UNIVERSITA’ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE ECONOMICHE ED AZIENDALI “M.FANNO”

CORSO DI LAUREA MAGISTRALE IN ECONOMICS AND FINANCE

TESI DI LAUREA

“THE EFFECTS OF PARENTS’ EDUCATION ON ASSORTATIVE MATING: EVIDENCE FROM NINETEEN COUNTRIES”

RELATORE:

CH.MO PROF. GIORGIO BRUNELLO

LAUREANDO: STEFANO GALLINARO

MATRICOLA N. 1128087

ANNO ACCADEMICO 2016 – 2017

Il candidato dichiara che il presente lavoro è originale e non è già stato sottoposto, in tutto o in parte, per il conseguimento di un titolo accademico in altre Università italiane o straniere.
Il candidato dichiara altresì che tutti i materiali utilizzati durante la preparazione dell'elaborato sono stati indicati nel testo e nella sezione "Riferimenti bibliografici" e che le eventuali citazioni testuali sono individuabili attraverso l'esplicito richiamo alla pubblicazione originale.

Firma dello studente

____

“It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.”

Jane Austen, *Pride and Prejudice*

Abstract

The literature on assortative mating usually focuses on how individuals cope on the basis of their characteristics and on which effects the match has on the two people. This paper tries to investigate whether assortative mating is influenced by parental background more than by individuals' peculiarities. Focusing on education we run a cross-countries OLS estimation in order to prove the presence of an effect of parents' education on partner's and investigate how this effect differs from country to country. Later on, we also use IV regression models in order to get rid of any endogeneity affecting our independent variables. We found evidence of a strong effect of parents' education on the characteristics of their children's partners; this effect is bigger in countries where social mobility is lower and vice-versa, former USSR countries have the lowest coefficients. IV estimation confirms our results and suggests the real effect might be even greater than the one estimated by OLS. Because of the lack of data on parents' characteristics, this paper cannot provide a precise estimate of the magnitude of the effect of parental background on assortative mating but it can be considered as an initial step toward a more consistent explanation of the phenomenon.

Contents

Index of Tables

Index of Figures

Introduction

Economists have been investigating assortative mating and the choice of the partner since the last decade of the XX century. Despite it could appear as topic closer to sociology than to economics, several consequences on social mobility, equality and distribution of income are proved to exist.

Stratification of society in classes is, in fact, a phenomenon which dates back to the origin of human civilizations and still persists. Even if we can firmly claim our society to be much more mobile than in past, many steps are still to be made in order to get an environment where economic background does not matter in determining who you are going to become.

Marriages have always been a way through which families try to improve or at least maintain their status, to the point that very often marriages were more of an economic than a romantic issue. For example, Glückel of Hameln made a purpose of her life to marry her sons and daughters in the most prominent Jewish families of several European countries in order to get insurance against shocks that could affect a region instead of another (Kaufmann, 1896).

In this kind of world, it is easy to understand how the accumulation of a dowry for his daughter's wedding was one of the most important and difficult purposes of a man. Depending on what a father could save for her, a girl's perspective could radically change.

For these reasons, we think that the economic literature on assortative mating, though broad and widespread, still misses an important aspect: the role of parental background in the choice of the partner.

In fact, even if a great number of authors investigated how individuals pair in terms of personal and economic characteristics (i.e. education, income, status, etc …) and returns people get from getting married, very few papers deal with the influence of parents in this decision.

Probably, this is due to the idea of wedding and partners as something regarding the most intimate dimension of a person so that it might sound absurd other people's characteristics can matter. On the other hand, if we meditate on how such a great part of who we are is determined by factors not depending on us (e.g.: where we were born) it does not seem so strange external factor could affect who we choose.

All of the literature on intergenerational persistence is indeed based on this hypothesis: much our scope of action depends on conditions already set before our birth.

After these considerations, it appears unreasonable to have plenty of papers on how parental background affects education, earnings or occupation but only a few studies considering its effect on partner's characteristics.

The lack of relevant literature, and the importance we think this aspect has, made us choose this topic. Our aim is to put the light on this issue so that other researchers would find interesting to inspect the subject deeper in the near future.

In this paper, we want to show a preliminary investigation on the effects of parental background on the choice of the partner. In particular, we want to determine if parents' education can predict partner's years of schooling. We are focusing on education as it can be considered as an easy measurable proxy for many other observable and unobservable. Our analysis will cover different countries and try to solve those endogeneity problems which may have stopped previous academics from going deep on the subject.

After this brief introduction, we are going through a review of the literature related to assortative mating and intergenerational mobility. Then, we divided the main part of this dissertation into four parts we called Chapters containing the description of data, models and results. In Chapter 1, we present a comparison among seventeen European countries basing our analysis on SHARE database through a multivariate OLS regression model. In Chapter 2, we restrict the scope of our analysis only on countries once part of the Nazi Reich in order to exploit the exposure of individuals to World War II as the instrument in an IV model. Later, in Chapter 3, we run the same regressions on a different a database, pairfam, in order to see whether our findings can get external validity. In Chapter 4, finally, we put the two databases together to exploit an even greater variability. The last part of the paper is dedicated to sum up the findings and conclude.

Related Literature

This paper relates to the literature on assortative mating and to the one on intergenerational persistence too.

Actually, our paper is quite unique in its aim: identify the effect of parental background on partner's quality. To our knowledge, the only study with a similar focus is the one by Hu (2006) investigating the relationship between father's occupation and spouse's education in China. He finds a strong influence of parents' characteristics on the choice of the partner. Moreover, he finds a high correlation between parents' characteristics and parents in law's.

A much broader literature has instead developed on assortative mating in recent years. The most famous study on this topic has been led by Lefgren and McIntyre (2006), they found a strong correlation between women's education and husbands' income and marital status. Through Data obtained by the 2000 Census, they claim that about half of the increase in a woman's available income predicted by higher education is attributable to the marriage market.

Another interesting research is the one by Kaufmann et al. (2013) where they investigate the relationship between education and partner's quality in Chile. According to their paper returns of education in terms of partner's quality is twice for women than for men.

In fact, the vast majority of authors looked for a correlation between spouses' characteristics as education or earnings. This is, for example, the case of Quian & Quian (2017) who relates marriage outcomes to education and province of origin in China.

On the other hand, several authors also investigated the effects of marriage on the returns of education. E.g. Liu (2011) proved that getting married boost returns of education for women with less than two years of college in Sweden. A similar conclusion has been reached by Huang, Li et al. (2009) in their paper on cross-productivity between spouses exploiting Chinese twins data in order to control for unobserved characteristics.

Lam & Schoeni (1994) found evidence of an influence of parents' and parents in law's characteristics on individual's earnings in both the United States and Brazil.

As the main channel through which the parental background affects partner's choice is individual's own characteristics, this paper also relates to the literature on intergenerational educational persistence.

On this topic, a very impressive work is the one by Hertz et al. (2007) where they look for the correlation between parents and children's education in 42 countries in a 50-year long period.

They found the presence of a decreasing-in-years predictive power of parents' education on their children's achievements.

Daouli, Demoussis, and Giannakopoulos (2010) explore the educational persistence through three Greek cohorts finding a quite high rate of social mobility but the presence of little educational persistence in women.

Another interesting contribute comes from Azam (2016) examining the intergenerational educational persistence among daughters in India. According to his conclusions, daughters' probability to attend secondary education is highly correlated to father's schooling and this predictive power is not declining in time.

Dustmann (2004) looked at how parental background influences the choice of the school track in Germany and its consequences in terms of occupation and earnings.

This paper contributes to the literature in the following ways.

First, as it is very difficult to separate parental background and marriage patterns from country-specific culture, the great part of the quoted researches focus on only one country. Differently, our study aims to make a comparison among 17 European countries.

Second, besides Lofgren & McIntyre (2006) only DeSilva & Bakthiar (2011) use an Instrumental Variable approach to solve the endogeneity of education in assortative mating. They chose to use, respectively, the quarter of birth and the order of birth to instrument individual's education. In order to solve the endogeneity of parents' education, we are using the exposure to the World War II as an exogenous variable.

Finally, to our knowledge, nobody have ever tried to estimate the effect of parental background on partner's quality in Europe.

Chapter 1 – Cross-Country OLS Estimation

Data

To investigate how parental background affect individuals' chances in the marriage market, we need data on individuals' characteristics, on their partners' and on their parents'. Unfortunately, it is very difficult to find such a database as most of the surveys conducted by research institutes usually focus either on the present condition or on lifelong perspective. Which is why in most of the publicly accessible databases we can find information either on partners or on parents but not both.

The Survey on Health, Ageing, Retirement and Education (SHARE) is a unique multidisciplinary and cross-national panel database of micro data on health, socio-economic status and social and family networks of more than 120,000 individuals aged 50 or older. The first wave of SHARE has been released in 2004 and since then it improved in the depth of analysis and dimension. Through the six waves of data which have been published until now, researchers collected information from 27 countries regarding different aspects of interviewed households.

Given the fact that any wave refers to a different survey, not all they contain the same information. For this reason, I only referred to information collected in waves 5 and 6, which were collected through surveys in 2013 and 2015.

Inside SHARE database, I was able to get information about education of individuals, their partner and their parents. I considered years of education for both individuals and their partners while, due to the way in which observations had been collected, I used ISCED 1997 coding for their parents. As ISCED is a reclassification of educational attainments through common standards, I could compare individuals of different countries despite every country having a specific and idiosyncratic educational system.

ISCED 1997 codification is composed of six categories corresponding to:

- ISCED 0   Pre-primary level of education
  Initial stage of organised instruction, designed primarily to introduce very young children to a school-type environment.

- ISCED 1 Primary level of education

  Programmes normally designed to give students a sound basic education in reading, writing and mathematics.

- ISCED 2 Lower secondary level of education.

  The lower secondary level of education generally continues the basic programmes of the primary level, although teaching is typically more subject-focused, often employing more specialised teachers who conduct classes in their field of specialisation.

- ISCED 3 Upper secondary level of education

  The final stage of secondary education in most countries. Instruction is often more organised along subject-matter lines than at ISCED level 2 and teachers typically need to have a higher level, or more subject-specific, qualification that at ISCED 2. There are substantial differences in the typical duration of ISCED 3 programmes both across and between countries, typically ranging from 2 to 5 years of schooling. Therefore, these programmes lead directly to labour market, ISCED 4 programmes or other ISCED 3 programmes.

- ISCED 4 Post-secondary, non-tertiary education

  These programmes straddle the boundary between upper secondary and post-secondary education from an international point of view, even though they might clearly be considered as upper secondary or post-secondary programmes in a national context. These programmes are often not significantly more advanced than programmes at ISCED 3 but they serve to broaden the knowledge of participants who have already completed a programme at level 3. The students are typically older than those in ISCED 3 programmes. They typically have a full-time equivalent duration of between 6 months and 2 years. These programmes lead directly to labour market or other ISCED 4 programmes.

- ISCED 5 First stage of tertiary education

  Programmes with an educational content more advanced than those offered at levels 3 and 4.

- ISCED 6 Second stage of tertiary education
  This level is reserved for tertiary programmes that lead to the award of an advanced research qualification. The programmes are devoted to advanced study and original research.

When coming to consider years of education for individuals' parents I had to impute them, as this kind of information is not included in SHARE database. In order to do that I considered the average schooling for any ISCED category in each country of the sample. If for example, the average schooling for individuals with an ISCED 3 was 12 in Italy, I imputed to fathers and mothers with ISCED 3 12 years of education.

The countries represented in the sample are 17 (Austria, Germany, Sweden, Netherlands, Spain, Italy, France, Belgium, Denmark, Switzerland, Greece, Israel, Czech Republic, Luxembourg, Slovenia, Estonia and Croatia). Observation for Greece and Croatia are present only in wave 6 while observations for Netherlands belong exclusively to wave 5.

