

Table C7: Italian Socialization Probabilities by Ethnic Group and Marital Status

Italian Socialization Probabilities				
	Homogamous Families		Heterogamous Families	
	Married	Separated	Married	Separated
Italian	1	1	0.936	0.736
Europe-EU15	0.410	0.546	0.885	0.750
Other Europe	0.389	0.472	0.940	0.786
North Africa-Middle East	0.268	0.357	0.919	0.619
Sub-Saharan Africa	0.398	0.238	0.927	0.600
East Asia	0.198	0.242	0.856	0.375
Latin America	0.493	0.426	0.927	0.750

Notes: This table shows Italian socialization probabilities by ethnic group of spouses and marital status. The outcome variable is an indicator for whether the child speaks Italian within the family. Estimates are reported separately for homogamous and heterogamous married and separated families.

Table C8: Separation Rates in Marriages With and Without Children

	Separation Rates			
	Homogamous		Heterogamous	
	$\pi_{hh}(n > 0)$	$\pi_{hh}(n = 0)$	$\pi_{hj}(n > 0)$	$\pi_{hj}(n = 0)$
Italian	0.054	0.095	0.045	0.097
Europe-EU15	0.024	0.025	0.041	0.061
Other Europe	0.016	0.040	0.039	0.093
North Africa-Middle East	0.023	0.072	0.073	0.127
Sub-Saharan Africa	0.017	0.037	0.063	0.108
East Asia	0.010	0.021	0.040	0.080
Latin America	0.026	0.061	0.053	0.114

Notes: This Table reports the separation rates by ethnic group of spouses in families with and without children, separately for homogamous and heterogamous couples.

Table C9: Singles by Cultural-Ethnic Group and Education

	Single Males				Single Females			
	Unskilled (%)		Skilled (%)		Unskilled (%)		Skilled (%)	
Italian	1994972	51,73	3366578	60,87	996934	40,60	2898654	54,93
Europe-EU15	24963	55,59	31161	54,11	15739	42,89	33036	48,96
Other Europe	79040	69,63	85751	70,46	56212	40,45	111006	49,70
Middle-East	37887	69,43	22818	63,24	12098	57,58	10923	59,60
Sub-Saharan Africa	18024	70,31	14029	61,50	11051	47,57	9759	51,41
East Asia	35551	79,94	20331	78,82	17549	58,06	16259	55,09
Latin America	17128	58,70	21777	57,24	20356	36,65	32726	41,98

Notes: This Table reports the distribution of singles by gender and cultural-ethnic group, separately for individuals with low and high education. Singles are defined as adult individuals not legally married over the 90th percentile of the age at marriage distribution. We report absolute numbers and shares, computed as the number of singles over the total number of individuals by gender, cultural-ethnic group, and education.

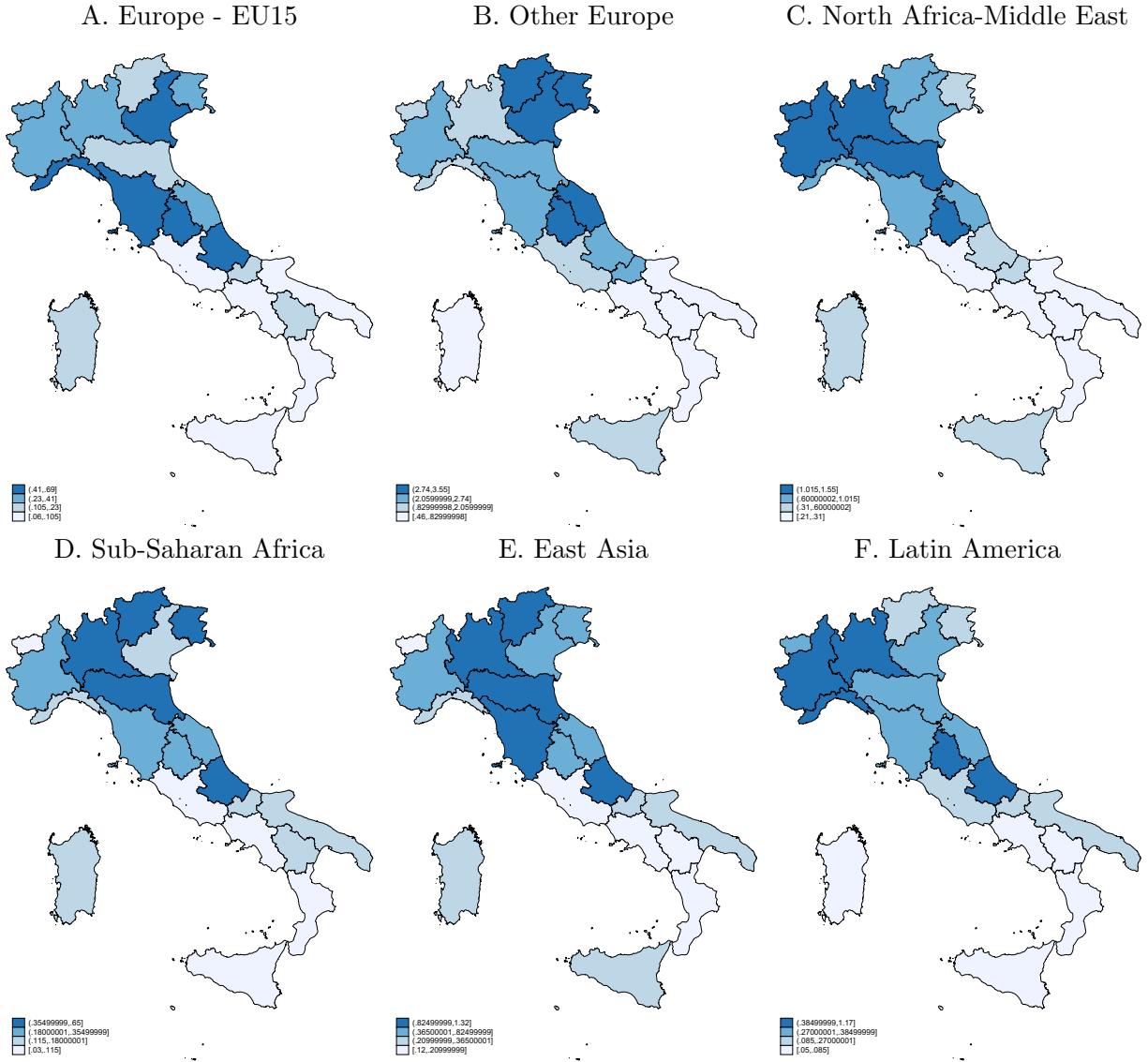


Figure C6: Migrants' Distribution across Regions

Notes: This Figure shows the distribution of the immigrant population by cultural-ethnic group and region. Population shares by cultural-ethnic group and region are computed over the total resident population at the regional level. The ethnic group classification is defined in Table C1. The color classification corresponds to the quartiles of the population distribution.

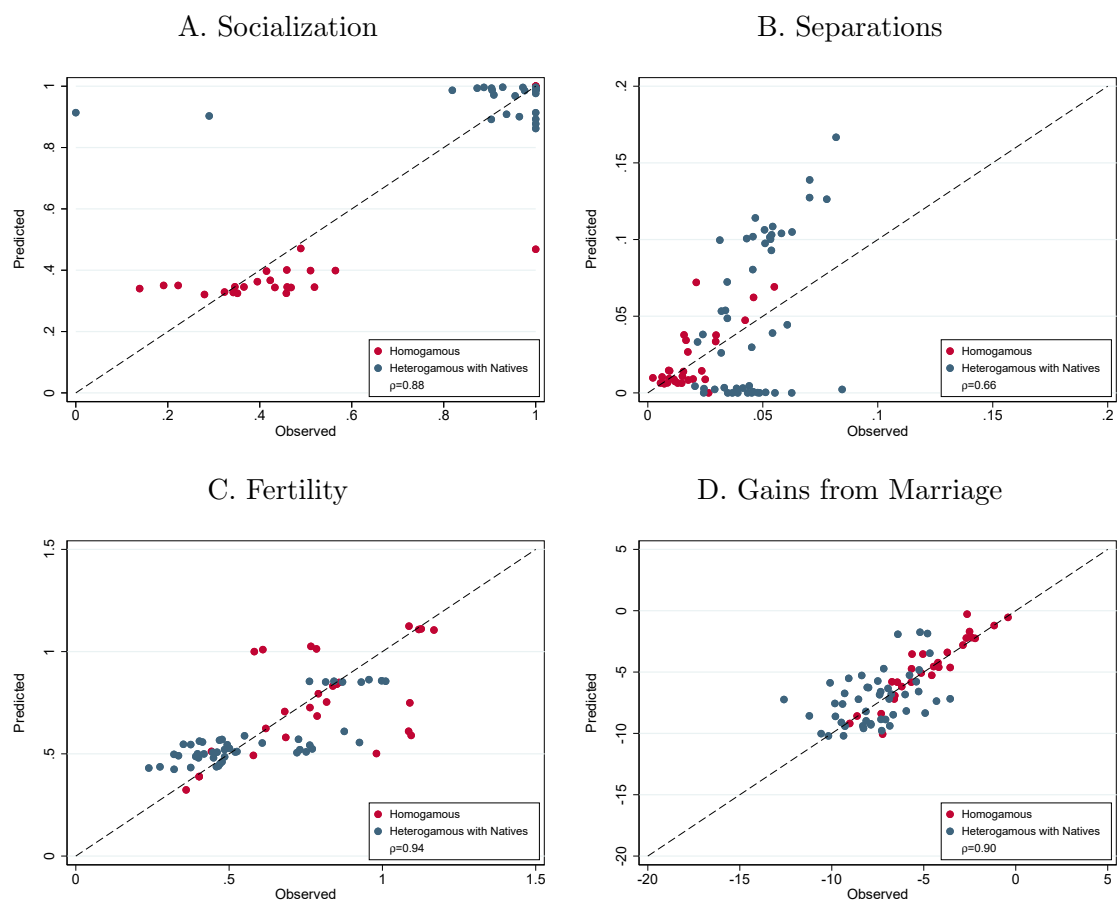


Figure C7: Fit of the model

Notes: This Figure shows the average fit of the model by h_j household, considering socialization rates to native Italian culture (panel A), fertility rates (panel B), separation rates (panel C), and gains from marriage (panel D), for homogamous (red) and heterogamous families with natives (blue).

