Siblings and human capital.

A comparison between Italy and France

Giulia Ferrari

Dept. of Demographic Sciences, University of Rome "La Sapienza"

Gianpiero Dalla Zuanna

Dept. of Statistical Sciences, University of Padova

Working Paper

Abstract

Does growing up in a large family affect people's educational chances? Previous studies on the argument proved the dilution effect theory validity: the higher the number of siblings, the less the parental resource availability for each child and –extending to educational opportunities- the less the probability that each child will carry on successfully his/her educational career. The paper investigate the dilution effect in two European countries – Italy and France – for which family policies are very different but comparable data are available. We assume (1) a stronger impact of dilution effect in Italy than in France, (2) an absence of the negative effect among wealthier families and (3) birth order differentials in terms of educational chances. For this purpose, we analyze the probability of achieving high levels of education by means of a simple logistic model. We find a negative correlation between family size and human capital and a higher probability for firstborns and only children in both countries, but when we include the interaction with the parents' occupations, it consistently weakens among French upper classes. However, when we use the siblings' sex composition as an instrument people belonging to a larger family have higher probabilities to achieve secondary school and university degrees.

Key Words: dilution effect, education, family size, siblings

Table of contents

1. Introduction

Investigations on the determinants of educational attainment, occupation and income usually consider the familiar background as a general effect. Less attention, on the other hand, has been paid to investigating family size as an independent effect. In this research project, the attention is focused on a set of fundamental questions relative to social inequalities that may be linked with family background characteristics, especially with family size. Based on the *dilution effect theory*, the assumptions are that the higher the number of siblings, the less the possibility for an individual to achieve the top levels of education. Does growing up in large families affect people educational chances? If so, does the negative effect of a large family size on educational attainment decrease among the more advantaged social classes? Has the negative effect of large family size on educational attainment changed over the course of the 20th century? Are there any significant educational and life chances differences between first and lastborns?

Large families often mean a dilution of resources available for each child. This research aims therefore at investigating, in light of national family policies, whether the "dilution effect" differs between countries. We approach this issue by concentrating our analysis on two European countries – Italy and France – for which family policies are very different but comparable data are available.

Under another point of view, some unobserved factors could play a role in both influencing family size and children outcomes. Criticisms from the literature (e.g. Angrist & Evans, 1998) argue that one possible source of unobserved heterogeneity might be linked to the parent's preference for the *quality* of each child. In fact, couples who have a strong proneness for their children's outcomes would try to keep the family size small in order to allow each child to reach a high level of education. Hence, the endogeneity issue should be taken into account when the *true* causal effect of family size on education has to be evaluated. In this research, an instrumental variable approach will be applied, using the sibling sex composition as an exogenous determinant of family size variation (Conley, Glauber, 2004).

The paper is organized as follows. The next section summarizes briefly the current literature debate on the subject of interest and discusses the main hypothesis of the project. Section 3 describes the datasets and the methods used for the purposes of the analysis. Afterwards, section 4 illustrates the set of variables included in the analysis together with some descriptive statistics on the relationship

between the variables. In section 5 the main results are presented. Section 6 introduces the issue of assessment causality and in section 7 results of the whole analysis are interpreted. Finally, section 8 concludes and suggests directions for further work.

2. Background and hypotheses

Scholars' interest on the topic under analysis has increased in the last decades and one common observed pattern in the literature is that as the number of sibling increases the children educational achievement is weakened.

Many authors have detected the relationship between quality and quantity of children, but Judith Blake primarily assessed the question we currently deal with in her book "*Family Size and Achievement"*. She was the first scholar who coined the term *dilution effect*, referring to the increasing disadvantages an individual encounter in growing up in large family size. The basic idea is that since the amount of the family's material and non-material resources is limited, children reared in larger families are more likely to suffer from its reduction. Further the differential allocation among each child might affect educational outcomes as well as intellectual development. The total amount of resources depends in turn on the number of children and in how they are spread out in age. In her book she presents results for the US, using a number of longitudinal and cross-sectional surveys, on the relationship between educational attainment and number of siblings as well as ethnicity, religion and social status. Moreover, she detects the association between sibship size and intelligence as well as the effect of birth order on educational outcomes and intellectual ability. She concludes that the reduction of family sizes that US is experiencing in the last decades can only improve opportunities for individuals to proceed in their education. In fact, evidence shows that even after controlling for major parental background characteristics, the larger the sibship size, the lower the probability for children to achieve high levels of education, the higher the tendency of dropping school and the lower the children's IQ.

Besides family size, one should take into account another crucial characteristic of the siblings group, namely the birth order. Evidence from literature shows no general impact of birth order on educational attainment arguing that firstborns and lastborns can both benefit from their rank position. On the one hand, some authors conclude that firstborn children are advantaged with respect to their later born siblings because of their priority in obtaining parental time, energy, and interest on children's lives. On the other hand, these benefits do not generally extend to economic resources. In fact, later born children are more likely than the firstborn to have older parents, who in turn are more likely to be in a better financial position to support their children. In addition, they could exploit their older siblings' knowledge for being helped in homework, consequently increasing their probability of succeeding at school. Recognizing their disadvantageous position in the family hierarchy, later born children also develop alternative *strategies of survival that often entail risk-taking and daring behaviour* (Black, Devereux, Salvanes, 2005). Other scholars (Gary-Bobo, Prieto, Picard, 2006) on the contrary, found a significant and negative impact of child rank on educational achievement. These negative impacts remain also after controlling for father's occupational status.