The variety of countries in our sample it is very important because, even though we could think that Europeans share a lot in terms of cultural heritage and social norms, during the period in which the individual of our sample and their parents studied and got married Europe was parted in two completely different spheres of influence.

In fact, there is no doubt that whether living in a NATO member state or in a Communist country did make a huge different in terms of education, job opportunities and way of life.

Descriptive Patterns

With the aim to describe our sample, in Table 1.1 we see the composition of the sample by country of origin and gender. Globally our sample contains 25 452 men and 30 956 women for a total of 56 408 individuals aged between 50 and 75.

Looking at Table 1.2 we see how individuals' average years of education is slightly higher than 11 and the value for their partners is very similar but when differentiating by gender we see that men have half a year of education more than women on average. The gap between genders is bigger than one year in Germany, Greece and Croatia while in Sweden and Estonia women are on average more educated than men (Table 1.3).

Table 1.4 describes parental education. We reported the percentage of fathers and mothers having an ISCED equal or higher than 3 i.e. having completed at least secondary education. Big differences can be observed both among countries and gender. North Europe countries have generally a higher fraction of high-educated parents but in South Europe countries, we can see a smaller difference between men women. In no countries, women are on average more educated than men.

Table 1.1 – Composition of the sample by gender

Country	Male	Female	Total
Austria	1412	1865	3277
Germany	2065	2287	4352
Sweden	151	1795	3305
Netherlands	1403	1738	3141
Spain	2103	2398	4501
Italy	208	2521	4601
France	1555	1911	3466
Denmark	1622	1847	3469
Greece	925	1107	2032
Switzerland	1031	1249	228
Belgium	2249	2611	486
Israel	722	947	1669
Czech Republic	1844	2526	437
Luxembourg	774	859	1633
Slovenia	1471	1774	3245
Estonia	1723	2371	4094
Croatia	916	1115	2031

This pattern is not surprising because, as we know from Barro and Lee, women caught up in education only from the cohorts around 1950. As in our sample individuals were born from 1940 until 1965 for sure their parents belonged to a generation where boys were expected to be more educated than girls (Barro and Lee, 2010).

Moreover, as we can see from Table 1.5 and 1.6 for some individuals' parents we could not retrieve information about their education. These individuals will not take part in the sample we are using for our regression.

Finally, we have to state that, as we can see from Table 1.7, people born later in time tend to be more educated in every country but Austria and Switzerland. This could easily be explained with better access to education due to increasing wealth and welfare systems.

Table 1.3 – Summary of the variables of interest

	Observations	Average	Standard Deviation	Min	Max
Years of Education	56 408	11 570	4 223	0	25
Partner's Education	44 358	11 584	4 204	0	25
Father's Education	56 408	1 906	1 547	0	6
Mother's Education	56 408	1 534	1 312	0	6

Table 1.2 – Average schooling by gender and country

Country	Male	Female
Austria	9,427	9,179
Germany	13,312	12,352
Sweden	11,713	12,071
Netherlands	12,512	11,571
Spain	9,580	9,123
Italy	9,798	8,944
France	12,195	11,888
Denmark	13,809	13,407
Greece	10,858	9,756
Switzerland	9,184	8,417
Belgium	12,949	12,324
Israel	12,852	12,651
Czech Republic	12,733	11,894
Luxembourg	11,114	10,546
Slovenia	11,263	9,999
Estonia	9,427	9,179
Croatia	13,312	12,352

Table 1.4 – Percentage of parents with at least secondary education by country

Country	Father	Mother
Austria	60%	29%
Germany	79%	43%
Sweden	22%	13%
Netherlands	21%	10%
Spain	5%	3%
Italy	6%	5%
France	22%	15%
Denmark	57%	28%
Greece	13%	5%
Switzerland	59%	32%
Belgium	30%	21%
Israel	38%	33%
Czech Republic	68%	41%
Luxembourg	34%	15%
Slovenia	34%	31%
Estonia	10%	6%
Croatia	60%	29%

Table 1.5 - Missing value for father's Education

Country	Refusal	Don't Know	Still in education	Other	Total
Austria	3	100	0	3	106
Germany	5	200	0	5	210
Sweden	14	108	0	25	147
Spain	8	46	1	1	56
Italy	23	52	0	1	76
France	5	120	0	1	126
Denmark	1	76	0	3	80
Greece	2	15	0	2	19
Switzerland	1	58	0	6	65
Belgium	1	255	2	18	276
Israel	25	10	0	1	36
Czech Republic	1	4	0	7	12
Luxembourg	1	56	0	3	60
Slovenia	64	134	1	1	200
Estonia	7	77	0	12	96
Croatia	6	30	0	0	36
Total	167	1341	4	89	1601

Table 1.6 - Missing values for mother's Education

Country	Refusal	Don't Know	Still in education	Other	Total
Austria	2	31	0	2	35
Germany	0	81	0	2	83
Sweden	6	46	0	13	65
Spain	2	22	0	0	24
Italy	22	34	0	0	56
France	3	71	0	1	75
Denmark	0	22	0	0	22
Greece	1	8	0	0	9
Switzerland	2	32	0	3	37
Belgium	1	155	0	6	162
Israel	22	7	1	2	32
Czech Republic	1	1	0	2	4
Luxembourg	0	34	1	1	36
Slovenia	52	42	0	0	94
Estonia	0	36	1	1	38
Croatia	3	11	0	0	14
Total	117	633	3	33	786

Table 1.7 – Average schooling by cohort and country

Country	'41-'45	'46-'50	'51-'55	'56-'60	'61-'65
Austria	9.377	9.113	9.414	9.601	9.318
Germany	12.531	13.026	13.236	13.227	13.359
Sweden	13.087	11.592	12.297	12.932	12.973
Netherlands	11.138	12.156	12.489	13.374	13.190
Spain	8.461	9.091	10.359	11.328	11.849
Italy	8.099	8.970	10.219	10.810	12.056
France	11.812	12.047	12.516	12.703	13.471
Denmark	12.902	13.722	14.021	14.419	14.153
Greece	8.354	9.348	10.428	11.652	12.116
Switzerland	8.651	9.109	8.874	8.645	7.724
Belgium	12.062	12.694	13.022	13.292	13.036
Israel	12.211	13.009	13.195	13.310	13.643
Czech Republic	11.999	12.508	12.334	12.600	12.738
Luxembourg	11.786	12.004	11.735	12.267	13.443
Slovenia	10.865	11.099	10.978	11.277	11.278
Estonia	11.839	12.247	12.549	12.846	12.791
Croatia	9.904	11.057	10.925	11.102	11.685

Methodology

The aim of our research is trying to identify the effect of parental background on the choice of the partner. In order to do so, we have to define which characteristic of the individuals, their parents and their partners we want to take into consideration. The characteristic we selected is education. We made this choice for two main reasons:

- It is a good proxy for many other characteristics as for example earnings and health;
- It is a variable that can be easily compared across countries.

We expect high-educated parents to be correlated with high-educated partners for their children. With this hypothesis, we try to link the findings of Lefgren about assortative mating and those of Dustmann about effects of parental background on education. In fact, the question driving our research is whether parental background affects outcomes in the marriage market.

It is common in several cultures, both in the past and at present day, for the head of the family to choose his children's spouses on the base of economic convenience. These arrangements between families often prevented individuals from marrying up or down and hindered intergenerational social mobility. The idea underlying our quest is that, even if arranged marriages are not common in Europe since the XIX century, parents can still influence their marriage perspective through other channels.

In the past, when girls entered the marriage market were evaluated on the base of the dowry they could bring plus a bunch of other personal characteristics. Men, instead, were ranked on the basis of their expected income. Trying to replicate this pattern at the present day we could think as female education also as a dowry making girls able to get a better husband.

Consistent with this thesis are the results obtained by Lefgren and McIntyre (2006) showing that a higher education affects women's well-being boosting both the probability of getting married and their expected return on the marriage market. Going on the same the direction, we are expecting to find that the effect of parental education on the quality of the partner is higher for women than for men.

Ideally, we would like to run a simple OLS regression of this kind:

$$y_i^p = \alpha_0 + \alpha_1 mother_i + \alpha_2 father_i + \boldsymbol{\beta X}_i + \varepsilon_i \tag{1}$$

Where the dependent variable y_i^p is years of education of the i^{th} individual's partner. $father$ is a dummy equal to one if the ISCED code of the i^{th} individual's father is at least 3 and $mother_i$ is an identical indicator for mother's education. $\boldsymbol{X}_i$ is a set of dummy variable controlling for which of the five-year cohorts we divide the century the individual was born and his age when data had been collected and the wave to which the observation belongs.

Of course, running this simple model will put us in the same position of Dustmann (2004): finding a high and significant effect of parental education on partner's education but not being able to address how much of that coefficient represents the direct effect and how much is instead a bias due to other individual's characteristics correlated with the education of his parents.

In fact, there could be many channels through which parental education can affect the choice of the partner. E.g: better parents have better children who look for a better partner.

We think that the main channel through which these two variables are connected is individual's education. Several authors discussed the relation between parents' education and children's (Chevalier, 2004, Dubow, 2009) and no fewer economists investigated educational assortative mating (Eika, 2014, Mare, 1991).

Taking into consideration this issue, it would come naturally to add a term for individual's education in our model. Thus, it would like:

$$y_i^p = \alpha_0 + \alpha_1 mother_i + \alpha_2 father_i + \alpha_3 y_i \; + \boldsymbol{\beta X}_i + \varepsilon_i \qquad (2)$$

Where y_i is years of education of the individual i.

However, this model could be a good starting point if we wanted to estimate α_3[1]. As we are looking for the value of α_1 and α_2, controlling for the education of the individual will bias our estimate. This is because y_i comes later in time with respect to $mother_i$ and $father_i$ and most likely it is an outcome of these two variable. We are facing a classical problem of bad control (Angrist & Pischke, 2008).

In this first chapter, we are going to exploit equation (1) using both the dummy and the years of the education.

In the first case, we decide to use a dummy variable of this kind instead of adding a vector for the ISCED because we are not interested in the changing of the outcome depending on the level of parental education but we want to infer the difference made by having "a certain kind of family".

We put the threshold at level 3 considering that at the beginning of the XX century only a few people could afford secondary education. Putting the threshold at 5 would have restricted the sample too much while choosing level 2 would have probably been meaningless for those countries where the attendance rate of elementary school was already high.
Later we also ran a regression using a variable mother and one using a variable $parent$ instead of $father$. $mother$ is an indicator built exactly as $father$ but considering the education of the ith individual's mother. Parent is a dummy for at least one of the parents with ISCED higher than 3.

[1] Actually, y_i is endogenous too. Thus, we would need a different model even to estimate α_3.

In running this kind of regression, we are aware of two main problems. First, the sample is not random as the higher the level of education the higher the probability of getting married. Second, parental education can affect the education of the partner both directly and through the education of the individual.

These issues can be solved by using other estimation techniques as IV regression, which is what we are going to show in the following sections. In this part of the analysis, we prefer to use OLS because it gives us more precise estimates, even if biased. In fact, here we are not interested in finding the absolute value of *father* but we want to describe how it varies across countries. For this reason, we can assume the error term to be composed of two parts: one common to all the countries and country-specific one.

Assuming the bias to be the same across countries of our sample is not that unrealistic if we look at the work of Hajnal (1965) in identifying a common pattern of marriage for Western Europe. Major differences in families formation and composition can be found if comparing different continents like Europe and Eastern Asia (Lund and Kurosu, 2014).

Thus, we suppose the error term to have this form:

$$\varepsilon_i = \delta + \epsilon_i \tag{3}$$

Where δ is the part of the bias which is equal among the countries and ϵ_i is instead a term such that $E[\epsilon_i | y_i, father_i, X_i] = 0$.