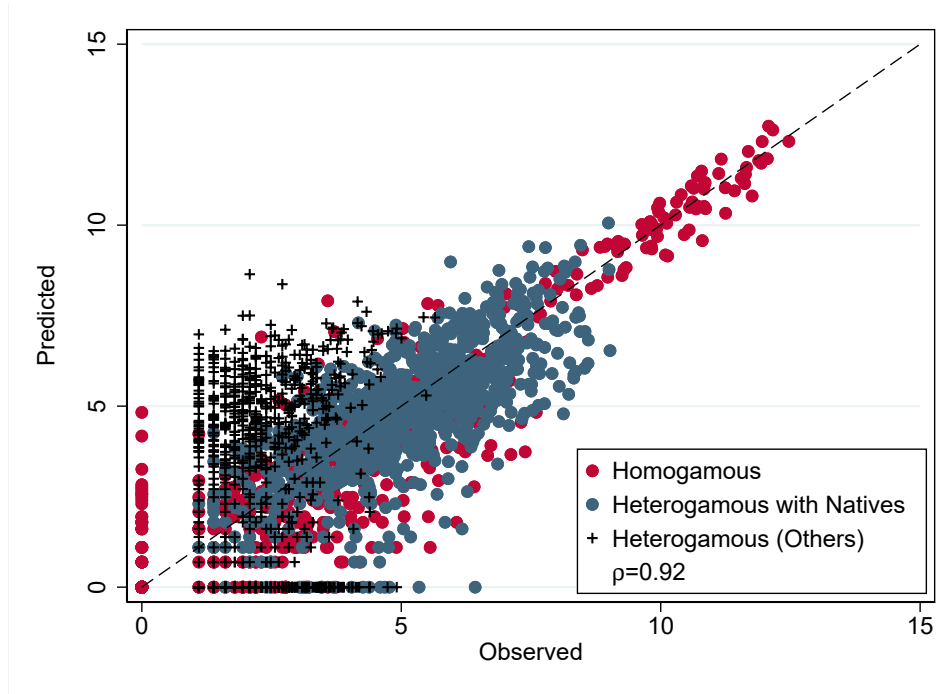
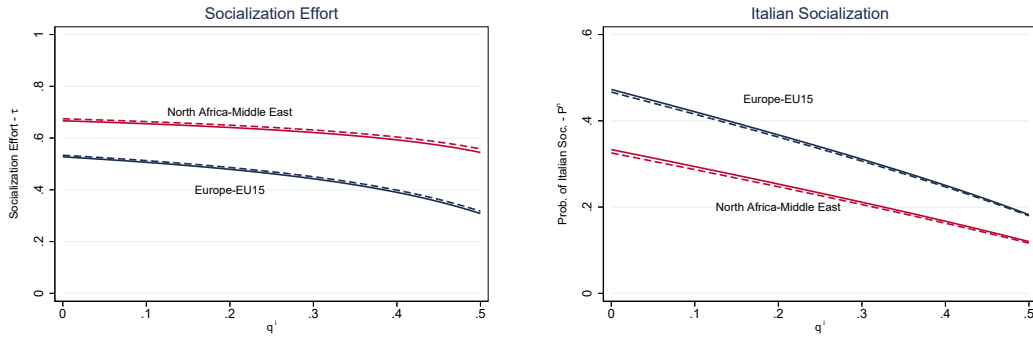


Figure C8: Fit of the Model - Number of Marriages by Match and Region

Notes: This Figure shows the relationship between the number of marriages observed in the data (in log) and the number of marriages predicted by the model (in log) by region for homogamous families (red), heterogamous families with natives (blue), and all other heterogamous matches (black).

A. Homogamous



B. Heterogamous

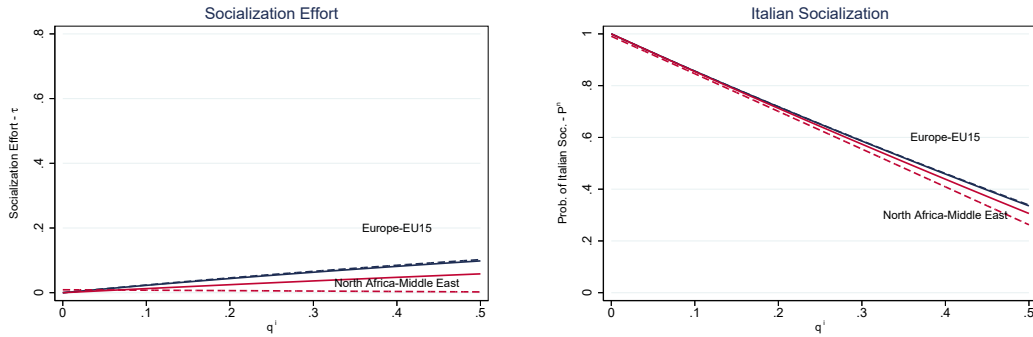


Figure C9: Estimates of Socialization Effort and Italian Socialization Rates by Minorities

Notes: This Figure reports estimates of socialization effort, τ , and the socialization probability to Italian culture, P^n , over the potential population share, q^i , for European-EU15 and North African-Middle East minorities. Results for homogamous and heterogamous marriages with natives are, respectively, in panel A and B.

Table C10: Outside Option of Singlehood Parameters

Cultural group c:	Italian	Europe-EU15	Other Europe	Middle East	Sub-Sah Africa	East Asia	Latin America
Homogamous households							
Outside option, High-High Educated ω_c	75.97 (1.893)	59.08 (2.097)	14.48 (1.097)	24.63 (2.207)	36.78 (1.782)	13.33 (1.972)	12.047 (1.250)
Outside option, Mix educated ω_c	74.85 (1.859)	63.45 (2.431)	22.87 (1.446)	30.69 (2.374)	38.13 (2.042)	20.38 (2.128)	8.53 (1.664)
Outside option, Low-Low educated ω_c	71.86 (1.785)	63.28 (2.282)	26.36 (1.152)	33.32 (1.974)	35.01 (1.688)	25.57 (1.430)	5.17 (1.401)
Economic complementarity in marriage α_{hom}	0.425 (0.017)						
Heterogamous households							
Outside option for het, ω_c	47.12 (0.029)	85.48 (0.029)	7.51 (0.030)	0.90 (0.029)	14.02 (0.029)	5.36 (0.029)	7.76 (0.028)
Outside option ω_e High-High Educated	0.696 (0.018)						
Outside option ω_e Mix Educated	0.682 (0.017)						
Outside option ω_e Low-Low Educated	0.629 (0.015)						
Economic complementarity in marriage α_{het}	0.427 (0.039)						

Notes: This Table shows structural parameter estimates of the outside option of being single ω_{cz} and $\bar{\omega}_{ez}$. Standard errors are reported in parentheses.

Table C11: Structural Model Parameters - Model of Marriage by Culture Only

Cultural Intolerance Parameters							
c:	Italian	Europe-EU15	Other Europe	Middle East	Sub-Sah Africa	East Asia	Latin America
ΔV_n^c , Italian		33.38	58.60	67.88	78.23	61.53	19.27
ΔV_{IE}^c , Europe-EU15	10.21		52.60	4.77	6.69	18.54	0.33
ΔV_{IO}^c , Other Europe	39.97	0.05		69.32	56.63	30.42	23.02
ΔV_{IM}^c , North Africa-Middle East	65.35	7.00	58.98		97.85	52.12	52.25
ΔV_{IA}^c , Sub-Saharan Africa	55.00	28.37	58.42	96.37		81.22	42.50
ΔV_{IS}^c , East Asia	40.02	0.30	87.08	54.70	47.29		93.13
ΔV_{IL}^c , Latin America	38.95	10.69	20.77	58.99	46.96	29.13	
Cost Function and Extra Parameters							
Socialization Cost Parameters	σ_τ hom	10.82	Fertility Cost Parameters			σ_n hom	67.62
	λ_τ hom	0.549				λ_n hom	0.006
	σ_τ het	21.61				ϵ hom	1.021
	λ_τ het	0.571				σ_n het	99.86
Extra Marital Gain per Child	δ	0.820				λ_n het	0.021
Segregation Parameter	ρ	1.765				ϵ het	1.229
Economic complementarity in marriage	α	0.484					

Notes: This Table shows structural parameter estimates from a model of marriage along cultural-ethnic lines only; that is, with $\gamma_c = 1$ and $S_c = 0$ for all c .