Nowadays, families may vary more in the age spacing than in the number of siblings. Scholars (Blake J., 1980) generally agree in saying that longer intervals allow parents to recover economic losses before the next child requires further investments. Relative to sibling's sex-composition, evidence shows a wide and inconsistent set of findings. Different effects for boys and girls might exist if one analyzes sex differences according to father's occupation. As a matter of fact, for some professions and in particular the self employed ones, fathers would rather be more orientated in encouraging their children towards the conduction of their own activity (Zarca 1995a,1995b)

Nonetheless, in the last decades there has been an increasing concern of scholars towards the question of whether the association between family size and children's outcomes could represent a real

causal influence. Criticism mainly derives from the fact that the observed relationship could be spurious because of the simultaneous determination of outcome and "treatment". Specifically, a considerable source of unobserved heterogeneity might arise from the parent's preference for each child's *quality*. In fact, couples who are strongly concerned with their children's education would try to keep the family size small in order to allow each child to reach a high level of education.

In light of the common research direction, the key hypothesis which has been developed concerns the negative impact of large family sizes on children's educational attainment. As the number of siblings increases, the individual's opportunities of achieving high levels of education decrease. The so called dilution effect is a concept very close to the one of children's quantity-quality trade off. The so called dilution effect is assumed to be *more intense in Italy than in France* because families with many children are less frequent. Furthermore, in France the instrument of *"quotient familial"* (i.e. a coefficient which operates in reducing household's income taxation accounting for the number of children) works as a financial incentive for increasing births and at the same time for indirectly decreasing the educational expenses burden.

The second assumption is based on the observed strong impact of socio-economic background characteristics on people's chances to reach the top educational degrees. As a matter of fact, children's education involves a large amount of direct and indirect expenses which rise more than proportionally as advanced educational levels are considered (it implies indeed a later entering in the labour market). Hence, children reared in favourable socio-economic conditions are assumed to be more likely to reach university and secondary school degrees whereas, on the contrary, those who have grown up in less advantaged families might face much more difficulties in achieving the desired educational goals.

Further, over the last century Western countries have experienced a progressive expansion of participation into the school system. Consequently, a growing number of people have reached a secondary school or a university degree. Therefore, the expected negative impact of family size is assumed to weaken for the youngest cohorts.

Finally, the last hypothesis concerns the differential advantages and drawbacks of birth order. As previously said, evidence from literature does not seem to support preferences for any specific position. However, in this research project the birth ranking is supposed to act negatively. In other words, following Zarca's findings on intergenerational mobility, it is assumed that firstborns are more likely to achieve top education then second-borns, which in turn have more chances than the third-borns, and so forth (Zarca, 1995a, 1995b).

3. Data and Methods

Data

Data for Italy come from the *Indagine Multiscopo sulle famiglie: Famiglia e soggetti sociali (2003)* (The Multipurpose Survey on the Family), a representative national survey conducted by the Italian National Statistics Institute (Istat) on a sample of 24,000 Italian families and 50,000 individuals. Data for France have been borrowed from the *Etude de l'Histoire Familiale (1999)* (the Family History Survey), a representative national survey conducted by the French National Institute of Statistics and Economic Studies (Insee) and the French National Institute for Demographic Studies (INED) on a sample of 380,000 respondents. Despite differences between the two surveys, a weighty correspondence has been maintained, allowing for strong comparability in terms of topics and observed time period.

Additionally, both surveys include detailed information on *individual education*, i.e. the response variable, as well as on the *number* and *birth order* of siblings, i.e. the main covariates. Data allowed also creating homogeneous categories concerning the *socio-economic background* of the family, based on parents' occupation and mother's activity status.

Methods

Several logistic regression models have been employed in order to analyze the impact of the number of siblings and birth order on the probability of achieving at least a secondary school (or university) degree, controlling for other covariates.
Henceforth, results displayed in the next pages are obtained by means of a logistic model for the dummy response variable which juxtaposes high and low educational levels: the "success" is to have completed at least the secondary school. The logistic regression's function is specified as follows:

$$Logit[\pi(x)] = \beta_0 + \sum_{i=1}^{q} x_i \beta_i$$

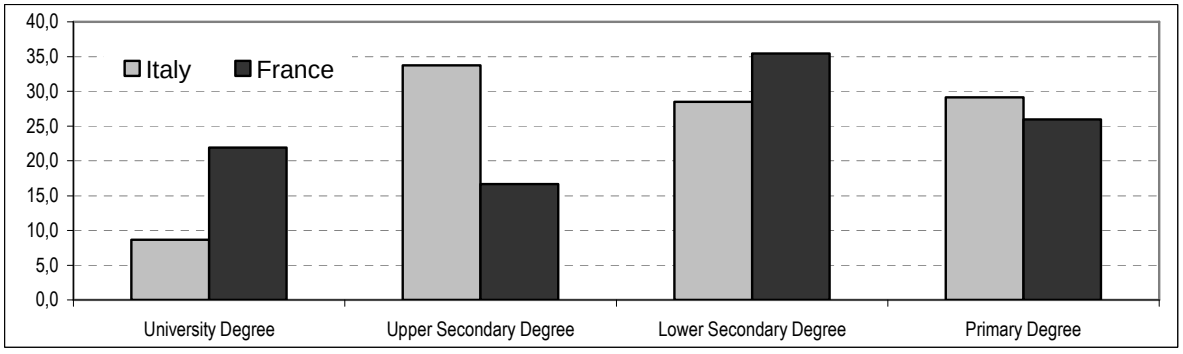
where $Logit[\pi(x)] = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right]$ denotes the natural logarithm of the ratio between the probability of success and the probability of failure, given the set of *q* covariates X. The probability that Y takes the value 1 can be written as a logistic function: $\pi(x) = \frac{e^{X\beta}}{1+e^{X\beta}}$ and hence the ratio $\left(\frac{\pi(x)}{1-\pi(x)}\right)$ can be interpreted as a measure of the risk of success. The latter is commonly called *odds ratio* and it is equal to the exponential of β in the logistic regression.