Given the assumptions above, we are no more interested in the absolute value of coefficient α_2 but in:

$$\bar{\alpha}_2 = \alpha_{2j} - \alpha_{2k} \text{ , with } j \neq k \tag{4}$$

being j , k countries of our sample.

Moreover, as we can notice in Table 1.6 and Table 1.7, for many countries the differences in the probability of marriage between those with high-educated parents and those without is not even significantly different from zero. Thus, at least for some countries, we could assume that education of father is orthogonal to the probability of marriage.

Table 1.7 - Differences in Probability of Marriage for Men

	Father=1	Father=0	Difference in Probability	
Austria	0,782	0,816	0,033	
Germany	0,842	0,813	-0,028	
Sweden	0,841	0,851	0,010	
Netherlands	0,820	0,848	0,029	
Spain	0,790	0,875	0,085	***
Italy	0,874	0,855	-0,019	
France	0,841	0,793	-0,048	***
Denmark	0,849	0,823	-0,026	
Greece	0,837	0,863	0,026	
Switzerland	0,831	0,865	0,033	
Belgium	0,779	0,807	0,029	
Israel	0,918	0,899	-0,019	
Czech Republic	0,853	0,850	-0,003	
Luxembourg	0,861	0,859	-0,002	
Slovenia	0,856	0,867	0,011	
Estonia	0,847	0,814	-0,033	
Croatia	0,825	0,879	0,053	

Table 1.8 – Differences in Probabilty of Marriage for Women

	Father=1	Father=0	Difference in Probability	
Austria	0,619	0,677	-0,058	***
Germany	0,813	0,794	0,019	
Sweden	0,716	0,776	-0,059	***
Netherlands	0,756	0,811	-0,055	***
Spain	0,754	0,827	-0,074	***
Italy	0,743	0,763	-0,020	
France	0,714	0,687	0,027	
Denmark	0,792	0,793	-0,001	
Greece	0,605	0,722	-0,118	
Switzerland	0,759	0,763	-0,004	
Belgium	0,692	0,733	-0,041	***
Israel	0,789	0,794	-0,005	
Czech Republic	0,660	0,664	-0,004	
Luxembourg	0,779	0,782	-0,004	
Slovenia	0,776	0,779	-0,003	
Estonia	0,652	0,617	0,035	
Croazia	0,726	0,770	-0,044	

Results

In this section, we are going to go through the results of the estimation of Equation 1. Taking into account the possible problems arising because of the endogeneity of our parameters of interest we are going to talk about our estimates in relative terms.

Before going to estimates of coefficients in Equation 1, we think it would be interesting to show results when the dummy states simply if at least one of the parents has secondary education. Thus having a model built in this way:

$$y_i^p = \alpha_0 + \alpha_1 parent_i + \boldsymbol{\beta X}_i + \varepsilon_i \tag{5}$$

Where $parent_i$ is equal to 1 when at least one of the parents as an ISCED equal or higher than 3 and 0 otherwise.

Results of this estimation are shown in Table 1.9.

In Table 1.10. we can find the estimates of the dummy father and mother in all of the countries in our sample for both males and females. Looking at the first two columns, those reporting estimates for men, we can see the highest coefficients belongs to Southern European countries as Italy, Spain and Greece. One first hypothesis is that these results depend on the lower percentage of high-educated parents we can find in these countries. In fact, the advantage of having a secondary education is more relevant if the population of high-educated people is smaller.

Anyway, this is not said to be the only possible explanation. If it were, in fact, also Estonia, which its percentage of secondary-educated parents is among the lowest, should have a relatively high coefficient for both mother and father. An interesting explanation could be the belonging of Estonia to the USSR: during Communist Era, especially in the early years, social classes were weakly related to education as often people were assigned to schools, jobs or places to live by the central government (Fitzpatrick, 2002).

This hypothesis is strengthened by the low coefficients of Czech Republic, another country under the influence of the Soviet Union from 1948 until 1989.

Thus, we can imagine that the completing of at least secondary education is directly linked to the belonging to a high social class. In this case, bigger coefficients in countries with few people going to high school could mean that the children of highly educated parents are more likely to marry high educated partner, not because of education itself but because only children belonging to a certain social class went to high school. Consistent with this high hypothesis is the fact that the countries with big coefficients in our model are the those with low intergenerational mobility and vice-versa in the literature. (Comi, 2009, Jäntti et al., 2006, Blanden, 2005)

Another thing we can notice in Table 1.8 is that coefficients of $father$ are higher for women than for men in almost all of the countries. This is consistent with our previous hypothesis of female education more as an investment in relational capital than in human capital as women participation to the labour market, especially in Mediterranean Europe, was quite poor.

Comparing the results of this table with those of the previous one, we see that coefficient of Table 1.9 are closer to those of $father$ for women and to those of $mother$ for men. This suggests a sort of cross-gender effect we are going to discuss later on in this paper.

When looking at Table 1.9 we see a similar estimation but using years of education instead of a dummy stating whether parents completed at least secondary education. Here the differences between countries are smaller and it is not easy to find a pattern as we did for Table 1.8. Ideally, we would expect countries on the top of the chart in the previous table to be the same at the top here. The reason for this inconsistency could be the fact the education has a similar effect in all of the countries but affecting different individuals. In fact, here we are considering the effect of one additional year of education irrespective of this being one year of primary, secondary school or college.

Table 1.9 - OLS estimation of parents' education on partner's education

Country	Men		Women	
Austria	0,846	***	1,486	***
Germany	1,13	***	1,72	***
Sweden	1,544	***	1,595	***
Netherlands	1,686	***	2,095	***
Spain	3,573	***	4,952	***
Italy	4,166	***	4,35	***
France	1,804	***	2,317	***
Denmark	0,894	***	1,211	***
Greece	2,685	***	2,585	***
Switzerland	1,712	***	2,082	***
Belgium	1,538	***	2,13	***
Israel	2,784	***	2,496	***
Czech Republic	1,345	***	1,173	***
Luxembourg	2,374	***	3,899	***
Slovenia	1,946	***	1,827	***
Estonia	0,984	***	1,631	***
Croatia	2,593	***	2,453	***

Table 1. – OLS estimation of parent’s education on partner’s education

Country	Men				Women			
	father		mother		father		mother	
Austria	0,067		1,709	***	0,531		2,001	***
Germany	0,546	***	1,009	***	1,332	***	0,906	***
Sweden	1,266	***	0,983	***	1,230	***	1,097	***
Netherlands	1,304	***	1,375	***	1,745	***	1,437	***
Spain	3,015	***	1,844	***	4,639	***	1,947	***
Italy	2,963	***	3,356	***	3,786	***	1,939	***
France	1,366	***	1,188	***	1,590	***	1,584	***
Denmark	0,635	***	0,850	***	1,064	***	0,757	***
Greece	1,835	***	2,591	***	2,029	***	1,689	***
Switzerland	1,484	***	0,392		1,629	***	0,982	***
Belgium	1,216	***	1,027	***	1,454	***	1,379	***
Israel	1,575	***	1,829	***	1,537	***	1,674	***
Czech Republic	0,928	***	0,928	***	0,522	***	1,079	***
Luxembourg	1,869	***	1,442	***	3,256	***	2,162	***
Slovenia	1,465	***	1,206	***	1,316	***	1,670	***
Estonia	0,309		1,003	***	0,983	***	1,145	***
Croatia	1,731	***	2,239	***	2,081	***	0,697	

Table 1.9 - OLS estimation of parent's years of schooling on partner's education

Country	Men				Women			
	father		mother		father		mother	
Austria	0,352	***	0,163		0,392	***	0,129	
Germany	0,199	***	0,295	***	0,318	***	0,247	***
Sweden	0,191	***	0,173	***	0,194	***	0,196	***
Netherlands	0,197	***	0,366	***	0,387	***	0,321	***
Spain	0,385	***	0,384	***	0,517	***	0,377	***
Italy	0,254	***	0,518	***	0,411	***	0,284	***
France	0,300	***	0,203	***	0,315	***	0,336	***
Denmark	0,122	***	0,122	***	0,169	***	0,130	***
Greece	0,228	***	0,372	***	0,207	***	0,363	***
Switzerland	0,128	***	0,098		0,171	***	0,190	***
Belgium	0,190	***	0,215	***	0,221	***	0,267	***
Israel	0,166	***	0,361	***	0,262	***	0,242	***
Czech Republic	0,380	***	0,377	***	0,270	***	0,300	***
Luxembourg	0,320	***	0,259	***	0,514	***	0,404	***
Slovenia	0,303	***	0,224	***	0,277	***	0,297	***
Estonia	0,096	***	0,140	***	0,156	***	0,228	***
Croatia	0,255	***	0,171	***	0,155	***	0,210	***

Chapter 2 – IV Estimation

As already said, assuming parent's education to be exogenous when estimating individuals' or their partner's education is not straightforward and several authors take it as endogenous (Lefgren, 2006, Havari, 2014). For this reason, besides the OLS estimates we retrieved in the previous section we also want to try to solve this possible endogeneity using Instrumental Variable estimation.

To our knowledge, only two paper try to estimate the effect of parents' education on spousal outcomes. In their 1994 work, Lam and Schoeni do not use instrumental variables only considering the total effect of parental education on spousal earnings before and after controlling for individuals' own income (Lam & Schoeni 1994). In the same way, also Hu in his model estimates the effect of parental background on assortative mating without correcting for possible endogeneity of the model (Hu 2016).

Both of the authors admit they might overestimate or underestimate the effect of parental background but they cannot avoid the omitted variable bias because of the lack of the data. Poor information about parents' individuals is actually a problem we faced in our estimation too. Anyway, we tried to do our best in order to propose an alternative solution to the problem.

We chose to use World War II as an instrument in the fashion of Ichino & Winter-Ebmer (2004) as the year of birth was the only information we could infer for parents' individuals in both waves 5 and 6 of SHARE dataset. How we imputed the year of birth when missing and the assumption of our IV model are going to be discussed in the following paragraph.

Data

Unfortunately, SHARE dataset is not generous when talking of information about parents. In fact, the only useful information we can exploit is the age of these people if they are still alive. Taking into consideration that our individuals are aged between 50 and 75 the chance their parents are both alive is not that high.

Being our data set made up by more than 56 thousand individuals, we could think about having at least 100 thousand observations for parents' year of birth. Actually we have a little more than 10 thousand observations, definitely too few for our cross countries analysis. For this reason, we decided to impute the year of birth of parents for those individuals with missing observations.

Imputation method

For those individuals with at least one of the parents alive, we decided to ascribe the year of birth of the dead one on the base of the age of his significant other. We computed the average difference between spouses for each year of birth of the wife and summed to the age of women for those whose husband was dead.

When both of the parents were already dead at the time of the interview we computed the average difference between individuals' age and their parents' one for those with parents alive. In this way, we find the average age of people when becoming fathers or mothers. Summing this number to individuals age we have a hypothetical age of parents if they were alive. Our imputation method is supported by the fact that the average age at birth we find along the cohorts is very close to those reported by statistics (Lappegård, 2000).

In Table 2.1, we can see the father's age imputation method in details. On the base of mother's or individual's age we impute father's age adding the average difference for every cohort.

Table 2.1 - Imputated father's age by cohort

When Mother Alive	
Mother's Age	Imputated Age for Father
Lower than 75	Mother'age + 1
From 75 to 81	Mother'age + 2
Higher than 81	Mother'age + 3
When Both Parents Dead	
Individual's Age	Imputated Age for Father
Lower than 52	Individual's Age + 28
From 52 to 58	Individual's Age + 26
From 58 to 63	Individual's Age + 25
Higher than 63	Individual's Age + 24

Instrument

Being the charge of the year of birth the only relevant information about parents in our possession, we had to fish for a good instrument among those historical events which can affect individuals' education. Given that we are dealing with people born in Europe between the 1916 and 1946, it comes quite natural to refer to the World War II as the real big thing influencing the lives of our sample.