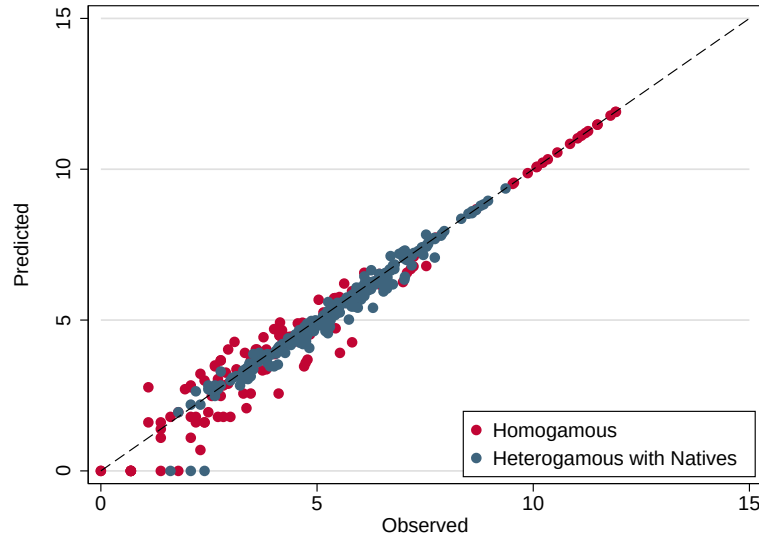


Figure C10: Model Validation - Non-Targeted 2013-2019 Marriages by Match and Region

Notes: This Figure displays the scatterplot of the relationship between the number of marriages observed in the out-of-sample data (in log) in the years from 2013 to 2019 and the number of marriages predicted by the model (in log) by region, for homogamous families (red) and heterogamous families with natives (blue). Out-of-sample aggregate marriage data doesn't provide details on heterogamous marriages between different immigrant groups.

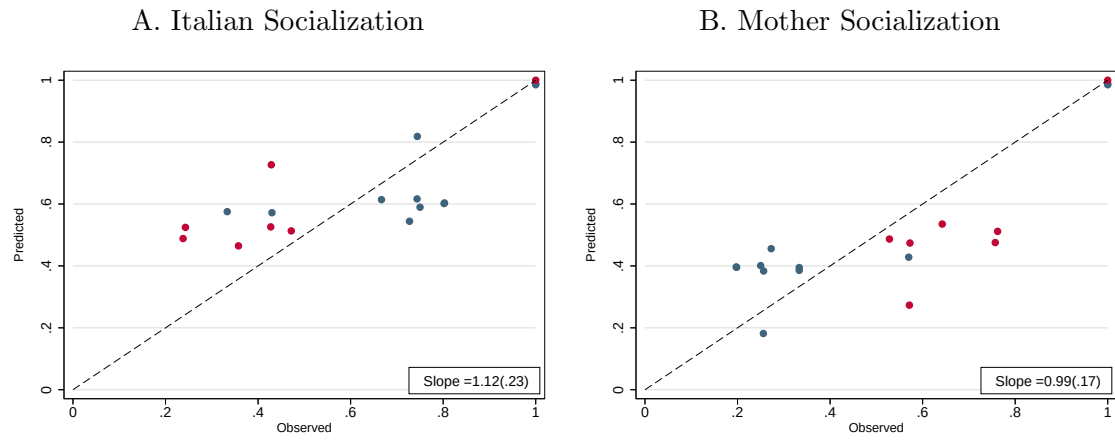


Figure C11: Model Validation - Non Targeted Socialization Rates for Divorced Couples

Notes: This Figure displays the scatterplot of the relationship between the observed and predicted Italian and mother socialization probabilities for the subsample of marriages ending in divorce, separately for homogamous families (red), and heterogamous families with natives (blue); 45-degree dashed line also reported.

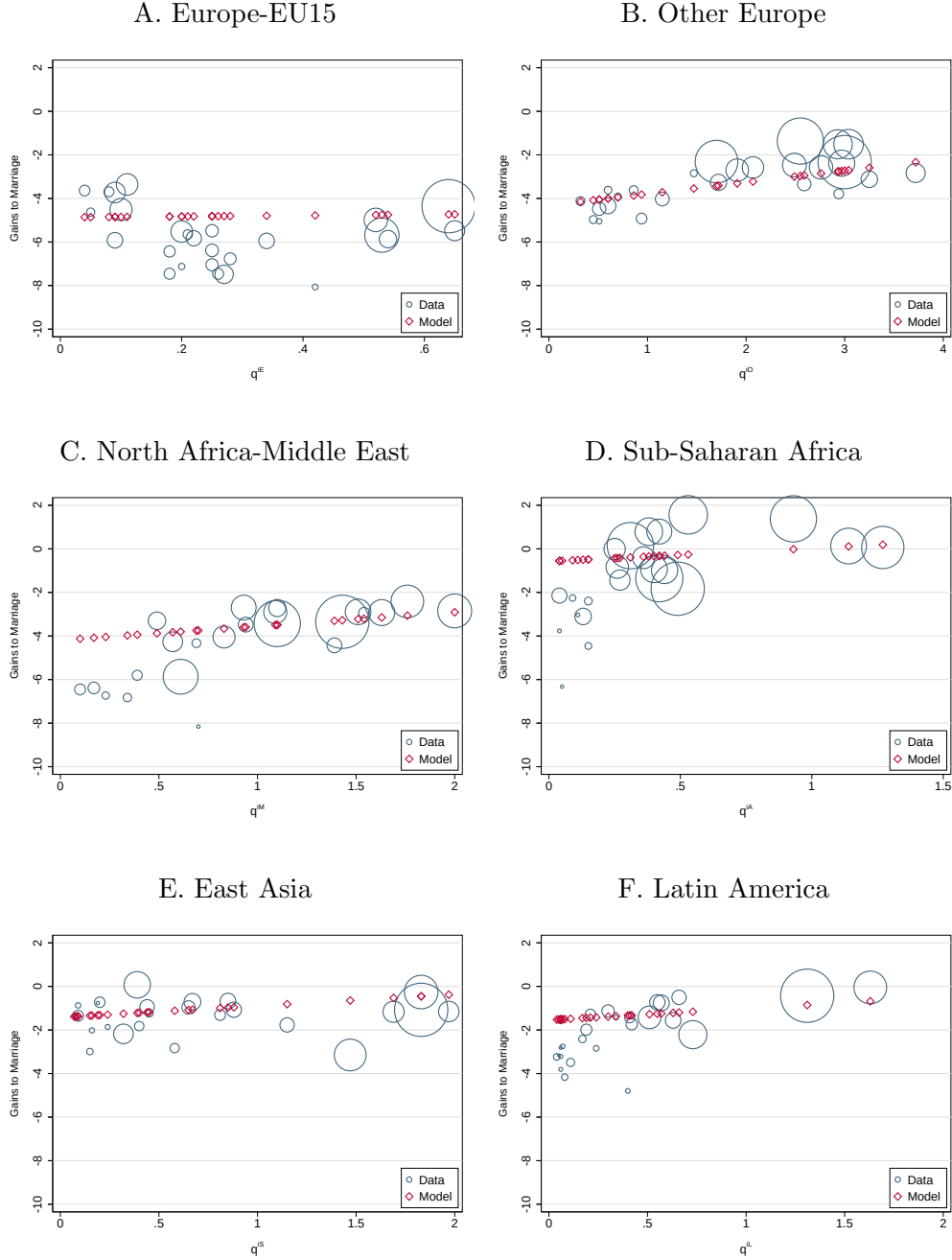
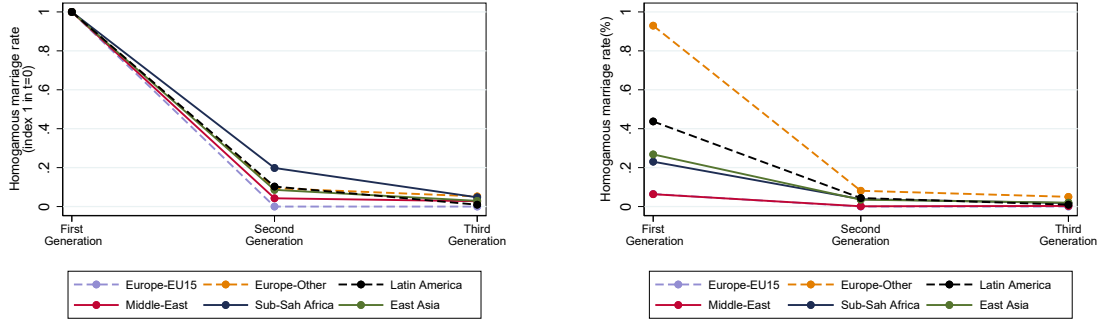


Figure C12: Model Validation - Gains from Marriage for Homogeneous Families by Province

Notes: The Figure shows out-of-sample predicted and implied gains from marriage for homogamous families of ethnic group minorities over the corresponding population share, q^i (in percentage), by province of residence (average over the time period). Empirical moments are weighted by the observed number of marriages per province. We select the most representative provinces across the northern, central, and southern parts of the country. The provinces are: Torino, Valle d'Aosta, Genova, Varese, Milano, Bergamo, Brescia, Trento, Verona, Venezia, Padova, Bologna, Ancona, Firenze, Perugia, Roma, Benevento, Napoli, Salerno, l'Aquila, Bari, Taranto, Potenza, Catanzaro, Palermo, and Cagliari.