4. Variables and sample description

Dependent variable

Since the objective of the study is to model people's chance to obtain high levels of education according to their family dimension by comparing two countries, it is crucial to handle a standard outcome variable, which can be easily applied to both the contexts. Hence, the educational level has been divided into four categories based on the last degree achieved: (1) university degree, (2) upper secondary school degree, (3) lower secondary school degree and (4) primary degree. Despite the different educational system characterizing France and Italy, the latter classification keeps strong links with national-specific levels. Additionally, when the analysis concerns the probability of reaching a university degree, only individuals aged over 25 years are considered whereas, when modeling the probability of achieving an upper secondary degree, the whole over 18 years sample is inspected. The following graph depicts the frequency distribution of school degrees comparing the two countries.

Figure 1: Educational degrees' frequency distribution, Italy and France



The main difference is found in the highest levels of education: in France the proportion of those who achieved a university degree is double the Italian one, while in Italy upper secondary degrees' frequency is twice the French one.

Covariates

The principal explanatory variable related to the sibling's characteristics is the *family size*, meant as the total amount of brothers and sisters. As it can be seen in the table 1, the distribution of large and small families has changed considerably over the last century towards a prevalence of the smaller one in the youngest cohort. The latter trend implies the mandatory inclusion of age classes when investigating the effect of number of siblings on educational achievement.

Table 1: Frequency distribution of family sizes according to age, Italy and France (%)

Ages (cohorts)	**Mean (all ages)**		**18-24** *(ITA:1979-1985) (FRA: 1975-1981)*		**25-44** *(ITA:1959-1978) (FRA: 1955-1974)*		**45-64** *(ITA:1939-1958) (FRA: 1935-1954)*		**65 and more** *(ITA:1901-1938) (FRA: 1900-1934)*	
N. of siblings	**Italy**	**France**	**Italy**	**France**	**Italy**	**France**	**Italy**	**France**	**Italy**	**France**
Only Child	11.7	11.1	13.3	10.8	11.9	9.0	12.0	10.7	10.3	15.9
1 sibling	29.6	22.8	51.2	35.9	36.9	24.1	24.4	18.6	15.6	20.5
2 siblings	21.8	21.6	25.1	28.4	24.9	23.5	20.6	19.4	17.1	18.2
3 siblings	13.3	14.3	7.3	11.7	12.1	14.8	15.2	15.0	15.2	13.4
4 siblings	8.0	9.5	1.8	5.6	6.0	9.3	9.1	10.9	12.5	9.7
5 and more	15.6	20.7	1.4	7.7	8.3	19.4	18.7	25.4	29.3	22.3
Total	100	100	100	100	100	100	100	100	100	100
N	41 559	366 231	4 378	34 898	15 280	140 798	12 529	113 393	9 372	77 142

Further, a measure of the socio-economic background of the family has been included in the model as a necessary control covariate, since children's educational attainment is known to be strongly affected by parent's financial constraints. In this analysis, *parents' occupation* has been used as a proxy of socio-economic conditions. The variable has been created combining father's and mother's jobs, considering the most remunerable between the two as indicator of the general family situation. Specifically, occupations have been grouped into four categories, distinguishing (1) upper class, (2) white collars, (3) self employers, (4) blue collars and unemployed. The rational decision for analyzing separately the self employers' category relies on the common strength of the intergenerational transmission of that particular profession: there are reasons to believe that the sons of self employers are likely to keep parent's activity (Zarca B., 1995a). This in turn could indirectly affect children's educational process in the direction of driving them to leave school earlier for taking part in the activity or, on the contrary, encouraging their permanence into the school system in order to acquire the knowledge required for a specific occupation.

Table 2 shows the distribution of educational levels of individuals for each parent's socio-economic category. In the two countries about 2/3 of people belonging to upper classes achieve at least a secondary school degree; in Italy the same pattern holds for those whose parent's were white collars, while in France only about the half of whose belong to this category achieve at least a secondary school degree. Concerning self-employers and blue collars categories in Italy, only about 1/3 of people achieve the two top levels of education. In France 40% of individuals whose parents were self-employers reach high levels of education, while only 20% of sons of blue collars or unemployed people succeed in that goal.