In their famous paper *The long-run educational cost of World War II*, Ichino and Winter-Ebmer prove that individuals born in Germany and Austria during the '30s had fewer years of schooling due to a poorer supply of education (Ichino & Winter-Ebmer, 2004). As we are looking for an exogenous event able to affect the access to education of our individuals, we decide to use a dummy stating whether parents of our individuals were adolescent during wartime.

Of course, not all the countries in our sample took part in World War II and, even among the fighting ones, the consequences of war were different. Moreover, we have to consider that some of the countries in our sample did not even exist in 1945 i.e. Israel, Croatia, Slovenia and the Czech Republic. For this reason, the sample we are going to exploit for our IV estimation is reduced to only four countries: Germany, Austria, Czech Republic and Estonia as from 1939 until 1945 they happened to be part of the same political entity: the German Reich. Thus, our sample is now made up by 16 093 individuals: 7 044 men and 9 049 women.

For what concerns our analysis, being this countries part of the same state is very useful as they shared identical institutions and laws. In fact, besides the call to the arms and the lack of schools supply due to Ally's bombing, also Nazi laws against confessional schools played a role in limiting the access to education.

As we can notice in Figure 2.1, the trend in people getting secondary education changes during the '30s and then turns to its original slope. In fact, those are the individuals who were between 10 and 15 years old during wartime and whose education was most likely to be conditioned by the war. Our instrument is thus going to be a dummy equal to one if the father's individual was born from 1927 until 1939.

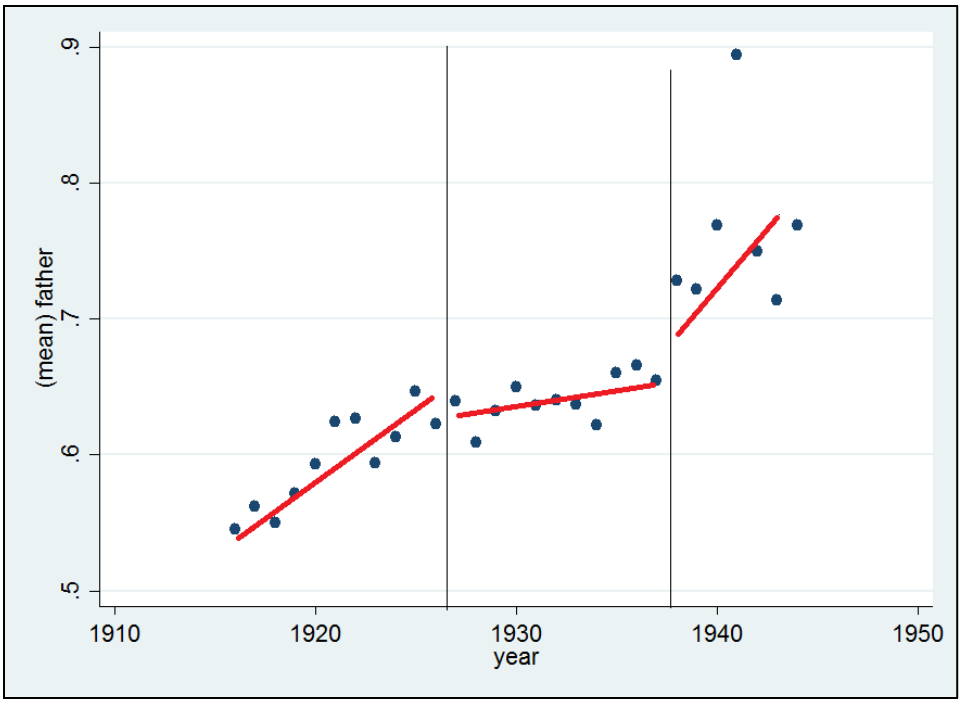

Figure 2.1 - Percentage of fathers with secondary education by year

Model

Consequently to what we said before the first stage of our 2SLS model is:

$$y_i^f = \theta + \gamma war_i + \vartheta_i + \mu year_i + \delta wave_i\, \epsilon_i \quad (5)$$

Where y_i^f is the years of education of the father, war_i is the dummy we chose as instrument, ϑ_i is a second degree polynomials controlling for individual i's age and $year$ is the year of birth of the father. And $\delta wave_i$ is equal to 1 if the individual belongs to wave 5 and 0 if it belongs to wave 6.

The war_i indicator is equal to 1 when the parent of our individual was born between the 1927 and the 1939 and equal to 0 otherwise. ϑ_i is instead constructed in this way: $\vartheta_i = \gamma_1 age_i + \gamma_2 age_i^2$

And the second stage is:

$$y_i^p = \alpha + \beta y_i^f + \vartheta_i + \mu year_i + \delta wave_i + \varepsilon_i \quad (6)$$

Where y_i^p represents the years of education of the partner.

We assume all of the conditions for the IV estimation to be respected. In particular, we have a strong first stage as the value of F is above 14. In the same way in which Ichino and Winter-Ebmer consider their exclusion restriction not to be violated, we think our first stage is the only way through which the World War II can affect the education of the partner.

In fact, it is very unlikely that other consequences of the war, as psychological disorders or malnutrition of children growing during the war, can affect the schooling outcome of people attending classes at least forty years after the end of the war.

In addition, one might argue that a way in which exclusion restriction could be violated is through a possible effect of the instrument on the education of partner's parents. The optimal strategy to get rid of this doubt would be adding a control variable similar to the one stating if father reached secondary education.

Unfortunately, this imply losing about half the observation thus threatening the validity of our estimates. Another strategy could be considering only partners with parents born after the end of the World War II, however also this is not feasible as it will reduce the size of the sample even more than the other method.

What we could is regressing war_i on a dummy stating whether partner's fathers has an ISCED equal or higher than 3 and check if there a significant effect.

For this reason, we are running this regression, pretty similar to the First Stage of our IV model:

$$father_i^p = \theta + \gamma war_i + \vartheta_i + \mu year_i + \epsilon_i \tag{7}$$

Where $father_i^p$ is equal to 1 if partner's father has completed higher education and 0 otherwise.

As we can see in Figure 2.3, the coefficient of war_i is not significant and the value of the F test is very low. For this reasn, we can claim that our instrument has no effect on the dependent variable of our IV model and thus our exclusion restriction hypothesis is valid.

As we did with our previous OLS estimations, we run the model for men and women separately and then jointly but adding a dummy for gender.

Source	SS	df	MS			
				Number of obs	=	5,100
				F(5, 5094)	=	5.63
Model	6.53178339	5	1.30635668	Prob > F	=	0.0000
Residual	1182.01802	5,094	.232041229	R-squared	=	0.0055
				Adj R-squared	=	0.0045
Total	1188.5498	5,099	.233094686	Root MSE	=	.48171

father_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
war	-.0147975	.0234963	-0.63	0.529	-.0608603	.0312653
year	-.0089267	.0052271	-1.71	0.088	-.0191741	.0013206
age	-.0047536	.0195903	-0.24	0.808	-.043159	.0336518
agesq	-.0000698	.0001627	-0.43	0.668	-.0003889	.0002492
wave5	-.0156518	.0140983	-1.11	0.267	-.0432906	.0119869
_cons	18.43172	10.20264	1.81	0.071	-1.569835	38.43327

Figure 2.3 - OLS estimation of exposure to World War II on the education of partner's father

Results

Looking at the estimates for men only, the first thing we must say is that value of F in the first stage equation is much lower than ten. Thus, one of two fundamental assumptions of IV estimation is violated. In Figure 2.2 we report the results, none of the coefficients is significant both in the first and in the second stage. As we expected the coefficient of war_i, even if not significant, is negative; the difference between the coefficient of y_i^f estimated by OLS and IV is quite wide, 0,184 in front of 0,981, and we tried to give an explanation to this fact in the following lines. Nothing changes if instead of years of education we use the dummy $father$ defined in the first chapter: we have no significant coefficients, a value of F very low and a big difference between OLS and IV estimates

Source	SS	df	MS			
				Number of obs	=	5,845
				F(6, 5838)	=	35.36
Model	2947.27679	6	491.212798	Prob > F	=	0.0000
Residual	81094.8163	5,838	13.8908558	R-squared	=	0.0351
				Adj R-squared	=	0.0341
Total	84042.0931	5,844	14.3809194	Root MSE	=	3.727

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.2347723	.1409255	-1.67	0.096	-.5110384	.0414939
agesq	.0014125	.0011587	1.22	0.223	-.0008589	.0036838
year	.0113439	.033743	0.34	0.737	-.0548049	.0774927
wave5	-.4817202	.0982381	-4.90	0.000	-.6743032	-.2891372
yedu_father	.1814023	.0174958	10.37	0.000	.1471041	.2157006
war	-.1078441	.1660824	-0.65	0.516	-.4334271	.2177389
_cons	-2.610603	65.79357	-0.04	0.968	-131.5904	126.3692

Figure 2.4 - OLS Estimation for Men in Germany, Austria, Czech Republic and Estonia

First-stage regressions

Number of obs	=	5,845
F(5, 5839)	=	5.11
Prob > F	=	0.0001
R-squared	=	0.0044
Adj R-squared	=	0.0035
Root MSE	=	2.7878

yedu_father	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.0627848	.105408	-0.60	0.551	-.2694235	.1438539
agesq	-.0000438	.0008667	-0.05	0.960	-.0017428	.0016552
year	-.0461521	.0252323	-1.83	0.067	-.0956167	.0033126
wave5	-.172264	.0734468	-2.35	0.019	-.3162468	-.0282811
war	-.1348868	.1242158	-1.09	0.278	-.3783958	.1086223
_cons	104.4717	49.19411	2.12	0.034	8.033018	200.9104

Instrumental variables (2SLS) regression

Number of obs	=	5,845
Wald chi2(5)	=	77.19
Prob > chi2	=	0.0000
R-squared	=	.
Root MSE	=	4.3402

yedu_p	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
yedu_father	.9809178	1.433681	0.68	0.494	-1.829045	3.790881
age	-.1845749	.199252	-0.93	0.354	-.5751017	.205952
agesq	.0014475	.0013383	1.08	0.279	-.0011756	.0040705
year	.0482432	.0804342	0.60	0.549	-.1094049	.2058913
wave5	-.3439925	.2721605	-1.26	0.206	-.8774172	.1894322
_cons	-86.13733	175.5854	-0.49	0.624	-430.2785	258.0038

Figure 2.2 - IV Estimation for Men in Germany, Austria, Czech Republic and Estonia

In Figure 2.6 we reported the results for a sample made up of only women. We see the effect of being part of the cohort 1926-1939 on years of education is negative. This is exactly what we were expecting: individuals get less education because of a lack of supply. The coefficient, - 0,315, is not that big but it is actually consistent with the magnitude of the effect found by Ichino and Winter-Ebmer.

Looking at the second stage, we notice the effect of father's education on husband's education is 1.461 which is way bigger than the effect we found through OLS for any of the countries included in this regression.

In order to make a fairer comparison, we also ran this regression with a simple OLS and, as we could see in Figure 2.5, the coefficient of the education of the father on husband's year of education is only 0.275.

Usually, having IV estimate bigger than the OLS makes us wonder whether there is any problem with the identification. Actually, it is not that strange IV estimates to be bigger than OLS when we are estimating a Local Average Treatment Effect.

In fact, we can hypothesize that those the war influenced the most were the poorer. The shutdown of a nearby school or rise in the cost opportunity of attending classes due to the war is likely to be more dramatic for poor families than for rich ones. For this heterogeneity of the effect, our LATE is higher than the OLS as it measures the marginal returns of individuals who suffered more than the average because of the constraints generated by the war.