A. Homogamous Marriages



B. Heterogamous Marriages

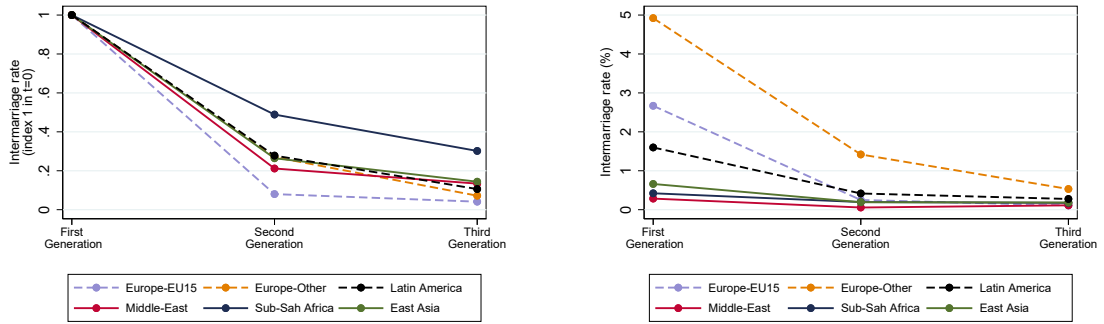


Figure C13: Inter-generational of Marital Matching

Notes: This Figure shows the inter-generational dynamics of matching patterns for homogamous marriages and heterogamous marriages with natives, in panel A and B, respectively, over successive generations.

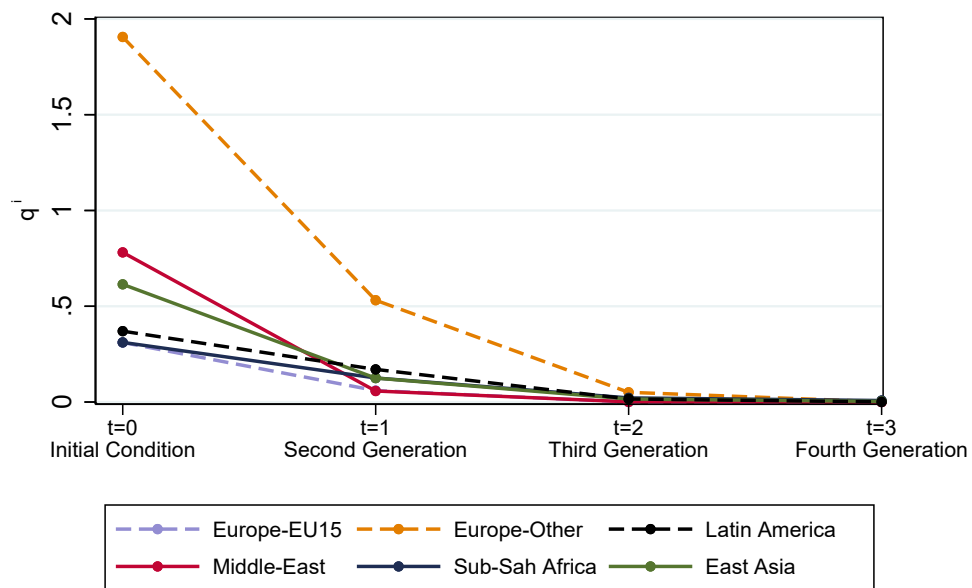


Figure C14: Inter-generational Dynamics of Cultural Traits

Notes: This Figure shows the inter-generational dynamics of the distribution of cultural traits in the population for all minorities.

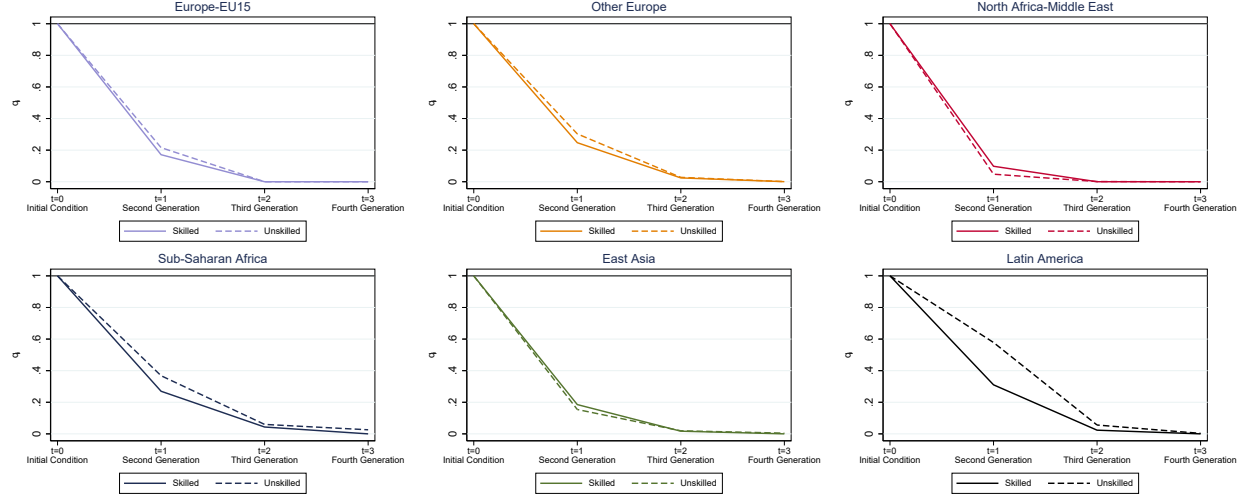


Figure C15: Inter-generational Dynamics of Cultural Traits by Education

Notes: This Figure shows the inter-generational dynamics of the distribution of cultural traits in the population for minority groups by education, normalized so that $q_t^{ie} = 1$ in $t = 0$ for $e \in \{s, u\}$. The solid line represents the dynamics for highly-educated immigrant minorities, while the dashed line represents the dynamics for immigrant minorities with low education.

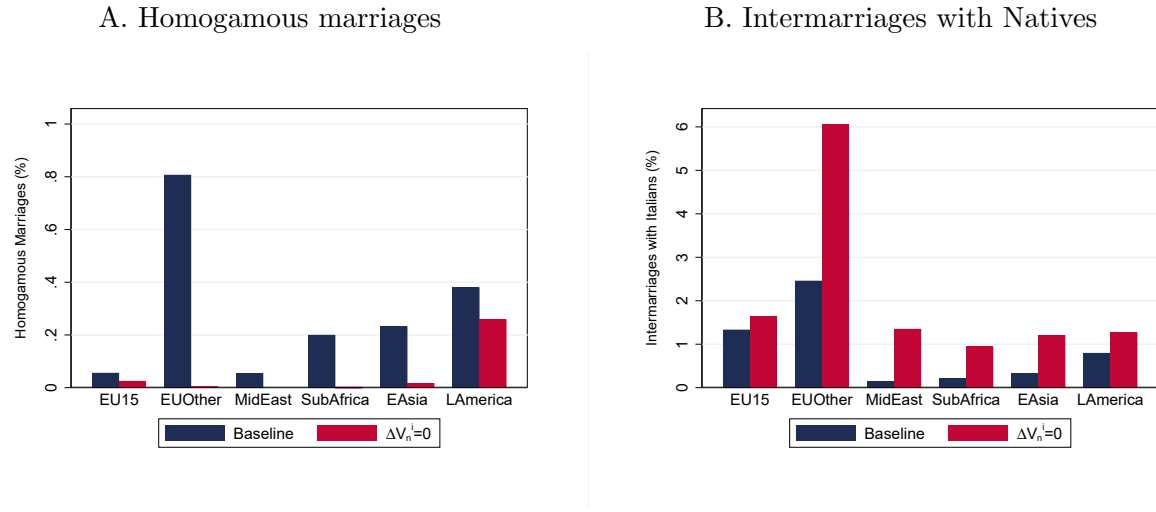


Figure C16: Change in Matching Patterns with Italians Fully Tolerant

Notes: This Figure shows the percentage change in homogamous and heterogamous marriages, in panel A and B, respectively - with full tolerance of natives towards minorities with respect to baseline; that is, $\Delta V_{ns}^{is} = 0$.

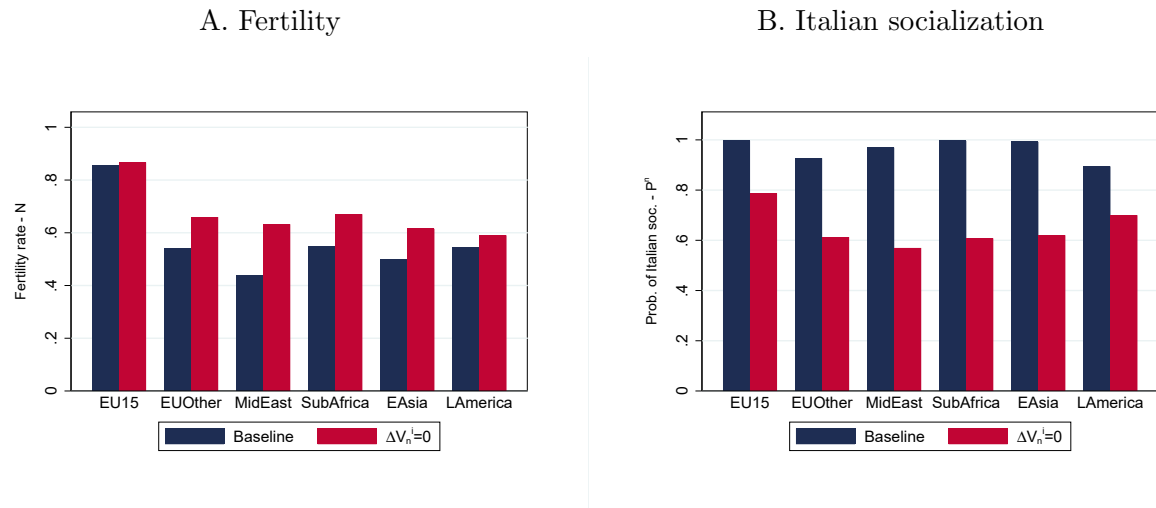


Figure C17: Change in Intra-household Patterns with Italians Fully Tolerant

Notes: This Figure shows the variation in fertility rate and socialization probability to Italian culture in panel A and B, respectively - in intermarriages with natives at the baseline and in case of complete tolerance of Italian majority towards minorities; that is, $\Delta V_{ns}^{is} = 0$.