Table 2: Frequency distribution of levels of education according to socio-economic origin, Italy and France (%)

Parents' job	Upper Class		White Collars		Self-Employers		Blue Collars	
Education	Italy	France	Italy	France	Italy	France	Italy	France
University Degree	27.65	52.66	21.12	27.42	6.18	19.64	3.79	8.83
Upper Secondary Degree	44.70	21.06	52.26	20.07	27.43	19.21	24.99	11.03
Lower Secondary Degree	16.84	20.75	18.92	37.97	26.45	36.94	32.54	39.06
Primary Degree	10.81	5.52	7.70	14.54	39.94	24.2	38.68	41.08

Moreover, models include a variable expressing the *mother's activity status*, for better characterizing the socio-economic background as well as the presence of parents at home. Indeed, in a study on the effects of number of siblings, birth order and social origin on children's educational outcomes Gary-Bobo et al. found that having a retired father increases the children's educational success (Gary-Bobo et al., 2006). This is probably linked to the additional support available for children at home, which in the present analysis is considered as the mother's one.

Besides the family size, the *birth order* is another important variable related to the sibship composition and strictly connected to the number of siblings. As it has been pointed out in the literature review, evidence shows its controversial role on children outcomes and it is thus worthy to be taken into account. In this article, it will be included in model 4 in interaction with family size in order to investigate how children's opportunities change according to different sub-groups.

As far as demographic characteristics are concerned, *age* and *gender* were also included. The age has been centered to its median value to obtain more reliable estimates. The units of analysis are individuals over the age of 18, when the outcome represents the secondary school achievement, and people aged over 25 when the university degree achievement is detected.

The last tables of this section aim in completing the data description. The frequency distribution of the main variables has been summarized for particular sub-populations. Specifically, educational degrees, parent's occupation and mother's activity categories are classified for each age class, gender and family size. The latter was divided into (1) 2 siblings and less and (2) 3 siblings and more in order to discriminate between small and large sizes.

Table 3: Some socio-economic characteristics according to age classes, sex and family size. Italy %

Italy %	25-44 years old				45-64 years old				65 and more years old			
	Males		Females		Males		Females		Males		Females	
	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more
Education												
University Degree	13.1	6.3	16.6	7.6	12.8	5.6	11.7	4.3	7.5	3.9	3.3	1.3
Upper Secondary Degree	50.2	31.3	53.6	33.0	36.1	19.8	32.1	17.0	18.3	9.1	12.9	5.6
Lower Secondary Degree	34.1	51.4	27.3	48.1	31.9	33.7	28.0	26.7	19.7	12.1	14.2	7.6
Primary Degree	2.6	11.1	2.5	11.4	19.2	40.9	28.1	52.0	54.5	74.9	69.7	85.6
Parents' occupation												
Upper Class	15.2	9.6	13.9	8.7	9.8	7.1	8.7	6.4	7.2	5.9	7.6	5.1
White-Collars	21.0	11.8	23.4	13.2	13.6	7.9	15.8	8.3	9.7	4.6	8.9	5.6
Self-Employers	20.2	22.8	20.7	21.3	24.1	28.0	23.8	25.3	28.6	35.2	28.4	31.3
Blue-Collars	43.6	55.8	42.1	56.8	52.6	75.1	51.8	60.0	54.5	54.3	55.1	58.0
Mother's activity												
Non-Active	59.0	74.9	57.8	64.1	67.5	75.2	64.1	74.1	66.3	69.8	67.8	69.2
Active	41.0	25.1	42.2	35.9	32.5	24.8	35.9	25.9	33.7	30.3	32.2	30.8

Table 4: Some socio-economic characteristics according to age classes. sex and family size. France %

France %	25-44 years old				45-64 years old				65 and more years old			
	Males		Females		Males		Females		Males		Females	
	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more	2 siblings or less	3 siblings or more
Education												
University Degree	36.7	19.3	42.5	19.6	23.5	13.8	26.2	10.4	12.4	6.5	4.9	2.5
Upper Secondary Degree	18.6	13.9	21.9	18.6	18.0	11.4	34.1	13.6	11.7	6.8	12.0	6.7
Lower Secondary Degree	41.1	57.8	32.2	51.2	35.1	36.6	21.3	32.2	24.7	18.6	26.6	17.4
Primary Degree	3.5	9.1	3.3	10.6	23.5	38.1	18.4	44.0	51.3	68.1	56.5	73.4
Parents' occupation												
Upper Class	18.1	9.0	18.7	9.0	11.7	7.8	11.8	7.5	7.9	4.7	7.0	4.4
White-Collars	41.8	27.7	42.0	28.4	29.0	21.4	29.4	21.7	22.6	14.3	21.1	15.2
Self-Employers	9.0	8.0	9.6	8.0	12.6	9.4	12.9	9.2	14.9	12.2	15.5	11.6
Blue-Collars	31.2	55.4	29.7	54.6	46.7	61.3	45.9	61.5	54.7	68.9	55.4	68.8
Mother's activity												
Non-Active	36.2	60.3	35.2	59.9	52.7	65.7	51.1	64.8	55.7	61.7	55.2	60.9
Active	63.8	39.7	64.8	40.2	47.3	34.3	48.9	35.2	44.3	38.3	44.8	39.1

From the descriptive analysis, the dilution effect seems to work in the two countries: proportions of individuals who achieve secondary school and university degrees are always higher among those with 2 siblings or less than among those with more than 3 siblings. Further, females appear to advance more than males in education, at least concerning the youngest cohorts. As far as socio-economic background is concerned, larger families sizes are more common among white collars and unemployed parents and inactive mothers, which are supposed to be the less advantaged socio-economic categories.