Source	SS	df	MS			
				Number of obs	=	6,211
				F(6, 6204)	=	45.90
Model	4177.72582	6	696.287636	Prob > F	=	0.0000
Residual	94103.048	6,204	15.1681251	R-squared	=	0.0425
				Adj R-squared	=	0.0416
Total	98280.7738	6,210	15.8262116	Root MSE	=	3.8946

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.147658	.139348	1.06	0.289	-.1255124	.4208284
agesq	-.0022072	.0011687	-1.89	0.059	-.0044983	.0000839
year	-.0937026	.032494	-2.88	0.004	-.1574021	-.0300031
wave5	-.2662408	.099318	-2.68	0.007	-.4609384	-.0715431
yedu_father	.2734593	.0181451	15.07	0.000	.2378886	.30903
war	-.3906476	.1683325	-2.32	0.020	-.7206375	-.0606576
_cons	189.4934	63.37462	2.99	0.003	65.25717	313.7296

Figure 2.5 – OLS Estimation for women in Germany, Austria, Czech Republic and Estonia.

Source	SS	df	MS			
				Number of obs	=	6,211
				F(6, 6204)	=	48.70
Model	4420.8732	6	736.812199	Prob > F	=	0.0000
Residual	93859.9006	6,204	15.128933	R-squared	=	0.0450
				Adj R-squared	=	0.0441
Total	98280.7738	6,210	15.8262116	Root MSE	=	3.8896

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.1305631	.1391719	0.94	0.348	-.1422621	.4033884
agesq	-.0020765	.0011673	-1.78	0.075	-.0043648	.0002118
year	-.0969215	.0324533	-2.99	0.003	-.1605412	-.0333019
wave5	-.2730542	.0991867	-2.75	0.006	-.4674944	-.0786139
father	1.599967	.1024724	15.61	0.000	1.399086	1.800849
war	-.3812216	.1681295	-2.27	0.023	-.7108136	-.0516296
_cons	198.3084	63.29485	3.13	0.002	74.22859	322.3882

Figure 2.8 – OLS Estimation of women in Germany, Austria, Estonia and Czech Republic

First-stage regressions

Number of obs	= 6,211
F(5, 6205)	= 8.83
Prob > F	= 0.0000
R-squared	= 0.0071
Adj R-squared	= 0.0063
Root MSE	= 2.7248

yedu_father	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.0017662	.0974923	-0.02	0.986	-.1928849	.1893524
agesq	-.0003316	.0008177	-0.41	0.685	-.0019345	.0012713
year	.0054443	.0227337	0.24	0.811	-.0391217	.0500104
wave5	-.0473332	.0694834	-0.68	0.496	-.1835447	.0888784
war	-.3185462	.1177013	-2.71	0.007	-.5492815	-.0878109
_cons	2.325147	44.33889	0.05	0.958	-84.59443	89.24472

Instrumental variables (2SLS) regression

Number of obs	= 6,211
Wald chi2(5)	= 27.85
Prob > chi2	= 0.0000
R-squared	= .
Root MSE	= 5.129

yedu_p	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
yedu_father	1.499804	.6955085	2.16	0.031	.1366326	2.862976
age	.149824	.1838531	0.81	0.415	-.2105214	.5101695
agesq	-.0018005	.0014827	-1.21	0.225	-.0047065	.0011056
year	-.1003792	.0427418	-2.35	0.019	-.1841516	-.0166068
wave5	-.208194	.1337688	-1.56	0.120	-.4703759	.053988
_cons	186.642	83.58898	2.23	0.026	22.81057	350.4733

Figure 2.6 IV Estimation for women in Germany, Austria, Czech Republic and Estonia

This estimated LATE is not the real effect of parental background on partner's education but it is still interesting as it proves us that real effect is likely to be bigger than the one we found using OLS.

We obtain a similar pattern when instrumenting the indicator $father$ instead of the years of schooling. In figure 2.7, we see the result of IV estimation. The first stage is strong as the value of F is similar to the previous one and the exclusion restriction assumption holds just as well as for the previous model.

As in the other case, we are estimating a Local Average Treatment Effect of a heterogeneous population thus we are not surprised in finding such a big coefficient for $father$. In fact, in both the models the IV estimated coefficient is around five times bigger than the OLS estimated one.

```
First-stage regressions
-----------------------

                                                Number of obs     =      6,211
                                                F(   5,    6205)  =      11.10
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0089
                                                Adj R-squared     =     0.0081
                                                Root MSE          =     0.4819
```

father	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0103827	.0172409	0.60	0.547	-.0234156	.0441809
agesq	-.0001383	.0001446	-0.96	0.339	-.0004218	.0001451
year	.0029424	.0040203	0.73	0.464	-.0049388	.0108236
wave5	-.0038315	.0122877	-0.31	0.755	-.0279197	.0202567
war	-.0603358	.0208148	-2.90	0.004	-.10114	-.0195316
_cons	-5.112105	7.841075	-0.65	0.514	-20.48333	10.25912

```
Instrumental variables (2SLS) regression          Number of obs     =      6,211
                                                  Wald chi2(5)      =      30.06
                                                  Prob > chi2       =     0.0000
                                                  R-squared         =          .
                                                  Root MSE          =     4.9368
```

yedu_p	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
father	7.918295	3.534434	2.24	0.025	.9909309	14.84566
age	.0649621	.1702523	0.38	0.703	-.2687262	.3986505
agesq	-.0012024	.0014062	-0.86	0.393	-.0039586	.0015537
year	-.1155125	.0418977	-2.76	0.006	-.1976305	-.0333946
wave5	-.2488455	.1261542	-1.97	0.049	-.4961032	-.0015878
_cons	230.6084	81.09526	2.84	0.004	71.66458	389.5522

Figure 2.7 - IV Estimation for women in Germany, Austria, Czech Republic and Estonia

Finally, we report the results obtained by running the regression over the whole sample and inserting a dummy equal to 1 if the individual is female and to 0 if it is male. Thus, the model becomes:

First Stage:

$$y_i^f = \theta + \gamma war_i + \vartheta_i + \mu year_i + \tau female_i + \epsilon_i \quad (8)$$

Second Stage:

$$y_i^p = \alpha + \beta y_i^f + \vartheta_i + \mu year_i + \tau female_i + \varepsilon_i \quad (9)$$

As in the previous case, we have a value of F higher than 10 and so our first stage assumption is valid. Coefficients are similar to the previous model when we exploited the only-women sample. However, the coefficient of father's years of education is not significant in the second stage as the value of p is exactly 0,05. In fact, the only significant coefficient in the second stage is the one of the dummy for the gender, being female gets you a more educated partner and this is consistent with the findings of the previous chapter where the coefficients for women were almost always higher than those for men.

The OLS estimation with this sample of both men and women gets us result similar to the previous model too. All of the coefficients are reported in Figure 2.9 and 2.10 and it is interesting to notice that the coefficient of the gender dummy is a little higher than the one for father's years of education. When substituting years of education with the dummy for secondary education as we did before, we obtain the exactly the same changing with the absolute magnitude of the coefficients of interest rising but keeping the same proportion with the IV estimate five times bigger than the OLS one (Appendix).

Source	SS	df	MS			
				Number of obs	=	12,056
				F(7, 12048)	=	67.54
Model	6894.43068	7	984.918669	Prob > F	=	0.0000
Residual	175680.999	12,048	14.5817562	R-squared	=	0.0378
				Adj R-squared	=	0.0372
Total	182575.429	12,055	15.1452036	Root MSE	=	3.8186

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0015106	.0981256	0.02	0.988	-.1908314	.1938526
agesq	-.0007658	.0008155	-0.94	0.348	-.0023643	.0008328
year	-.0447169	.023418	-1.91	0.056	-.0906199	.0011862
wave5	-.3616322	.0699601	-5.17	0.000	-.4987653	-.224499
yedu_father	.2274155	.0126241	18.01	0.000	.2026701	.2521608
war	-.2470398	.1182254	-2.09	0.037	-.4787805	-.0152991
female	.2265504	.0700062	3.24	0.001	.0893271	.3637738
_cons	98.76356	45.67313	2.16	0.031	9.236874	188.2902

Figure 2.9 -- OLS Estimation for both women and men in Germany, Austria, Czech Republic and Estonia.

```
First-stage regressions
```

Number of obs	=	12,056
F(6, 12049)	=	11.58
Prob > F	=	0.0000
R-squared	=	0.0057
Adj R-squared	=	0.0052
Root MSE	=	0.4816

father	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0099344	.0123761	0.80	0.422	-.0143247	.0341936
agesq	-.0001515	.0001029	-1.47	0.141	-.0003531	.0000501
year	-.0017856	.0029535	-0.60	0.545	-.007575	.0040037
wave5	-.0068138	.0088222	-0.77	0.440	-.0241067	.0104792
female	-.0142528	.0088292	-1.61	0.106	-.0315594	.0030538
war	-.0493684	.0149069	-3.31	0.001	-.0785883	-.0201485
_cons	4.091198	5.760098	0.71	0.478	-7.199521	15.38192

```
Instrumental variables (2SLS) regression
```

Number of obs	=	12,056
Wald chi2(6)	=	107.88
Prob > chi2	=	0.0000
R-squared	=	.
Root MSE	=	4.4771

yedu_p	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
father	6.073422	2.806873	2.16	0.030	.5720526	11.57479
age	-.0691048	.1122255	-0.62	0.538	-.2890627	.1508532
agesq	.0001382	.0009396	0.15	0.883	-.0017033	.0019797
year	-.0379699	.0283025	-1.34	0.180	-.0934418	.017502
wave5	-.3447987	.083895	-4.11	0.000	-.5092299	-.1803676
female	.2994269	.091055	3.29	0.001	.1209623	.4778915
_cons	85.12358	55.75485	1.53	0.127	-24.15391	194.4011

Figure 2.10 - IV Estimation for both women and men in Germany, Austria, Czech Republic and Estonia.

Chapter 3 - Pairfam Database

In this third and last section, we try to replicate our model exploiting a different database in order to check the external validity of our results. We expect to get similar results using a different database so that we could claim our findings to be valid for the whole population of European individuals and not only for those similar to ones present in SHARE Database.

Data

The German Family Panel pairfam ("Panel Analysis of Intimate Relationships and Family Dynamics") is a multi-disciplinary, longitudinal study for researching partnership and family dynamics launched in 2008 in Germany. Survey data are annually collected from a nationwide random sample of more than 12,000 persons of the three birth cohorts 1971-73, 1981-83, 1991-93 and their partners, parents and children. Thus, it offers unique opportunities for the analysis of partner and generational relationships as they develop over the course of multiple life phases.

At the present day, there are eight waves of data in pairfam database where individuals are followed in times. For our research, we exploited the eighth and latest wave as we want the highest number of individuals to have completed education and found a partner. In fact, being the last cohort made up of people born between the 1991 and 1993 it is likely that a great part of this boys and girls is still studying and without a partner.

Descriptive Patterns

Our sample is composed of 5 460 individuals, 2 527 men and 2 933 women. All of the people are aged between 21 and 45 and this is a key difference from SHARE Database where all of the interviewed were over 50 years old.

As we can see in Table 3.1, the average years of education are about 13 for both men and women and about 12 for their parents. We got similar educational background for German people in SHARE.

Table 3.1 - Average Years of Education

	Average years of education			
	Individual	Father	Mother	Partner
Men	13,207	12,732	12,339	13,606
Women	13,427	12,643	12,213	13,529

In the same fashion of our previous estimations, we are running regressions using both years of education of individuals' parents and a dummy stating whether parents have at least secondary education. However, both because of the average high-level education of German people and the fact that we are referring to recent cohorts of individuals when compulsory school reforms were already in place, the number of people not having an ISCED equal or higher than 3 is very low (Table 3.2).