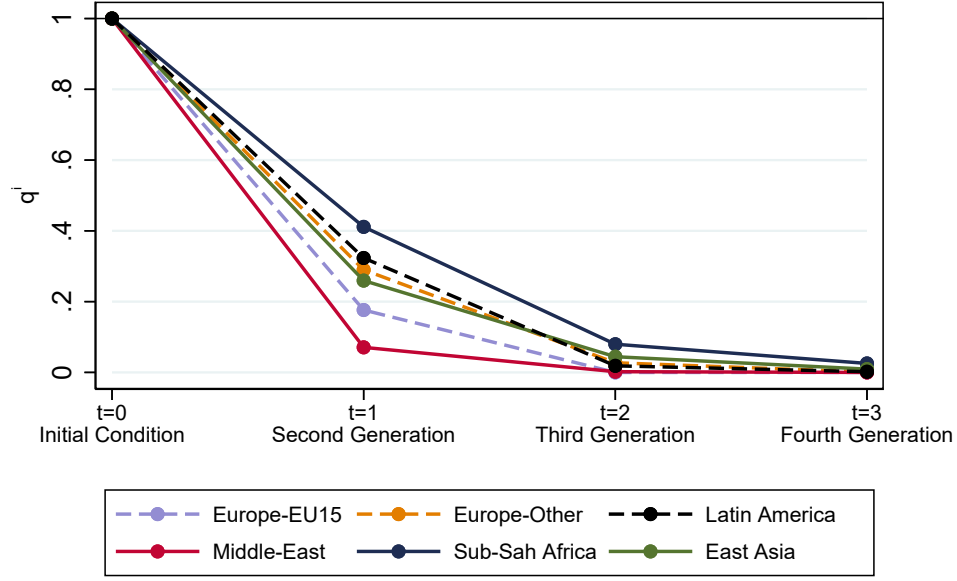


Figure C18: Inter-generational Dynamics of Cultural Traits with Italians Fully Intolerant, $\Delta V_{ns}^{is} = 100$

Notes: This Figure shows the inter-generational dynamics of the distribution of cultural traits in the population for all minorities i , assuming the case of complete intolerance of Italian majority towards minorities.

Table C12: Italian Language Socialization and Additional Measures of Integration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Dependent variable:	Social network and language			Educational achievement			Italian language proficiency				
	Speaking ITA w/ school mates	Having Italian friends	Speaking ITA w/ friends	High education	Pass all years	Aspiration university	Reading	Writing	Speaking	Dialogue	Media
Italian at Home	0.077 (0.01)	0.164 (0.02)	0.249 (0.01)	0.080 (0.02)	0.039 (0.01)	0.065 (0.02)	0.141 (0.01)	0.155 (0.01)	0.142 (0.01)	0.137 (0.01)	0.139 (0.02)
Observations	2,661	2,661	4,273	8,007	2,927	1,661	4,273	4,273	4,273	4,273	2,151
R-squared	0.099	0.154	0.181	0.082	0.065	0.112	0.124	0.126	0.144	0.150	0.192
Dep. var. mean	0.948	.328	0.838	0.518	0.909	0.533	.723	.712	.797	.803	.802
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This Table shows estimates of the correlation between our measure of linguistic socialization to Italian culture ("Italian at home") and various measures of socio-cultural integration concerning social networks in columns (1)-(3), educational achievement and aspiration in columns (4)-(6), and proficiency in the Italian language in columns (7)-(11). The sample is restricted to children and young adults (less than 25 years old), living with their parents at the time of the interview. The dependent variables include in column (1) an indicator for whether the child speaks Italian with his school mates; in column (2) an indicator for whether the child has at least some Italian friends out of the school; in column (3) an indicator for whether the child speaks Italian with his friends out of the school; in column (4) an indicator for high educational attainment (above high school); in column (5) an indicator for having passed all academic years; in column (6) an indicator for aspirations to university enrollment; in columns (7)-(11) a series of indicators for very good Italian proficiency in reading, writing, speaking, comprehension of interpersonal conversation and comprehension of media (television and radio newscast). Unconditional means of the dependent variables are reported below. All specifications control for province fixed effects, gender and age fixed effects. Standard errors clustered at the province level are reported in parentheses.

Table C13: Italian Language Socialization and Educational Outcomes - Robustness

	(1)	(2)	(3)	(4)	(5)
Panel A. Dependent variable: Reading standardized test score, 5th grade					
Italian at Home	0.371 (0.004)	0.298 (0.004)	0.203 (0.004)	0.202 (0.004)	0.193 (0.004)
Observations	337,096	336,369	330,739	330,778	330,739
R-squared	0.055	0.079	0.144	0.142	0.210
Panel B. Dependent variable: Math standardized test score, 5th grade					
Italian at Home	0.239 (0.004)	0.192 (0.004)	0.109 (0.004)	0.110 (0.004)	0.103 (0.004)
Observations	352,895	352,121	345,980	346,021	345,980
R-squared	0.036	0.048	0.098	0.100	0.190
Province and Cohort fixed effects	Yes	Yes	Yes	Yes	Yes
Student Controls	No	Yes	Yes	Yes	Yes
Family Controls	No	No	Yes	Yes	Yes
School fixed effects	No	No	No	No	Yes
Panel C. Dependent variable: Choosing the high-track, 10th grade					
Italian at Home	0.064 (0.003)	0.051 (0.003)	0.027 (0.003)	0.027 (0.003)	
Observations	93,000	92,477	90,656	90,691	
R-squared	0.018	0.023	0.058	0.059	
Province and Cohort fixed effects	Yes	Yes	Yes	Yes	
Student Controls	No	Yes	Yes	Yes	
Family Controls	No	No	Yes	Yes	

Notes: This Table shows how our measure of linguistic socialization to Italian ("Italian at home") influences the achievement and educational choices of immigrant students. The dependent variables include the reading and math standardized test scores in grade 5 in panel A and B, respectively, and a dummy equal to one for students choosing the high-track (academic or technical schools) in panel C. The sample includes all students with at least one immigrant parent. Student controls include gender, regular schooling, a dummy for first-generation immigrants, and a dummy for kindergarten. Family controls include mother's and father's education and a set of dummies for deciles of the socio-economic status distribution. Estimates are weighted by a cheating correction factor, except in column (4). Province and cohort fixed effects included in all specifications. School fixed effects included in column (5). Robust standard errors clustered at school level are in parentheses.

D Additional Robustness

We present three robustness exercises to address the possibility that: (i) fertility differences may reflect systematic variation in spouses' observable characteristics; (ii) the population distribution across regions may be endogenous; and (iii) children's educational outcomes may depend on their parents' marital status, that is, whether the parents are divorced.

First, we address the possibility that differences in fertility across cultural and education traits might be due to systematic variation in observables across households.¹⁰ We regress fertility on covariates such as marital duration, age at marriage, educational attainment (primary education, secondary education, high-school education level, college degree and above), a dummy variable for employment, and occupational standing (blue-collar, white-collar, director, self-employed, and entrepreneur). We compute fertility moments from regression residuals, which control for confounding factors. Figure D1 shows cultural intolerances consistent with our main estimates, with the exception of a slight underestimation of the demand for cultural identity among East Asian minorities.

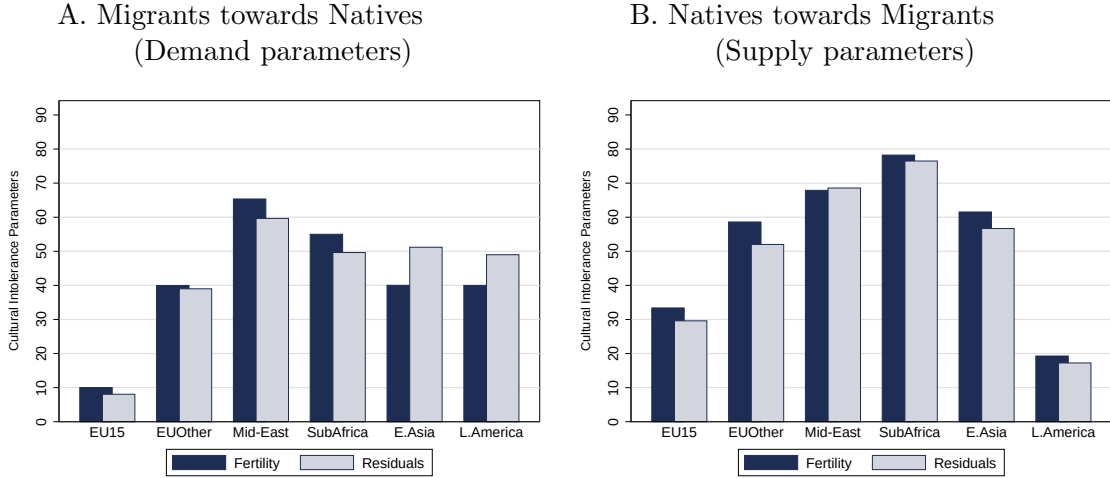


Figure D1: Cultural Intolerance Parameters - Robustness with Fertility Residuals

Notes: This Figure reports parameter estimates of the cultural intolerance of immigrants versus natives ΔV_i^n and natives versus immigrants ΔV_n^i - in panel A and B, respectively - for all cultural-ethnic groups i . The blue bars report baseline estimates. The grey bars, instead, report estimates exploiting fertility residuals from a linear regression model, to control for systematic differences in observables across households, in terms of marital duration, age at marriage of spouses, as well as education (primary education, secondary education, high-school education level, college degree and above), a dummy variable for employment, and occupational standing (blue-collar, white-collar, director, self-employed, and entrepreneur).