5. Results

This section describes the results of the logistic regression models transformed in odds ratios, referring to the next section for discussion. Firstly, relative risks of achieving university and secondary school degrees will be presented without including any interaction, comparing France and Italy (model 1). Afterwards, the probability of reaching the two considered levels will be analyzed including interactions with parent's occupation categories (model 2) and age classes (model 3). Lastly, an interaction between the family size and birth order will be included into the basic model (model 4).

Model 1

Table 5 displays the probabilities of achieving university and secondary school degrees according to different family sizes, controlling for parents' occupation, mother's activity, sex, and year of birth (model 1). In both Italy and France, there are no significant differences between only children and people with one sibling. On the contrary, people who have two siblings or more are clearly penalized. Despite expectations, there is a strong negative effect of family size in France, although it is weaker than in Italy. People who have experienced a disadvantaged socio-economic background show lower probabilities to achieve high levels of education in France and in Italy. Further, the mother's activity

seems to work negatively towards children's education, following the initial assumptions. Finally, sex differences, where they exist, are not much important in determining people opportunities.

Table 5: Effects of family size and other covariates on the probability of achieving university and secondary school education

	University degree				Secondary degree			
	Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)	
Number of siblings								
0	1.00	n.s.	0.95	**	0.99	n.s.	0.97	*
1 (Ref)	1		1		1		1	
2	0.75	***	0.82	***	0.67	***	0.78	***
3	0.62	***	0.64	***	0.50	***	0.61	***
4	0.42	***	0.53	***	0.33	***	0.50	***
5 and more	0.29	***	0.38	***	0.22	***	0.36	***
Paerents' Occupation								
Upper Class	9.34	***	9.00	***	5.92	***	9.17	***
White-collars	5.59	***	2.87	***	4.63	***	2.78	***
Self-Employers	2.00	***	2.55	***	1.52	***	2.55	***
Blue-Collars(Ref)	1		1		1		1	
Mother Activity								
Active	0.90	**	0.89	***	0.84	n.s.	0.83	***
Sex								
Males	1.00	n.s.	0.91	***	0.99	***	1.13	***
Age	0.99	***	0.96	***	0.96	***	0.96	***

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

Model 2

The next two tables (table 6 and 7) show estimates obtained from model 2**.** The probability of achieving university and secondary school degrees lowers significantly as the family enlarges, but the decrease is slower in France for almost all the socio-economic groups. Nonetheless, an interesting result could be found among people belonging to the large families at the top socio-economic group: in France, those who have 5 siblings or more have "only" 40% and half the probability of achieving high education (respectively for university and high school degree) than the reference group, while in all the other categories and also in Italy the same chance to reach top degrees is about 80% less likely. Moreover, there are no significant differences between only children and people with one sibling and this holds true for both countries, with the exceptions (1) of French only children whose parents' occupation belongs to the blue collar category, which seem to be less likely to reach top levels than the reference group and (2) of French persons belonging to self employers group, which turn out to be more advantaged than the reference.

Table 6: Effect of family size on educational outcomes by socio-economic origin, Italy and France, university degrees

	University degree															
	Upper Class				White Collars				Self-Employers				Blue Collars			
	Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)	
Number of siblings																
0	1.01	n.s.	0.92	*	1.17	n.s.	0.97	n.s	0.97	n.s.	0.97	n.s.	0.87	n.s.	0.83	***
1 (Ref)	1		1		1		1		1		1		1		1	
2	0.80	*	0.97	n.s.	0.82	*	0.75	***	0.77	**	0.87	***	0.70	***	0.80	***
3	0.70	**	0.89	***	0.83	n.s.	0.55	***	0.62	***	0.73	***	0.44	***	0.62	***
4	0.50	***	0.83	***	0.50	***	0.44	***	0.36	***	0.57	***	0.39	***	0.50	***
5 and more	0.22	***	0.59	***	0.30	***	0.33	***	0.27	***	0.47	***	0.32	***	0.36	***
Sex																
Males	1.00	***	0.92	***	1.10	***	1.07	***	1.05	***	0.79	***	0.86	***	1.00	***
Age	1.01	n.s.	0.98	***	1.01	n.s.	0.97	***	0.99	n.s.	0.97	***	0.99	n.s.	0.95	n.s.

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

Table 7: Effect of family size on educational outcomes by socio-economic origin, Italy and France, secondary school degree

	Secondary School Degree															
	Upper Class				White Collars				Self-Employers				Blue Collars			
	Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)	
Number of siblings																
0	0.95	n.s.	0.96	n.s.	1.03	n.s.	1.02	n.s	1.01	n.s.	1.10	**	0.98	n.s.	0.89	***
1 (Ref)	1		1		1		1		1		1		1		1	
2	0.73	***	0.86	***	0.76	***	0.74	***	0.70	***	0.89	***	0.63	***	0.78	***
3	0.55	***	0.77	***	0.60	***	0.54	***	0.53	***	0.74	***	0.46	***	0.62	***
4	0.31	***	0.74	***	0.36	***	0.45	***	0.38	***	0.54	***	0.32	***	0.50	***
5 and more	0.19	***	0.50	***	0.18	***	0.32	***	0.25	***	0.44	***	0.23	***	0.37	***
Sex																
Males	1.02	***	1.01	***	1.23	***	1.18	***	1.09	***	0.99	***	0.87	***	1.17	***
Age	0.98	n.s.	0.98	n.s.	0.97	***	0.97	***	0.95	*	0.97	n.s.	0.96	***	0.96	***

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

Model 3

The relationship between family size and education found before could be only indicative; for example, it could simply embody cohort effects, as family sizes have declined over time as educational attainment has increased (Black et al. 2005). The inclusion of an interaction with age group in modelling education of children on family size aims in better controlling the different cohort effects, which are anyhow taken into account in the previous models thought the *age* variable.