Table 3.2 - Percentage of Parents with at least Secondary Education

	Father	Mother
Men	93%	88%
Women	94%	87%

Many things we said in the previous section about educational patterns (e.g. men more educated than women) are not true anymore. This is because we are observing a population who was born on average 25 years later than the other one and the trend in education changed a lot in the second half of the XX century (Schofer & Meyer 2004).

Given that, we can try to create also a similar dummy stating whether parents got tertiary education instead of secondary (Table 3.3).

Table 3.3 - Percentage of Parents with Tertiary Education

	Father	Mother
Men	33%	22%
Women	29%	21%

Results

As we are simply applying the same OLS model to another sample, we skip the paragraph on methodology and go straight to the presentation of the results.

In Table 3.4, we see the coefficients when we use years of education for measuring both dependent and independent variables. Here numbers are not that different from those we could see for Germany in Table.

Using Pairfam instead SHARE we get slightly smaller coefficients but it is interesting to notice that the pattern with father's education counting the most for girl and mother's education for boys is exactly replicated.

These small differences can be the consequence of a rise in social mobility from one generation to the other or simply due to the composition of the samples. We have no way to verify that.

When running the regression with the dummy for secondary or tertiary education we get less encouraging results. In fact, when using the old dummy for ISCED equal or higher than 3 we get no significant coefficient for men's fathers and mother's education counting more than father's for girls. This is in open contrast with our previous findings.

However, this contrast is partly solved when we use a dummy equal to 1 if parents have an ISCED of at least 4. Results reported in Table 3.4, in fact, get close to those obtained with SHARE. Actually, they are even smaller as it happened for the OLS estimates. Moreover, the pattern "father matters for girls, mothers for boys" holds.

This is consistent with our hypothesis that, given the positive trend in education during the XX century, the premium once given by going to high school is now granted only by going to college.

Table 3.4 - OLS Estimation using pairfam Database

	Men				Women			
	Father		Mother		Father		Mother	
Years of Education	0,141	***	0,229	***	0,245	***	0,123	***
ISCED 3	0,208		1,703	***	0,857	***	1,016	***
ISCED 4	0,905	***	1,019	***	1,082	***	0,784	***

IV Estimation

At the same way we did for SHARE, we try to get rid of the endogeneity of parental education. In this case, besides using the exposure to World War II as an instrument we also a more classic instrumental variable as compulsory schooling reforms.

We have been able to adopt this second technique mainly because the cohorts of parents of individuals in pairfam better cover the century than those in SHARE so that we both have fathers being adolescent in the '30s and in the ‘60s.

World War II

When using the exposure to World War II as an instrument we did exactly the same steps described in the previous chapter. Unluckily, in this case, results are much less significant than when using SHARE.

When exploiting the only women sample or the one containing both the genders, we get a solid first stage but non-significant coefficients for the instrumented variable. Moreover, the coefficient of father’s education is negative: if it were significant, it would mean that better fathers lead to worse partner. (Figure 3.1)

On the other hand, if we use the sample with only men we get a positive but non-significant coefficient for father’s education and very weak first stage. (Figure 3.2 and Figure 3.3)

First-stage regressions

Number of obs =	3,205
F(5, 3199) =	16.15
Prob > F =	0.0000
R-squared =	0.0246
Adj R-squared =	0.0231
Root MSE =	2.7970

fyeduc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.2345459	.068602	-3.42	0.001	-.3690543	-.1000375
age2	.0023109	.001023	2.26	0.024	.0003052	.0043167
year	-.0562402	.009703	-5.80	0.000	-.075265	-.0372154
female	-.1401853	.0997542	-1.41	0.160	-.3357739	.0554033
war	-.6823734	.1936865	-3.52	0.000	-1.062136	-.3026112
_cons	127.7499	19.18844	6.66	0.000	90.12702	165.3728

Instrumental variables (2SLS) regression

Number of obs =	3,205
Wald chi2(5) =	115.97
Prob > chi2 =	0.0000
R-squared =	.
Root MSE =	3.0205

pyeduc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
fyeduc	-.1457621	.3065264	-0.48	0.634	-.7465428	.4550185
age	.3104173	.0939343	3.30	0.001	.1263094	.4945251
age2	-.0048902	.0012226	-4.00	0.000	-.0072864	-.0024941
year	-.0662073	.0145002	-4.57	0.000	-.0946271	-.0377875
female	.0129507	.1152631	0.11	0.911	-.2129608	.2388622
_cons	140.1563	32.71479	4.28	0.000	76.03654	204.2762

Figure 3.1 - IV estimation for both men and women (pairfam)

First-stage regressions

Number of obs	= 1,817
F(4, 1812)	= 12.28
Prob > F	= 0.0000
R-squared	= 0.0264
Adj R-squared	= 0.0242
Root MSE	= 2.7662

fyeduc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.2533077	.0893926	-2.83	0.005	-.428631	-.0779844
age2	.0025428	.0013372	1.90	0.057	-.0000798	.0051655
year	-.0581875	.0129054	-4.51	0.000	-.0834986	-.0328764
war	-.6708932	.2596727	-2.58	0.010	-1.180183	-.1616038
_cons	131.7637	25.47974	5.17	0.000	81.79093	181.7364

Instrumental variables (2SLS) regression

Number of obs	= 1,817
Wald chi2(4)	= 34.20
Prob > chi2	= 0.0000
R-squared	= .
Root MSE	= 3.5029

pyeduc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
fyeduc	-.4538071	.4901503	-0.93	0.355	-1.414484	.5068699
age	.0452496	.1565027	0.29	0.772	-.2614901	.3519892
age2	-.0016667	.0019987	-0.83	0.404	-.005584	.0022507
year	-.0853773	.0237633	-3.59	0.000	-.1319526	-.0388021
_cons	186.5596	53.75069	3.47	0.001	81.21013	291.909

Figure 3.2 - IV estimation for women (pairfam)

First-stage regressions

Number of obs	= 1,388
F(4, 1383)	= 7.55
Prob > F	= 0.0000
R-squared	= 0.0214
Adj R-squared	= 0.0185
Root MSE	= 2.8406

fyeduc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.208772	.1072124	-1.95	0.052	-.4190885	.0015446
age2	.0019925	.001593	1.25	0.211	-.0011325	.0051176
year	-.0535698	.0147708	-3.63	0.000	-.0825454	-.0245943
war	-.6940685	.2928471	-2.37	0.018	-1.268541	-.119596
_cons	122.0527	29.27267	4.17	0.000	64.62905	179.4763

Instrumental variables (2SLS) regression

Number of obs	= 1,388
Wald chi2(4)	= 117.29
Prob > chi2	= 0.0000
R-squared	= 0.1475
Root MSE	= 2.6655

pyeduc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
fyeduc	.210787	.395923	0.53	0.594	-.5652079	.9867819
age	.6363059	.1153436	5.52	0.000	.4102366	.8623752
age2	-.0088492	.0015452	-5.73	0.000	-.0118777	-.0058207
year	-.0438356	.0180802	-2.42	0.015	-.0792722	-.0083991
_cons	85.68517	40.60036	2.11	0.035	6.10993	165.2604

Figure 3.3 - IV estimation for men (pairfam)

Compulsory Schooling Reform

As pairfam gets us all the information we need to use compulsory schooling reforms as an instrument we wanted to see if using a different instrument for fathers education we could get similar results.

Table 3.5 - Compulsory Schooling Reform Introduction by Federal State

Federal State	First Birth Cohort Affected
Hamburg	1934
Schleswig-Holstein	1941
Bremen	1943
Lower Saxony	1947
Saarland	1949
North Rhine - Westfalia	1953
Hesse	1953
Rhineland - Palatinate	1953
Baden - Wuerttemberg	1953
Bavaria	1953

Compulsory schooling reforms have been introduced in Germany at different times depending on the federal state. In all of the states, the reformed modified the minimum level of education stating that pupils must attend no less than 9 years of education instead of 8. We retrieved information on reforms and their enforcement from Kemptner, Jurges & Reinhold (2010). In Table 3.5, we can see when reforms have been introduced in the different states and which cohorts were affected. In the next lines we are briefly describing the model:

First Stage:

$$y_i^f = \alpha + \beta school_i + \mu year_i + \gamma state_i + \epsilon_i \tag{10}$$

Where y_i^f, y_i^p and $year_i$ are the same as in Chapter 2, $school_i$ is a dummy equal to 1 if the individual is born after the schooling reform. To build the dummy variable we controlled for the year of birth and the state of residence, as we do not have the state where people attended school we assumed they spent the whole life in the same state. $state_i$ is indeed a control for state fixed effect.

Second Stage:

$$y_i^p = \alpha + \theta y_i^f + \mu year_i + \gamma state_i + \epsilon_i \tag{11}$$

In Figure 3.4, we see the estimates for both first and second stages. As we immediately notice, even in this case we have a strong first stage but no significance for the instrumented variable. Actually, this result is not that stunning as something similar has been found by Pischke & Watcher (2006). They found that compulsory schooling reforms had no effects on returns of education for Germans; at the same time, it could be possible that they do not have effects on returns of marriage.

```
First-stage regressions
-----------------------

                                                Number of obs     =      1,090
                                                F(   3,    1086)  =      30.44
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0776
                                                Adj R-squared     =     0.0750
                                                Root MSE          =     2.6975
```

paryeduc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
parage	.0013887	.0168616	0.08	0.934	-.0316966	.0344739
parbula						
2 Hamburg	.2550382	.7167526	0.36	0.722	-1.15135	1.661427
3 Niedersachsen	-.6806041	.4784199	-1.42	0.155	-1.619344	.2581358
4 Bremen	1.742866	1.445991	1.21	0.228	-1.09441	4.580141
5 Nordrhein-Westfalen	-.3884102	.4662683	-0.83	0.405	-1.303307	.5264861
6 Hessen	-.0175762	.5050339	-0.03	0.972	-1.008537	.9733847
7 Rheinland-Pfalz	.3100642	.5557448	0.56	0.577	-.7803999	1.400528
8 Baden-Württemberg	-.0240409	.494174	-0.05	0.961	-.9936928	.945611
9 Bayern (Bavaria)	-.9037161	.4745118	-1.90	0.057	-1.834787	.0273553
10 Saarland	1.601104	1.016069	1.58	0.115	-.3925926	3.5948
school_dummy	.6753997	.3141189	2.15	0.032	.0590459	1.291753
_cons	12.66414	1.319382	9.60	0.000	10.07529	15.25299

```
Instrumental variables (2SLS) regression          Number of obs     =      1,090
                                                  Wald chi2(11)     =     132.35
                                                  Prob > chi2       =     0.0000
                                                  R-squared         =     0.1272
                                                  Root MSE          =     2.7211
```

pyeduc	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
paryeduc	.1711418	.4581613	0.37	0.709	-.7268378	1.069121
parage	.1129029	.0160994	7.01	0.000	.0813486	.1444571
parbula						
2 Hamburg	.7961357	.7137708	1.12	0.265	-.6028292	2.195101
3 Niedersachsen	.8212358	.585044	1.40	0.160	-.3254294	1.967901
4 Bremen	.9042168	1.623116	0.56	0.577	-2.277031	4.085465
5 Nordrhein-Westfalen	.5460961	.5245122	1.04	0.298	-.4819289	1.574121
6 Hessen	.2520336	.4990086	0.51	0.614	-.7260053	1.230073
7 Rheinland-Pfalz	.5303714	.5495115	0.97	0.334	-.5466514	1.607394
8 Baden-Württemberg	.7406313	.4889762	1.51	0.130	-.2177444	1.699007
9 Bayern (Bavaria)	.4963899	.6921488	0.72	0.473	-.8601969	1.852977
10 Saarland	1.253776	1.210545	1.04	0.300	-1.118849	3.626402
_cons	4.500453	6.928457	0.65	0.516	-9.079073	18.07998

Figure 3.4 - IV estimation using compulsory schooling reform as instrument

To conclude, we can say that running the same model exploiting a different database allowed us to raise the consistency of our OLS estimates, especially those made using years of education for both dependent and independent variables. Unluckily, our IV estimates did not find any confirmation despite both of the instruments we chose respected exclusion restriction and first stage assumptions.