¹⁰As in the validation exercises in Section 5.3, we disregard educational variability. Theoretically, we impose $\gamma_c = 1$ and $S_c = 0$, and re-estimate the other parameters in our model; see footnote 39.

Our second robustness exercise tackles the possible endogeneity of the population distribution across regions. Endogenous moving or location decisions would be problematic for our estimates if these decisions were partly motivated by marriage, socialization, and unobserved heterogeneity. Consider the natural hypothesis that minorities that are particularly attached to their cultural identity choose to locate into more segregated areas. In this case, we would expect a positive correlation between direct vertical and indirect horizontal socialization. Figure D2 describes this relationship, showing instead a negative and, at times, statistically significant correlation. To further mitigate this concern, we adopt a shift-share approach (Card, 2001; Tabellini, 2020), predicting ethnic group distributions across regions based pre-existing variation in immigrants' settlements by ethnic group observed in 1993, interacted with subsequent inflows by origin. Given the persistent ethnic clustering of immigrants, these historical patterns strongly predict location choices of newly arriving immigrants. The resulting estimates remain consistent with our baseline findings.

In our final robustness analysis we investigate whether educational attainment may be

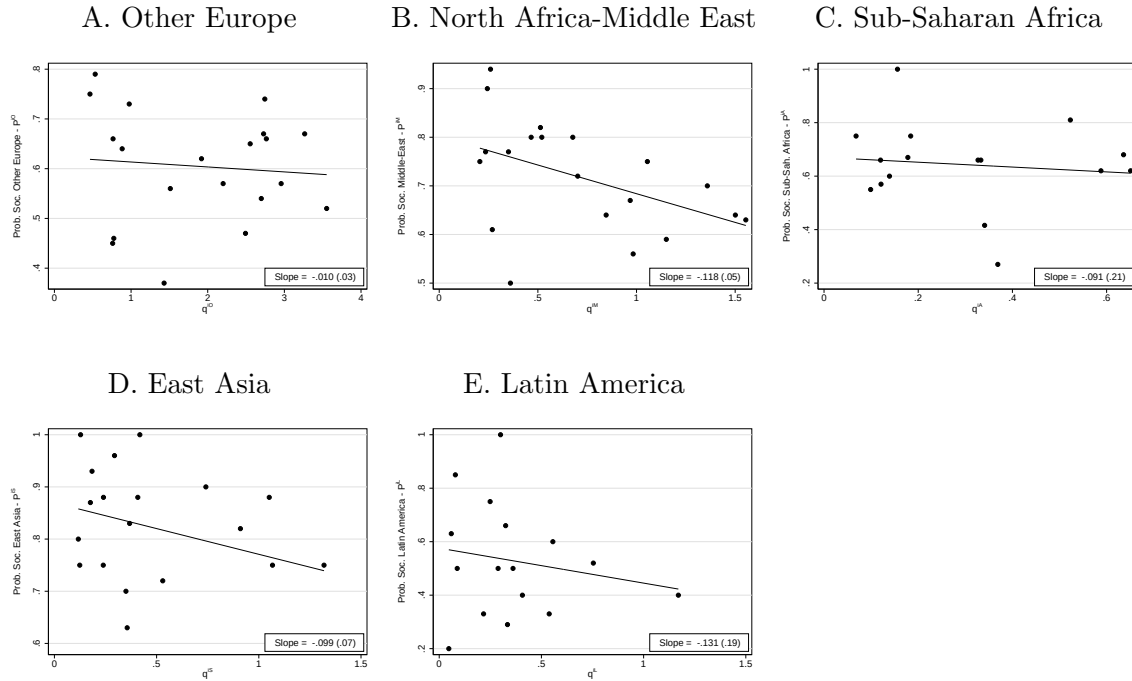


Figure D2: Minorities Socialization Probabilities and Horizontal Socialization

Notes: This figure shows the average socialization probability of each minority group, over the correspondent population share, q^i (in percentage), for all i by region of residence (average rate over the time period). The substitution pattern displayed by Europe-EU15 minority is in line with the other minorities. However, due to sample limitations and in compliance with the ADELE Laboratory agreement, we were not allowed to export the graph.

affected by the divorce decision and explore its implications for our model. We treat children’s educational attainment as exogenous, directly observed in the data. In this revision, we maintain the same approach in modeling children’s education conditional on divorce. Crucially, we rely on two data sources to estimate the probability that children have low education (i.e., enroll in low-track vocational schools as opposed to high-track schools), conditionally on parents’ cultural-ethnic traits and educational background, by region: INVALSI data and ISTAT data from the Condition and Social Integration of Foreign Nationals Survey; see Appendix B for details on the INVALSI data. Unfortunately, the INVALSI data do not provide information on parents’ marital status. However, the ISTAT survey does include this information - though for a relatively small sample of approximately 1,900 students between 15 and 25 years old (the target population for our analysis). With these data we estimate the probability of children having low education conditionally on parental divorce.

Results in column (1) of Table D1 show that children with divorced parents are, on average, more likely to have low education, i.e., they are 5 percent more likely to attend low-track (vocational) schools. The effect is positive and statistically significant at the 10% level. The estimates control for cultural-ethnic fixed effects as well as region (marriage market) fixed effects. We further investigate heterogeneity by parental education. Specifically, we report estimates separately for households in which both spouses are highly educated (column 2), the husband is highly educated and the wife has low education (column 3), and the husband has low education and the wife is highly educated (column 4). Unfortunately, the proportion of divorced couples in households with spouses with low education is insufficient to support any meaningful estimation. Crucially, the effect of interest is driven solely by families with mixed education levels, and only when the mother has low education. We find no significant effects for the other household types.

Thus, while in principle, educational attainment may be affected by the parental divorce decision, the evidence appears to be relevant only for a specific subset of households, namely, those with mothers with low education, who are more likely to retain custody of the children after divorce and raise them in a context with lower educational resources. This reasoning does not appear to generalize across all households in our data. However, we again emphasize that information on divorce is available only from the ISTAT survey, which covers a relatively small sample of divorced families. As a result, the estimated probabilities of educational attainment conditionally on divorce are subject to measurement error and should be interpreted with caution.

Despite these data limitations, we have re-estimated the model, allowing for different probabilities of educational attainment conditionally on divorce status. This approach accounts for the empirical variation in educational outcomes associated with parental marital

status. We report our estimates in Table D2. We highlight three points. First, the estimate of δ —the extra value of each child attributable to staying in a marriage, as opposed to divorcing, independently of socialization—more than doubles, from 0.82 to 1.98. This is to be expected, as this parameter now captures the fact that staying married leads to an education premium, i.e., it increases the probability that children attain a high level of education, which has value for parents. Second, cultural intolerance parameters are consistent with our main estimates, see Figure D3. Finally, the educational parameters S_c also remain in line with our main estimates, and the γ_c parameters are estimated to be close to or even greater than one. That is, cultural intolerance tends to be equal to or higher among immigrant parents with low education compared to those with high education.

Table D1: Estimates of Children with Low Education by Spouses' Education and Marital Status

	(1) All sample	(2) High-high	(3) High and Low (Low educ mother)	(4) Low and High (High educ mother)
Dependent variable:	Probability of "Low educated" Children			
Divorce (dummy)	0.051 (0.027) {0.061}	-0.039 (0.052) {0.454}	0.091 (0.047) {0.053}	-0.019 (0.067) {0.778}
Observations	1916	408	380	309
R-squared	0.062	0.129	0.060	0.087
Marriage market (region) fixed effects	Yes	Yes	Yes	Yes
Ethnic fixed effects	Yes	Yes	Yes	Yes

Notes: This Table reports OLS estimates of the the probability of low educated children by marital status of the parents. The dependent variable is a dummy equal to one for low educated children, i.e., students attending low-track (vocational) schools. The sample is restricted to students between 15 and 25 years old. The main explanatory variable is a dummy equal to one for separated parents. Column (1) reports average estimates. Columns (2)-(4) report estimates separately for households by spouses' education, distinguishing households with both highly-educated spouses (column 2), households with highly-educated husband and low-educated wife (column 3) and households with low-educated husband and highly-educated wife (column 4). All regressions include marriage market (at the regional level) fixed effects and ethnic-group fixed effects of both spouses. Robust standard errors clustered at the regional level are reported in parentheses; p-values are reported in curly brackets.

Table D2: Structural Model Parameters - Robustness with Education conditional on Divorce

Cultural Intolerance Parameters							
c :	Italian	Europe-EU15	Other Europe	Middle East	Sub-Sah Africa	East Asia	Latin America
ΔV_n^c , Italian		24.96	76.64	87.31	82.10	83.65	27.49
ΔV_E^c , Europe-EU15	20.45		48.51	14.76	87.31	87.02	17.43
ΔV_{iO}^c , Other Europe	45.62	25.73		95.66	88.14	89.64	68.82
ΔV_{iM}^c , North Africa-Middle East	71.56	9.31	97.73		54.55	59.90	79.29
ΔV_{iA}^c , Sub-Saharan Africa	70.83	95.56	80.94	46.11		59.73	59.08
ΔV_{iS}^c , East Asia	39.08	46.28	85.83	63.74	29.65		56.42
ΔV_{iL}^c , Latin America	33.20	15.48	67.52	90.64	76.62	62.93	
Education Parameters							
Surplus for skilled children, S_c	15.29	10.67	45.09	11.08	1.435	34.97	45.61
Educational preference of unskilled parents, γ_c	1.037	1.037	1.600	1.088	0.973	1.359	0.936
Cost Function and Extra Parameters							
Socialization Cost Parameters	σ_r hom	12.17	Fertility Cost Parameters			σ_N hom	67.27
	λ_r hom	0.921				λ_N hom	0.099
	σ_r het	23.46				ϵ hom	1.027
	λ_r het	0.433				σ_N het	82.19
Extra Marital Gain per Child	δ	1.988				λ_N het	0.199
Segregation Parameter	ρ	1.494				ϵ het	1.017

Notes: This Table shows the estimates of the structural parameters of the model.