In Tables 8 and 9, the dilution effect by individual's year of birth group is presented. In both countries, the impact of large family size is negative throughout the 20th century, despite tremendous changes in the educational systems and the diffusion of secondary education. This remains true even when controlling for father's occupation, mother's activity status, sex, and year of birth and for every

level of education which has been analyzed. As previously noted, however, the dilution effect is stronger in Italy than in France; this holds true for each cohort group considered.

Such stability is noticeably evident in light of other fundamental changes. For example, cohort after cohort, the relative impact of father's occupation becomes less and less important in determining children's educational outcomes. Moreover, the impact of mother's occupation and sex have reversed over time: girls with working mothers born in the second half of 20th century more frequently succeed in school.

Table 8: Effect of family size on educational outcomes by age classes, Italy and France, university degree

Age class *(COHORT)*	**25-44** *(ITA:1959-1978)* *(FRA: 1955-1974)*				**45-64** *(ITA:1939-1958)* *(FRA: 1935-1954)*				**65 and more** *(ITA:1901-1938)* *(FRA: 1900-1934)*			
	Italy exp(β)		**France exp(β)**		**Italy exp(β)**		**France exp(β)**		**Italy exp(β)**		**France exp(β)**	
Number of siblings												
0	0.95	n.s.	0.94	**	1.10	n.s.	0.98	n.s.	0.95	n.s.	1.08	n.s.
1 (Ref)	1		1		1		1		1		1	
2	0.74	***	0.79	***	0.75	***	0.87	***	0.73	n.s.	0.87	***
3	0.61	***	0.60	***	0.55	***	0.71	***	0.67	***	0.78	***
4	0.51	***	0.46	***	0.33	***	0.63	***	0.35	***	0.65	***
5 and more	0.20	***	0.35	***	0.30	***	0.44	***	0.37	***	0.42	***
Father's Occupation												
Upper Class	8.20	***	6.77	***	10.27	***	11.29	***	15.88	***	19.27	***
White-collars	5.04	***	2.38	***	6.51	***	3.28	***	6.52	***	4.75	***
Self-Employers	2.16	***	2.10	***	1.93	***	3.06	***	1.77	**	3.91	***
Blue-Collars (Ref)	1		1		1		1		1		1	
Mother Activity												
Active	1.09	n.s.	0.96	**	0.76	***	0.90	***	0.69	**	0.64	***
Sex												
Males	1.33	***	1.20	***	0.84	***	0.70	***	0.35	***	0.34	***
Age	1.02	***	0.95	***	0.96	***	0.95	***	0.98	*	0.99	**

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

Table 9: Effect of family size on educational outcomes by cohorts, Italy and France, secondary school degree

Age class (COHORT)	Secondary School Degree 25-44 (ITA:1959-1978) (FRA: 1955-1974)				45-64 (ITA:1939-1958) (FRA: 1935-1954)				65 and more (ITA:1901-1938) (FRA: 1900-1934)			
	Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)	
Number of siblings												
0	0.88	*	0.92	**	1.17	**	1.01	n.s.	1.07	n.s.	1.16	***
1 (Ref)	1		1		1		1		1		1	
2	0.65	***	0.76	***	0.68	***	0.81	***	0.64	***	0.83	***
3	0.48	***	0.57	***	0.47	***	0.65	***	0.59	***	0.71	***
4	0.30	***	0.44	***	0.34	***	0.54	***	0.33	***	0.61	***
5 and more	0.17	***	0.33	***	0.22	***	0.37	***	0.27	***	0.42	***
Father's Occupation												
Upper Class	5.03	***	6.93	***	7.26	***	10.76	***	12.44	***	16.94	***
White-collars	4.27	***	2.32	***	6.35	***	3.11	***	6.76	***	4.19	***
Self-Employers	1.55	***	2.16	***	1.73	***	2.96	***	1.45	***	3.61	***
Blue-Collars (Ref)	1		1		1		1		1		1	
Mother Activity												
Active	1.03	n.s.	0.96	**	0.75	***	0.85	***	0.53	***	0.54	***
Sex												
Males	1.33	***	1.41	***	0.77	***	0.92	***	0.49	***	0.63	***
Age	0.96	***	0.96	***	0.92	***	0.95	***	0.96	***	0.99	***

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

Model 4

The last model that will be presented includes the interaction between the number of siblings and birth order. This enables to investigate whether there are differences in educational outcomes according to birth position within same-size families. Firstborns in two sibling families are used as the reference category. Except for only children and second-borns of two, all of the positions are significantly disadvantaged in terms of educational achievement.

A further check on possible birth order disparities within each family size has been performed, always taking firstborns as the reference category. It has been tested whether differences between pairs of coefficients were significant (analytical results not shown). In Italy, within families of three or four children, only the firstborn have greater probabilities of obtaining high levels of education. On the other hand, in France, the effect of birth order is strong and regular, regardless of family size: the firstborn is favored, the lastborn is penalized, and those in the middle are in an intermediate position.