According to us, this could be explained by having the two samples a different target population: over 50 for SHARE and under 50 for pairfam. Having the trend in education changed widely during the last century, using historical events or modifications in the school law allows us to identify effects affecting only a very specific number of individuals in time.

Pairfam database has been very useful especially to confirm the validity of our imputation method. In fact, one of the flaws of our estimates with SHARE was the absence of the variables "years of education" and "year of birth" and the consequent imputation we had to do on our own. As pairfam contains this information and gives us similar results, we can be reassured about the validity of our imputations.

Chapter 4 – Combined Dataset

Finally, we decided to put together the two datasets in order to exploit a bigger sample. Thus, combining the observations we worked with in the two previous sections, we now have a very broad sample made up of 18 843 individuals, 8 223 men and 10 620 women.

As we can see from Table 4.1, individuals in this sample are 58 years old on average, their partners have about 12 years of education while their fathers have 11. As we could expect, average years of education are higher than in SHARE and lower than in pairfam for both fathers and partners. This is perfectly consistent with the rising trend in education, in fact, now we have both individuals over and under 50.

Table 4.1 – Average Age, Father's and Partmer's Eduction by gender

	Age	Partner's Education	Father's Education
Men	58,715	12,252	11,710
Women	58,215	12,498	11,638

What we want to do now is to run again the IV estimation using the exposure to World War II as an instrument. The war_i instrument is built exactly in the same way we did before and so the all the other variables are.

Thus, once again we have a First Stage:

$$y_i^f = \theta + \gamma war_i + \vartheta_i + \mu year_i + \epsilon_i \tag{12}$$

And a Second Stage:

$$y_i^p = \alpha + \beta y_i^f + \vartheta_i + \mu year_i + \varepsilon_i. \tag{13}$$

As in the previous sections, we chose to run the model splitting the sample by gender and then exploiting the whole sample but adding a dummy stating whether the individual is male or female.

In Figure 4.1, we see the results when using the only male sample. First, we notice the value of F is extremely high so we can be sure about our First Stage assumption. war_i is significant and negative as we expected. On the other hand, y_i^f is not significant, as always happened for only-men sample in the previous scenarios.

More interesting are results for the only-women sample (Figure 4.2), here in fact we have a strong first stage e significant coefficient for the instrumented variable in the second stage. In addition, just like in Chapter 2, the IV coefficient (1.017) is almost five times bigger than the OLS coefficient (.281). The coefficient is high but still realistic because as we explained before we are considering a subpopulation who is likely to be affected by the war more than the average.

Everything we said is confirmed in Figure 4.3; here we see the results when keeping men and women and adding a gender dummy variable: value of F above 100, negative and significant effect for the instrument and positive effect for the instrumented variable. Moreover, as we expected the coefficient of the gender dummy is positive consistent with the hypothesis that, inside a couple, men are more educated than women.

```
First-stage regressions
-----------------------

                                               Number of obs     =      7,233
                                               F(   5,    7227)  =      76.98
                                               Prob > F          =     0.0000
                                               R-squared         =     0.0506
                                               Adj R-squared     =     0.0499
                                               Root MSE          =     2.7989
```

yedu_father	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.0780194	.0243643	-3.20	0.001	-.1257806	-.0302582
agesq	.000078	.0002228	0.35	0.726	-.0003587	.0005147
year	-.0416886	.0110724	-3.77	0.000	-.0633937	-.0199835
pairfam	.4203352	.2088614	2.01	0.044	.0109058	.8297646
war	-.2045035	.1102602	-1.85	0.064	-.4206458	.0116387
_cons	96.28283	21.95315	4.39	0.000	53.24823	139.3174

```
Instrumental variables (2SLS) regression         Number of obs     =      7,233
                                                 Wald chi2(5)      =     345.75
                                                 Prob > chi2       =     0.0000
                                                 R-squared         =          .
                                                 Root MSE          =     4.1201
```

yedu_p	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
yedu_father	.9463367	.7936664	1.19	0.233	-.6092209	2.501894
age	.2681912	.0823733	3.26	0.001	.1067424	.4296399
agesq	-.0024815	.0003691	-6.72	0.000	-.003205	-.001758
year	-.005806	.0340366	-0.17	0.865	-.0725165	.0609045
pairfam	1.359524	.5633052	2.41	0.016	.2554665	2.463582
_cons	5.466531	77.31519	0.07	0.944	-146.0685	157.0015

Figure 4.1 – IV Estimation for Men (pairfam and SHARE)

First-stage regressions

Number of obs =	8,028
F(5, 8022) =	92.35
Prob > F =	0.0000
R-squared =	0.0544
Adj R-squared =	0.0538
Root MSE =	2.7354

yedu_father	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.0636788	.0211952	-3.00	0.003	-.1052268	-.0221308
agesq	-.000129	.0002009	-0.64	0.521	-.000523	.0002649
year	-.0399074	.0096942	-4.12	0.000	-.0589106	-.0209043
pairfam	.0984479	.1845834	0.53	0.594	-.2633834	.4602792
war	-.2923891	.1030152	-2.84	0.005	-.4943256	-.0904526
_cons	92.76906	19.21139	4.83	0.000	55.10975	130.4284

Instrumental variables (2SLS) regression

Number of obs =	8,028
Wald chi2(5) =	250.66
Prob > chi2 =	0.0000
R-squared =	.
Root MSE =	4.3976

yedu_p	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
yedu_father	1.161654	.5664255	2.05	0.040	.0514802	2.271827
age	.1373691	.0582139	2.36	0.018	.0232719	.2514662
agesq	-.001318	.0002838	-4.64	0.000	-.0018743	-.0007617
year	-.0309771	.0245399	-1.26	0.207	-.0790743	.0171202
pairfam	1.017617	.3326665	3.06	0.002	.3656028	1.669632
_cons	55.41139	55.01978	1.01	0.314	-52.4254	163.2482

Figure 4.2 – IV Estimation for women (pairfam and SHARE)

First-stage regressions

Number of obs	=	15,261	
F(6, 15254)	=	140.64	
Prob > F	=	0.0000	
R-squared	=	0.0524	
Adj R-squared	=	0.0520	
Root MSE	=	2.7654	

yedu_father	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	-.0709191	.0159816	-4.44	0.000	-.1022448	-.0395933
agesq	-.0000274	.000149	-0.18	0.854	-.0003195	.0002647
year	-.0407921	.0072967	-5.59	0.000	-.0550946	-.0264896
pairfam	.2442235	.1382464	1.77	0.077	-.0267561	.515203
female	-.0741579	.0450709	-1.65	0.100	-.1625022	.0141863
war	-.2505129	.0752728	-3.33	0.001	-.3980565	-.1029694
_cons	94.56813	14.46297	6.54	0.000	66.21899	122.9173

Instrumental variables (2SLS) regression

Number of obs	=	15,261
Wald chi2(6)	=	594.29
Prob > chi2	=	0.0000
R-squared	=	.
Root MSE	=	4.264

yedu_p	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
yedu_father	1.054798	.4633075	2.28	0.023	.1467317	1.962864
age	.1974327	.0479205	4.12	0.000	.1035101	.2913552
agesq	-.0018745	.0002222	-8.44	0.000	-.0023099	-.0014391
year	-.0199832	.019942	-1.00	0.316	-.0590688	.0191025
pairfam	1.143936	.2927164	3.91	0.000	.5702229	1.71765
female	.2598401	.0772451	3.36	0.001	.1084425	.4112377
_cons	33.58848	45.03597	0.75	0.456	-54.6804	121.8573

Figure 4.3 – IV Estimation for both Men and Women (pairfam and SHARE)

Source	SS	df	MS			
				Number of obs	=	7,233
				F(5, 7227)	=	129.64
Model	8218.31575	5	1643.66315	Prob > F	=	0.0000
Residual	91625.2424	7,227	12.6781849	R-squared	=	0.0823
				Adj R-squared	=	0.0817
Total	99843.5581	7,232	13.8058017	Root MSE	=	3.5606

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.1977677	.0287711	6.87	0.000	.1413679	.2541676
agesq	-.0022357	.0002239	-9.99	0.000	-.0026746	-.0017968
year	-.0338726	.0138485	-2.45	0.014	-.0610198	-.0067255
pairfam	1.827408	.2232357	8.19	0.000	1.389801	2.265015
yedu_father	.2047012	.0149611	13.68	0.000	.1753731	.2340293
_cons	71.33293	27.48622	2.60	0.009	17.4519	125.214

Figure 4.4 - OLS Estimation for Men (pairfam and SHARE)

Source	SS	df	MS			
				Number of obs	=	8,028
				F(5, 8022)	=	140.44
Model	9511.28967	5	1902.25793	Prob > F	=	0.0000
Residual	108660.451	8,022	13.5453068	R-squared	=	0.0805
				Adj R-squared	=	0.0799
Total	118171.741	8,027	14.7217816	Root MSE	=	3.6804

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0613187	.0264452	2.32	0.020	.0094793	.1131581
agesq	-.0011217	.0002128	-5.27	0.000	-.0015388	-.0007046
year	-.0609051	.0127478	-4.78	0.000	-.0858942	-.0359161
pairfam	1.37608	.2007872	6.85	0.000	.9824847	1.769675
yedu_father	.2810233	.0150147	18.72	0.000	.2515905	.3104561
_cons	126.9297	25.29077	5.02	0.000	77.35319	176.5061

Figure 4.5 - OLS Estimation for Women (pairfam and SHARE)

Source	SS	df	MS			
				Number of obs	=	15,261
				F(6, 15254)	=	223.58
Model	17649.57	6	2941.595	Prob > F	=	0.0000
Residual	200691.907	15,254	13.1566741	R-squared	=	0.0808
				Adj R-squared	=	0.0805
Total	218341.477	15,260	14.3080915	Root MSE	=	3.6272

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.1236807	.0194444	6.36	0.000	.0855673	.161794
agesq	-.0016489	.0001539	-10.71	0.000	-.0019506	-.0013471
year	-.0490792	.0093803	-5.23	0.000	-.0674657	-.0306926
pairfam	1.554259	.1492309	10.42	0.000	1.261748	1.846769
yedu_father	.2437938	.0106161	22.96	0.000	.222985	.2646027
female	.2008009	.0591194	3.40	0.001	.0849198	.316682
_cons	102.5177	18.61292	5.51	0.000	66.03412	139.0012

Figure 4.6 - OLS Estimation for both Men and Women (pairfam and SHARE)

Conclusions

The evidence in this paper suggests that the effect of parental background on partner's characteristics is significant and it varies across countries. Thanks to our OLS regression we found out that the predictability of partner's education on the base of parents' differs substantially among European countries.

In fact, having at least one parent with secondary education is definitely more important in countries of Southern Europe (i.e. Spain, Italy and Greece) than in the North (i.e. Denmark, Germany, Netherlands). Matching this information with those about intergenerational mobility, we can see that the correlation between the effect of parents' education and the intergenerational mobility in a country is strongly negative.

Our hypothesis is that in countries where social mobility and average education are low, élites fight harder to maintain their status and privileges avoiding to get too close to lower social classes.

Another interesting finding is that, in several countries, women's choices seem to be more affected by father's education while men's seem to be influenced by mother's. We think this is because of the different role that education had for boys and girls in the past century. This hypothesis is consistent with the fact that on average parents' education has higher coefficients for women than for men.