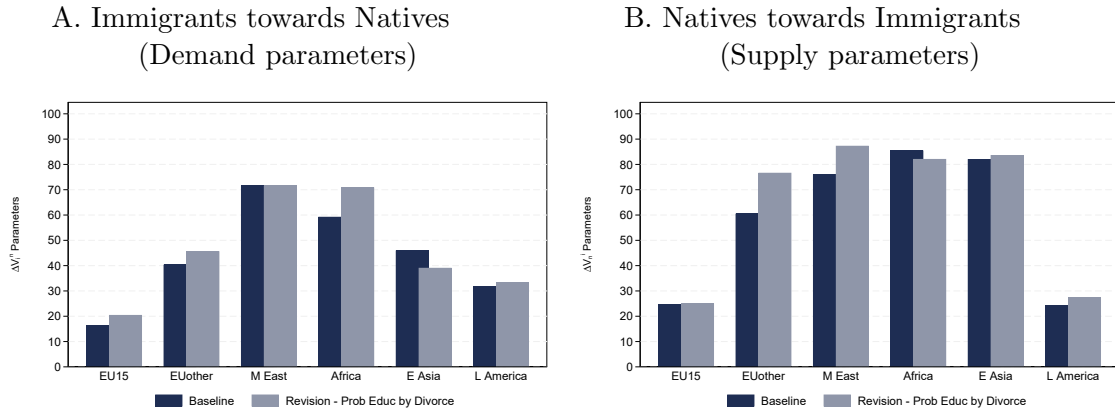


Figure D3: Cultural Intolerance Parameters

Notes: This Figure reports parameter estimates for the cultural intolerance of immigrants versus natives ΔV_{is}^{ns} and natives versus immigrants ΔV_{ns}^{is} , in panel A and B, for natives n and all cultural minorities i . The blue bars report baseline estimates. The grey bars, instead, report estimates allowing in the model for different probabilities of educational attainment conditionally on divorce status.

E Identification

We provide here proof of identification in our model. It is convenient to restrict ourselves to the set-up of the model in Section II. In particular, we restrict to dichotomous cultural-ethnics traits c , natives and immigrants $c \in (n, i)$. Furthermore, we assume that cost functions are quadratic, which in our set-up implies that τ and N can be solved for in closed form. The identification proof is easily extended to the more general model we estimate in Section III. In fact, allowing for multiple cultural groups adds over-identifying restrictions to the problem. Furthermore, we do not exploit, in the following, geographical variation across $R = 20$ regions in the data, which also adds over-identifying restrictions to the problem.

We will show that the model is over-identified given the available data by proceeding recursively, solving the moment equations for the parameters to be estimated as a function of the data (given the postulated functional forms). More specifically, the data contains the population fraction of immigrants q^i ; the socialization probabilities P_{hj}^c , the divorce rates π_{hj} , the fertility rates N_{hj} , and the marriage gains G_{hj} (which are a transformation of marriage fractions). The following detailed analysis of identification is summarized in Table E1.

From the socialization moments, we identify the cultural intolerances $\frac{\Delta V_{is}^{ns}}{\sigma_\tau}$, $\frac{\Delta V_{is}^{is}}{\sigma_\tau}$. With quadratic socialization costs, $c(\tau) = \frac{1}{2}\sigma_\tau\tau^2$, when the socialization effort is $\tau = (\tau_m, \tau_f)$ and divorce is $d = 0$, the socialization moments for homogamous immigrant households with skilled parents are:

$$P_{isis}^i = 2\frac{1}{\sigma_\tau}\Delta V_{is}^{ns}(1 - q^i) + (1 - 2\frac{1}{\sigma_\tau}\Delta V_{is}^{ns}(1 - q^i))q^i.$$

This moment induces a linear equation which is independent of γ_i, S_i , and hence it is uniquely solved for $\frac{\Delta V_{is}^{ns}}{\sigma_\tau}$, taking P_{isis}^i from data.¹¹ The corresponding moments for $P_{isiu}^i, P_{iuis}^i, P_{iuiu}^i$ induce linear equations in γ_i , given $\frac{\Delta V_{is}^{ns}}{\sigma_\tau}$.

$$P_{iuis}^i = (1 + \gamma_i)\frac{1}{\sigma_\tau}\Delta V_{is}^{ns}(1 - q^i) + (1 - (1 + \gamma_i)\frac{1}{\sigma_\tau}\Delta V_{is}^{ns}(1 - q^i))q^i.$$

Same as above (because of symmetric):

$$P_{isiu}^i = (1 + \gamma_i)\frac{1}{\sigma_\tau}\Delta V_{is}^{ns}(1 - q^i) + (1 - (1 + \gamma_i)\frac{1}{\sigma_\tau}\Delta V_{is}^{ns}(1 - q^i))q^i.$$

$$P_{iuiu}^i = 2\gamma_i\frac{1}{\sigma_\tau}\Delta V_{is}^{ns}(1 - q^i) + (1 - 2\gamma_i\frac{1}{\sigma_\tau}\Delta V_{is}^{ns}(1 - q^i))q^i.$$

¹¹The corresponding equation for $d = 1$ is linearly dependent from the first.

Table E1: Recursive Identification Structure of the Model

Parameter	Theoretical Expression
<i>Socialization in homogamous skilled immigrant households</i>	
$\frac{\Delta V_{is}^{ns}}{\sigma_\tau}$	$P_{isis}^i = \frac{2}{\sigma_\tau} \Delta V_{is}^{ns} (1 - q_i) + \left(1 - \frac{2}{\sigma_\tau} \Delta V_{is}^{ns} (1 - q_i)\right) q_i$
<i>Socialization in mixed-education homogamous immigrant households</i>	
γ_i	$P_{iuis}^i, P_{isiu}^i, P_{iuiu}^i$ depend on γ_i and ΔV_{is}^{ns}
<i>Socialization in immigrant-native heterogeneous couples</i>	
ΔV_{ns}^{is}	Based on $P_{isns}^i > q_i$, identify which parent socializes
<i>Socialization in native-immigrant heterogeneous households</i>	
γ_n	Linear in $\frac{\gamma_n \Delta V_{ns}^{is} - \Delta V_{is}^{ns}}{\sigma_\tau}$
<i>Divorce in homogamous native households</i>	
δ	$\pi_{hj}(N_{hj}) = F\left(N_{hj}(u_{hj}(\tau(1), 1) - c(\tau(1))) - N_{hj}(u_{hj}(\tau(0), 0) + \delta - c(\tau(0)))\right)$
<i>Divorce and socialization in homogamous immigrant households</i>	
σ_τ, S_i	Use known $u_{hj}(\tau, d) - c(\tau)$, solve recursively for S_i, σ_τ
<i>Fertility across all household types</i>	
σ_N	$N_{hj} = \frac{1}{\sigma_N} \left(\pi(u_{hj}(\tau(1), 1) - c(\tau(1))) + (1 - \pi)(u_{hj}(\tau(0), 0) + \delta - c(\tau(0))) \right)$
<i>Divorce probability without children across all household types</i>	
a_{hj}	$F(0; a_{hj}, 1) = \hat{\pi}_{hj}(0)$
<i>Gains from marriage across all household types</i>	
$\omega_{hz}, \omega_{jz}, \alpha_z$	$G_{hj} = U_{hj} - \omega_{hz} - \omega_{jz}$ and $U_{hj}^{ec} = \alpha_z(\omega_{hz} + \omega_{jz})$

Therefore each of these equations is uniquely solved for γ_i , given $\frac{\Delta V_{is}^{ns}}{\sigma_\tau}$ from the previous equation and $P_{isiu}^i, P_{iuis}^i, P_{iuiu}^i$ from data.

Consider now the socialization moments for heterogamous households with $d = 0$.

$$\begin{aligned} P_{isns}^i &= \tau_m + (1 - \tau_m - \tau_f)q^i, \\ \tau_m &= \frac{1}{\sigma_\tau}(\Delta V_{is}^{ns} - \Delta V_{ns}^{is})(1 - q^i), \quad \tau_f = 0 \quad \text{if } (\Delta V_{is}^{ns} - \Delta V_{ns}^{is}) > 0 \quad (\text{Het. in}) \\ \tau_f &= \frac{1}{\sigma_\tau}(\Delta V_{ns}^{is} - \Delta V_{is}^{ns})q^i, \quad \tau_m = 0 \quad \text{if } (\Delta V_{ns}^{is} - \Delta V_{is}^{ns}) > 0 \end{aligned}$$

$P_{isns}^i > q^i$ if $\tau_m > 0$ (which implies $\tau_f = 0$). Hence $\frac{\Delta V_{is}^{ns} - \Delta V_{ns}^{is}}{\sigma_\tau} > 0$. Once this sign is identified, then, the moment for P_{nsis}^i is linear in $\frac{\Delta V_{is}^{ns} - \Delta V_{ns}^{is}}{\sigma_\tau}$. Since we previously identified $\frac{\Delta V_{is}^{ns}}{\sigma_\tau}$, this identifies $\frac{\Delta V_{ns}^{is}}{\sigma_\tau}$.