Table 10: Odds Ratios of achieving high levels of education, model with interaction between family size and birth order.

	University Degree				Secondary School Degree			
	Italy exp(β)		France exp(β)		Italy exp(β)		France exp(β)	
Siblings*Birth Order								
Only Child	0.96	n.s.	0.97	n.s.	0.94	n.s.	0.95	*
1st between 2 (Ref)	1		1		1		1	
2nd of 2	0.90	n.s.	0.85	***	0.89	n.s.	0.84	***
1st of 3	0.86	*	0.85	***	0.70	***	0.81	***
2nd of 3	0.63	***	0.71	***	0.58	***	0.67	***
3rd of 3	0.69	***	0.69	***	0.68	***	0.64	***
1st of 4	0.81	n.s.	0.70	***	0.55	***	0.67	***
2nd of 4	0.53	***	0.58	***	0.44	***	0.52	***
3rd of 4	0.54	***	0.55	***	0.47	***	0.51	***
4th of 4	0.51	***	0.53	***	0.47	***	0.49	***
Father's Occupation								
Upper Class	9.38	***	8.09	***	5.91	***	8.46	***
White-collars	5.58	***	2.80	***	4.64	***	2.77	***
Self-Employers	2.06	***	2.43	***	1.50	***	2.47	***
Blue-Collars (Ref)	1		1		1		1	
Mother Activity								
Active	0.93	n.s.	0.87	***	0.87	***	0.80	***
Sex								
Males	1.04	n.s.	0.96	***	1.03	n.s.	1.18	***
Age	0.99	***	0.96	***	0.96	***	0.96	***

n.s. p>0.10; * 0.05<p<0.10; ** 0.01<p<0.05; *** p<0.01

6. The issue of assessment causality

The estimation of the probability of achieving a secondary school (and university) education by means of a logit model shows that as the family size increases, the children's opportunity of reaching high levels of education -*dilution effect*- decreases in Italy and in France, after controlling for both socio-economic backgrounds factors and demographic variables.

However, the model as it has been identified so far does not solve the issue of causality. Indeed, the observed dilution effect could be biased from unobserved heterogeneity. In particular, parental preferences towards their children education could jointly affect the number of children they decide to have and their children's future education. In other words, couples who are strongly prone towards having "high quality" children would try to keep the family size small in order to ensure to each of them the required amount of material and non-material resources. On the contrary, those who carry on having many children might be characterized by less interest towards their children education. Consequently, the presence of a parental heterogeneous preference needs additional "adjustments" in order to estimate the effect of family size correctly.

In this study the issue finds a solution in the use of an instrumental variable (IV) which exogenously accounts for family size variation. Among the set of possible instruments that might be

used for this purpose[1], the only one which is available from the data is the **same sex of the two eldest children**. Unfortunately, it can only be derived for families who have 2 or 3 children; for those characterized of more than 3 children, data do not allow for a distinction of sex composition.

In both countries' data, couples with two same sex children (either 2 boys or 2 girls) are more likely to move to a third child than those who already have 2 opposite sex children. Specifically, the proportion of the former exceeds the one for the latter of 7.8 and 8.2 percentage points, for Italy and France respectively.

Table 11: Proportion of families progressing from 2 to 3 offspring, according to the sex of the two oldest children

	Italy (%)	France (%)
(a) 2 boys	41.4	54.3
(b) 2 girls	43.2	49.7
Same-sex (average of a & b)	**42.3**	**52.0**
Mix-sex	**34.5**	**43.8**

Additionally, the selected variable fulfills the main conditions for instrument validity: (1) it must be correlated with the "treatment" (i.e. having a 3-children family) and (2) it must have an effect on the outcome only through the "treatment". These assumptions seem to be reasonable both under theoretical and empirical grounds: (1) if parents with two same sex children are more likely to have a third one than those who already have children with mixed sex composition, the transition from 2 to 3 children is for them more probable; (2) the parental preference for children with mixed sex composition does not have any impact on the children educational attainment by itself, but only by increasing the family size.

The appendix explains in details how the instrumental variable has been constructed and the empirical proof of its statistical validity; in this section, only 2SLS estimates of the dummy endogenous variable model will be showed.

As mentioned above, the instrument has been practically created as an indicator of whether the first 2 children among families of 2 or 3 children were of the same sex; it has been applied to the variable which discriminates between families with 2 or 3 children. The following table (Table 12) presents a comparison of logit and 2SLS results for the probability of achieving, respectively, a university or a secondary school degree of education.

Table 12: Logit and 2SLS estimates of the effect of family size on university and secondary degrees achievement

A. with covariates								
	Italy				France			
Dependent variable	logit		IV		logit		IV	
University Degree	-0.30	***	-0.21	n.s.	-0.21	***	**0.13**	**
s.e.	*(0.0597)*		*(0.2631)*		*(0.0158)*		*(0.0543)*	
Secondary Degree	-0.40	***	0.30	n.s.	-0.26	***	**0.22**	***
s.e.	*(0.0376)*		*(0.3750)*		*(0.0136)*		*(0.0567)*	
B. without covariates								
University Degree	-0.44	***	-0.30	n.s.	-0.25	***	**0.27**	***
s.e.	*(0.0570)*		*(0.2529)*		*(0.0142)*		*(0.0637)*	
Secondary Degree	-0.56	***	0.14	n.s.	-0.29	***	**0.38**	***
s.e.	*(0.0338)*		*(0.3841)*		*(0.0119)*		*(0.0684)*	

[1] A common variable that has been used in the present field of research is the twin variable (Black, Devereux, Salvanes 2005, Angrist Lavy Schlosser, 2006). Other authors (Merlier, Monso 2007 and Maurin, Moschion) proposed also the use of age distance between parents or the TFR of the mother's birth cohort.