Besides the cross-country OLS regressions, we also tried to get rid of the endogeneity in father's education exploiting the exposure to the World War II as an instrument in former Nazi Reich countries. The estimate of our LATE is way bigger than the OLS coefficient thus suggesting that the effect of father's education is even bigger than what reported in Chapter 1.

To be honest, our LATE can be just be taken as an evidence of a bigger effect and not as an absolute value valid for the whole population. In fact, due to the heterogeneity of our sample, we think our LATE captures the marginal effect on a subset of people affected by the war more than the average. For this reason, the real effect of father's education on partner's is probably higher than our OLS but lower than our LATE.

Finally, the simple robustness test we took exploiting a different database has been useful to confirm the accuracy of our imputation method. Unfortunately, because of the substantial differences in the choices of the sample and the changes in educational trends happened in the past century, we could not find any confirmation but only evidence suggesting our quest was developed in the right direction.

Still, some questions remained unsolved:

- Is there a predictivity not only in education and earnings but also in the occupation?
- Which is the precise and absolute magnitude of the effect of parent's background on partner's characteristics? And how does it changes through generation?

We hope this paper to be a starting point for further investigation in this subject exploiting richer samples and more complex econometric techniques.

It is indeed that economic science can limit itself no more to the traditional fields of application but need to be contaminated by other social sciences. It is, in fact, absolutely important to better understand those mechanism pushing people to take decisions in order to design a fairer system of scarce resources distribution.

Appendix

```
Instrumental variables (2SLS) regression          Number of obs   =      5,845
                                                  Wald chi2(4)    =      67.84
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =          .
                                                  Root MSE        =     3.9804
```

yedu_p	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
father	3.664028	4.910464	0.75	0.456	-5.960304	13.28836
age	-.2991502	.16296	-1.84	0.066	-.6185459	.0202456
agesq	.0021957	.001532	1.43	0.152	-.000807	.0051985
year	.0333269	.0541152	0.62	0.538	-.0727369	.1393907
_cons	-44.68224	106.7948	-0.42	0.676	-253.9963	164.6318

```
Instrumented:  father
Instruments:   age agesq year war

. reg yedu_p father age agesq year if gender==1
```

Source	SS	df	MS
Model	2068.42943	4	517.107358
Residual	81973.6636	5,840	14.0365862
Total	84042.0931	5,844	14.3809194

```
Number of obs   =      5,845
F(4, 5840)      =      36.84
Prob > F        =     0.0000
R-squared       =     0.0246
Adj R-squared   =     0.0239
Root MSE        =     3.7465
```

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
father	.8607059	.1018518	8.45	0.000	.6610386	1.060373
age	-.2609393	.1398566	-1.87	0.062	-.53511	.0132313
agesq	.0016616	.001142	1.45	0.146	-.0005772	.0039003
year	.0101702	.0337264	0.30	0.763	-.055946	.0762864
_cons	1.462139	65.7183	0.02	0.982	-127.3701	130.2943

Figure A.1 – IV Estimation for men in Germany, Austria, Czech Republic and Estonia.

Source	SS	df	MS			
				Number of obs	=	5,845
				F(4, 5840)	=	36.84
Model	2068.42943	4	517.107358	Prob > F	=	0.0000
Residual	81973.6636	5,840	14.0365862	R-squared	=	0.0246
				Adj R-squared	=	0.0239
Total	84042.0931	5,844	14.3809194	Root MSE	=	3.7465

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
father	.8607059	.1018518	8.45	0.000	.6610386	1.060373
age	-.2609393	.1398566	-1.87	0.062	-.53511	.0132313
agesq	.0016616	.001142	1.45	0.146	-.0005772	.0039003
year	.0101702	.0337264	0.30	0.763	-.055946	.0762864
_cons	1.462139	65.7183	0.02	0.982	-127.3701	130.2943

Figure A.2 - OLS Estimation for men in Germany, Austria, Czech Republic and Estonia.

First-stage regressions

Number of obs	=	12,056
F(5, 12050)	=	13.78
Prob > F	=	0.0000
R-squared	=	0.0057
Adj R-squared	=	0.0053
Root MSE	=	0.4816

father	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
age	.0099752	.0123758	0.81	0.420	-.0142833	.0342338
agesq	-.0001512	.0001029	-1.47	0.142	-.0003528	.0000504
year	-.0017192	.0029522	-0.58	0.560	-.007506	.0040676
female	-.014128	.0088275	-1.60	0.110	-.0314314	.0031754
war	-.0491741	.0149045	-3.30	0.001	-.0783894	-.0199589
_cons	3.955495	5.757322	0.69	0.492	-7.329781	15.24077

Instrumental variables (2SLS) regression

Number of obs	=	12,056
Wald chi2(5)	=	87.49
Prob > chi2	=	0.0000
R-squared	=	.
Root MSE	=	4.4311

yedu_p	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
father	5.873501	2.788609	2.11	0.035	.4079292	11.33907
age	-.0650464	.1110733	-0.59	0.558	-.2827461	.1526532
agesq	.0001227	.0009301	0.13	0.895	-.0017002	.0019456
year	-.0349508	.0279608	-1.25	0.211	-.0897531	.0198514
female	.3029154	.0900329	3.36	0.001	.1264542	.4793765
_cons	79.04734	55.067	1.44	0.151	-28.88199	186.9767

Instrumented: father
Instruments: age agesq year female war

Figure A.3 – IV Estimation for both women and men in Germany, Austria, Czech Republic and Estonia.

Source	SS	df	MS			
				Number of obs	=	12,056
				F(5, 12050)	=	80.91
Model	5930.12665	5	1186.02533	Prob > F	=	0.0000
Residual	176645.303	12,050	14.6593612	R-squared	=	0.0325
				Adj R-squared	=	0.0321
Total	182575.429	12,055	15.1452036	Root MSE	=	3.8288

yedu_p	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
father	1.239658	.072387	17.13	0.000	1.097768	1.381548
age	-.0606727	.095947	-0.63	0.527	-.2487443	.1273989
agesq	-.0001676	.0007893	-0.21	0.832	-.0017148	.0013796
year	-.0465701	.0233929	-1.99	0.047	-.0924239	-.0007163
female	.2383435	.070182	3.40	0.001	.1007755	.3759116
_cons	105.2795	45.58593	2.31	0.021	15.92375	194.6353

Figure A.4 - OLS Estimation for both women and men in Germany, Austria, Czech Republic and Estonia.

References

Barro, Robert J. and Lee, Jong-Wha 2010. A New Data Set of Educational Attainment in the World, 1950-2010, NBER.

Angrist, Joshua and Pischke, Steve 2008. *Mostly Harmless Econometrics*, Princeton University Press.

Azam, M. 2016. Intergenerational Persistence among Daughters: Evidence from India, he BE Journal of Economic Analysis & Policy

Blanden, Jo 2005. Intergenerational Mobility in Europe and North America, Center for Economic Performance, LSE.

Chevalier, Arnaud 2004. Parental Education and Child's Education: an experiment, IZA Discussion Papers.

Comi, Simona 2009. *Intergenerational mobility in seven European Countries*, Quaderni dell'Istituto di Impresa, Economia e Lavoro.

Daouli, J., Demoussis, M. and Giannakopoulos, N. 2010. *Mothers, fathers and daughters: Intergenerational transmission of education in Greece*, Economics of Education Review.

DeSilva, S. and Bakthiar, M. 2011. *Women, Schooling and Marriage in Rural Philippines*, Levy Economics Insititute.

Dubow, Eric F. 2009. *Long-term Effects of Parents' Education on Children's Educational and Occupational Success: Mediation by Family Interactions, Child Aggression, and Teenage Aspirations*, NCBI.

Dustmann, Christian 2004. *Parental Background, Secondary-Track School Choice and Wages*, Oxford Economics Papers.

Fitzpatrick, Sheila 2002. *The Commissariat of Enlightenment: Soviet Organization of Education and the Arts Under Lunacharsky, October 1917-1921*, Cambridge University Press.

Hajnal, John 1965. *European Marriage Pattern in Perspective*, in Glass, D. V. & Eversley, D. E.C. (eds.), Population in History (pp 101-143).

Havari, Enkelejda and Peracchi, Franco 2014. *Growing up in wartime: Evidence from the era of two world wars*, EIEF Working Papers.

Havari, Enkelejda and Savegnago, Marco 2014. *The causal effect of parents' schooling on children's schooling in Europe. A new IV approach*, CEIS Research Papers.

Hertz, T. et al. 2007. *The Inheritance of Educational Comparisons and Fifty-Year Trends*, The B.E. Journal of Economic Analysis and Policy

Hu, Yang 2016. *Marriage of Matching Doors: Marital Sorting on Parental Background in China*, Demographic Research.

Huang, C., Li, H. et al. 2009. *Why Does Spousal Education Matter for Earnings? Assortative Mating and Cross-Productivity*, Journal of Labour Economics.

Ichino, Andrea & Winter-Ebmer, Rudolf 2004. *The Long-Run Educational Cost of World War II*, Journal of Labour Economics.

Jäntti Markus et al., 2006. *American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States*, IZA Discussion Papers.

Kaufmann K., M. et al. 2013. *Return to Elite Higher Education in the Marriage Market: Evidence from Chile*, SSRN Electronic Journal.

Kaufmann, D 1896. *Die Memoiren der Glückel von Hameln, 1645–1719*

Lam, David and Schoeni, Robert F.1994. *Family Ties and Labour Markets in the United States and Brazil*, The Journal of Human Resources.

Lappegård, Trude 2000. *New Fertility Trends in Norway*, Demographic Research.

Lefgren, Lars and McIntyre, John 2006. *The Relationship Between Women's Education and Marital Outcomes*, Journal of Labor Economics.

Liu, Q. 2011. *Can we have it all? The Impact of Marriage on the Returns to Education in Sweden*, Journal of Labour Economics.

Lund Christer and Kurosu, Satomi 2014. *Similarity in Difference. Marriage in Europe and Asia, 1700-1900*, MIT Press.

Mare, Robert D. 1991. Five Decades of Educational Assortative Mating, Center for Demography and Ecology.

Pischke, J. and Von Watcher, T. 2006. *Zero returns to compulsory schooling in Germany: Evidence and Interpretation*, Centre for the Economics of Education.

Schofer, E. and Meyer, J. 2004. *The World Wide Expansion of Education in the Twentieth Century*, American Sociological Review.

Acknowledgments

I would first like to thank my thesis supervisor Professor Giorgio Brunello of the Departments of Economics and Management at Padua University. The door to Prof. Brunello office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work but steered me in the right the direction whenever he thought I needed it.

I would also like to acknowledge Prof. Marco Bertoni as the second reader of this thesis, and I am gratefully indebted to him for his very valuable comments on this thesis. He does not want me to call him “Professor” but he actually taught me most of the things I know.

My parents played an important role too, always supporting me and putting me in the best conditions to achieve my goals no matter what it could cost to them. Moreover, I would especially say thanks to my brother Marco for all the nights spent together celebrating successes and sharing defeats. You cannot choose your siblings but if I could, I would choose nobody but him.

Getting to the end of this amazing academic path would not have been possible without the company of real friends met along the way: Zeno, Jacopo, Giulia, Michela and Linda. We will not be seeing each other every day anymore but I really hope your life will be filled with the happiness you deserve.

I did not forget to mention Erin but a few lines are not enough to tell what you meant to me: you have been a colleague, a friend, a lover and you will always be a model to look at and admire.

Every important adventure needs the support of a great soundtrack. For this reason, I cannot help thanking *Sparkers* for composing mine day by day with a melody made up of laughs, jokes and Coldplay songs.

Finally, I would like to thank Sara who made this possible in the first place: she was the good reason to stay when I had a million to go. I truly hope she will always be.

Stefano Gallinaro