The moment for P_{nuiis}^i is as follows.

$$\begin{aligned} P_{nuiis}^i &= \tau_f + (1 - \tau_m - \tau_f)q^i, \\ \tau_m &= \frac{1}{\sigma_\tau}(\gamma_n \Delta V_{ns}^{is} - \Delta V_{is}^{ns})(1 - q^i), \quad \tau_f = 0 \quad \text{if } (\gamma_n \Delta V_{ns}^{is} - \Delta V_{is}^{ns}) > 0 \quad (\text{Het. ni}) \\ \tau_f &= \frac{1}{\sigma_\tau}(\Delta V_{is}^{ns} - \gamma_n \Delta V_{ns}^{is})q^i, \quad \tau_m = 0 \quad \text{if } (\Delta V_{is}^{ns} - \gamma_n \Delta V_{ns}^{is}) > 0 \end{aligned}$$

It is linear in $\frac{\gamma_n \Delta V_{is}^{ns} - \Delta V_{ns}^{is}}{\sigma_\tau}$ and hence identifies it. Since we previously identified $\frac{\Delta V_{is}^{ns}}{\sigma_\tau}$, and $\frac{\Delta V_{ns}^{is}}{\sigma_\tau}$, this identifies γ_n .¹²

As a consequence, we have identified $\frac{\Delta V_{is}^{ns}}{\sigma_\tau}$ and $\frac{\Delta V_{ns}^{is}}{\sigma_\tau}$. Note that this is the case, independently, for every region r , as long as the distribution of the population q^i varies across region. Assuming that $\Delta V_{is}^{ns}, \Delta V_{ns}^{is}, \sigma_\tau$ are constant across region r , we cannot exploit the variation across regions to identify them separately.

From the divorce moments, we identify the means of the distribution of the marriage quality shock θ , a_{hj} ; the utility of a child within marriage δ , the socialization costs σ_τ , and the values of education S_i, S_n . Recall that δ denotes the utility of having a child within marriage; while θ denotes the marriage quality shock. The distribution of θ_{hj} has a mean that depends on the household type hj . θ_{hj} follows a generalized logistic distribution cumulative distribution function $F(\theta_{hj})$, with location a_{hj} and scale parameter b . We set $b = 1$. The parameters a_{hj} are identified independently from any other parameter

¹²In fact, γ_n is over-identified as it can also be identified by the moment for P_{nuiiu}^i , which is linear in $\frac{\gamma_n \Delta V_{is}^{ns} - \gamma_i \Delta V_{ns}^{is}}{\sigma_\tau}$.

in the model with data on the dissolution probability of couples *without children*, $\hat{\pi}_{hj}(0)$, for all h, j ; that is, a_{hj} is the unique solution to

$$F(0; a_{hj}, 1) = \hat{\pi}_{hj}(0), \text{ for all } h, j.$$

The moments for the probability of divorce, for any household type hj , are:

$$\pi_{hj}(N_{hj}) = F\left(N_{hj}(u_{hj}(\tau(1), 1) - c(\tau(1))) - N_{hj}(u_{hj}(\tau(0), 0) + \delta - c(\tau(0)))\right). \quad (\text{Div})$$

Consider, first, equation (Div) for $h = ns; j = ns$. Normalizing $V = 100$,

$$\begin{aligned} u_{nsns}(\tau, 0) &= 2V - 2P_{nsns}^u S_n \\ u_{nsns}(\tau, 1) &= 2V - 2P_{nsns}^u S_n, \end{aligned}$$

and hence the equation (Div) is independent of S_i, S_n . It is only a function of δ , which is then identified from homogamous households of natives. Consider now equations (Div) for homogamous households of immigrants, with $h = is; j = is$ and with $h = iu; j = is$, where

$$\begin{aligned} u_{isis}(\tau, 0) - c(\tau(0)) &= 2VP_{ii}^i(\tau, 0) + 2V_{is}^{ns}(1 - P_{ii}^i(\tau, 0)) - 2P_{isis}^u S_i - \frac{1}{2}\sigma_\tau(\tau_m + \tau_f)^2 \\ u_{isis}(\tau_f, 1) - c(\tau_f(1)) &= 2VP_{ii}^i(\tau_f, 1) + 2V_{is}^{ns}(1 - P_{ii}^i(\tau_f, 1)) - 2P_{isis}^u S_i - \frac{1}{2}\sigma_\tau(\tau_f)^2, \\ u_{iuis}(\tau, 0) - c(\tau(0)) &= (1 + \gamma_i)VP_{ii}^i(\tau, 0) + 2V_{is}^{ns}(1 - P_{ii}^i(\tau, 0)) + (1 - \gamma_i)S_i - 2P_{iuis}^u S_i - \frac{1}{2}\sigma_\tau(\tau_m + \tau_f)^2 \\ u_{iuis}(\tau_f, 1) - c(\tau_f(1)) &= (1 + \gamma_i)VP_{ii}^i(\tau_f, 1) + 2V_{is}^{ns}(1 - P_{ii}^i(\tau_f, 1)) + (1 - \gamma_i)S_i - 2P_{iuis}^u S_i - \frac{1}{2}\sigma_\tau(\tau_f)^2. \end{aligned}$$

These equations are a function of $\sigma_\tau, \frac{\Delta V_{is}^{ns}}{\sigma_\tau}, \delta, S_i$ and hence identify σ_τ and S_i . Notice that, since we previously identified $\frac{\Delta V_{is}^{ns}}{\sigma_\tau}, \frac{\Delta V_{ns}^{is}}{\sigma_\tau}$, we have now identified the cultural intolerances $\Delta V_{is}^{ns}, \Delta V_{ns}^{is}$.

From the fertility moments, we identify the fertility cost σ_N . Assuming quadratic fertility costs, $\kappa(N) = \frac{1}{2}\sigma_N N^2$, the fertility moments for any household type hj , are:

$$N_{hj} = \frac{1}{\sigma_N} \left(\pi(u_{hj}(\tau(1), 1) - c(\tau(1))) + (1 - \pi)(u_{hj}(\tau(0), 0) + \delta - c(\tau(0))) \right).$$

Consider these moments for $h = ns; j = ns$ and for $h = is; j = is$. Using these equations uniquely identify σ_N and S_n with data on N_{nsns}, N_{isis} , given the parameters previously identified. The moments for heterogamous households allow for over-identification.

From the marriage fractions and marriage gains, we identify the outside options

of marriages, $\omega_{h_z}, \omega_{j_z}$ for $z \in \{hom, het\}$. For each household type hj , the moment equations for the implied marital surplus net of outside options are

$$U_{hj} - \omega_{h_z} - \omega_{j_z} = G_{hj}$$

where gains from marriage are defined as:

$$G_{hj} = \log \frac{(\mu_{hj})^2}{\mu_{h0} \cdot \mu_{0j}}; \quad (1)$$

and U_{hj} is the utility of household hj , which is identified given the parameters of the model we have already identified. In turn, the marriage distribution μ_{hj} and fractions of singles μ_{h0}, μ_{0j} are data. These moments then depend on the parameters already identified and linearly on ω_{h_z} and ω_{j_z} . In this particular case, with 4 possible individual types $h, j = ce$ and two household types $z \in \{hom, het\}$, we have $4 \times 2 = 8$ outside options ω_{ce_z} and $4^2 = 16$ gains from marriage, so that the outside options are identified.

In the more general model, with 7 cultural-ethnic traits in the population, we also identify the relative effects of marriage on spouses' economic opportunities, α_z , with $z \in \{hom, het\}$. Given our specification,

$$U_{hj}^{ec} = \alpha_z (\omega_{h_z} + \omega_{j_z});$$

for each household type hj , the moment equations for the implied marital surplus net of outside options are:

$$U_{hj} + (\alpha_z - 1)(\omega_{h_z} + \omega_{j_z}) = G_{hj}.$$

In this system, with 14 possible individual types $h, j = ce$ and two household types z , we have $14 \times 2 = 28$ outside options plus 2 α_z parameters and $14^2 = 196$ estimated gains from marriage, so that the outside options and the α_z parameters are overidentified.

References

- Adda, Jérôme, Paolo Pinotti, and Giulia Tura.** 2025. “There’s More to Marriage Than Love: The Effect of Legal Status and Cultural Distance on Intermarriages and Separations.” *Journal of Political Economy*, 133(4): 1276–1333.
- Brunello, Giorgio, and Daniele Checchi.** 2007. “Does School Tracking Affect Equality of Opportunity? New International Evidence.” *Economic Policy*, 22(52): 782–861.
- Card, David.** 2001. “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration.” *Journal of Labor Economics*, 19(1): 22–64.
- Carlana, Michela, Eliana La Ferrara, and Paolo Pinotti.** 2022. “Goals and Gaps: Educational Careers of Immigrant Children.” *Econometrica*, 90(1): 1–29.
- Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss.** 2017. “Partner Choice, Investment in Children, and the Marital College Premium.” *American Economic Review*, 107(8): 2109–67.
- Lundberg, Shelly, and Robert A Pollak.** 1993. “Separate Spheres Bargaining and the Marriage Market.” *Journal of Political Economy*, 101(6): 988–1010.
- Spolaore, Enrico, and Romain Wacziarg.** 2009. “The Diffusion of Development.” *Quarterly Journal of Economics*, 124(2): 469–529.
- Spolaore, Enrico, and Romain Wacziarg.** 2016. “Ancestry, Language and Culture.” In *The Palgrave Handbook of Economics and Language*. 174–211. Springer.
- Tabellini, Marco.** 2020. “Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration.” *Review of Economic Studies*, 87(1): 454–486.