The parameters estimates of the "true" effect of moving from 2 to 3 children on the probability of achieving a university or secondary school degree turn out to be non significant in Italian case. In the French context an increase of family size due to siblings' sex composition raises children's probability of achieving the analyzed levels of education.

7. Discussion and conclusions

The findings of the present analysis show that in Italy and in France being an only child or having only one sibling is associated with greater probabilities of obtaining at least a high school education, even when controlling for parent's occupation, mother's working status, sex, and year of birth. This result continues to hold when different social strata are considered separately, with the exception of the French bourgeois with four children or less. Moreover, in France the effect of birth order is stronger than in Italy, and lastborns are more strongly penalized. Finally, the interaction with cohort groups results in an unchanged pattern of the dilution effect over the course of the 20th century.

However, the risk of spuriousness in the relationship between family size and educational achievement is embodied in the fact that parental preferences for their children future education are heterogeneous. In fact, if parents are strongly interested in having "high quality" children, they will try to keep the family size relatively small in order to ensure to each of them enough resources for their studies. If, on the contrary, parents are not interested in their children "quality", they will have as many children as they desire, without considering any educational constraint.

Once the siblings' sex composition instrument is introduced by means of the *dummy endogenous variable model*, in Italy the negative effects of family size on education are not significant anymore. Therefore, the causal inference on the effect of moving from 2 to 3 children is in this case supporting an absence of any dilution effect. Therefore, people whose parents decided to have an additional child after having the first 2 children of the same sex are no more disadvantaged in terms of probability of achieving high levels of education.

Things are different in France, where an increase from 2 to 3 children *when the eldest two are of the same sex*[2], leads to higher probabilities of achieving university and secondary school degrees (the effects become positive). In particular, when the instrument is introduced, moving from 2 to 3 children increases siblings' probability of achieving a university degree by 13 percentage points and a secondary school degree by 22 percentage points. The new "true" effect is "depurated" from the sources of heterogeneity linked to parental preferences towards their children education, and it can derive from the positive impact of a family-friendly policy as well as from other external factors which all contribute in improving people's educational chances. In this context, stating an absence of dilution effect in France (at least for families with 3 children) and conversely a presence of educational advantages is not hazardous. Finally, the positive relationship can be strictly related to the financial help that French families receive from government ("quotient familial").

The use of an instrumental variable aims in keeping under control a possible source of bias derived from the presence of unobserved heterogeneity in the issue of assessing causality. Causality evaluation in turn distinguishes different effects, allowing for a precise intervention in terms of policy decision. The present results suggest, on the one hand, that policies simply aiming in discouraging fertility over a certain threshold without intervening in changing parental constraints towards fertility will not increase children's probability of achieving high levels of education. On the other hand, policies helping financially families with children would have first the effect of further encouraging fertility;

[2] The parameter estimates apply only to families who move to an additional child because the eldest 2 children are of the same sex (local average treatment). Thus, there is a lack of information for families with only two children and for those who move to 3 but having already two children of different sex.

subsequently, couples interested in their children education will have more resources to invest on each child schooling, whereas those who are not interested will just invest in having more children.

This investigation leaves some issues open to further consideration. (1) Which government policies should be implemented to reduce the negative effect of both family size and socio-economic inequalities in accessing higher education? (2) Are our results confirmed if we consider individual occupation as a response variable, rather than educational achievement? (3) Are there other determinants of dilution effect, others then family size (i.e. parental divorce, parentless condition)?

Acknowledgements

Special thanks go to Laurent Toulemon, without whose contributions this project would not have been possible. We also thank Enrico Rettore and Luca Pieroni for their useful econometrics suggestions.

References

Angrist, J. D., Lavy V. and Schlosser A. "New evidence on the causal link between quantity and quality of children." MIT working paper, 2005.

Angrist J. D., Evans W. N., *Children and their parents' labor supply : evidence from exogenous variation in family size*. The American Economic Review, Vol. 88, No. 3, (Jun., 1998), pp. 450-477

Angrist J. D., Imbens G. W., Rubin D. B., *Identification of causal effects using instrumental variables.* Journal of the American Statistical Association, Vol. 91, No. 434, (Jun., 1996), pp.444-472

Baum, C. F., Schaffer M. E., and Stillman S. (2003), "*Instrumental Variables and GMM: Estimation and Testing*" Boston College Working Paper No.545, Boston MA.

Black, S. E.; Devereux P. J.; Salvanes K. G.. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education," *Quarterly Journal of Economics*, May 2005, 120(2): 669-700.

Blake. J.. *Family Size and Achievement.* Berkeley: University of California Press. c1980. http://ark.cdlib.org/ark:/13030/ft6489p0rr/

Conley, Dalton. “Parental Education Investment and Children’s Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility,”
Conley, Dalton *The Pecking Order: Which Siblings Succeed and Why* (New York, NY: Pantheon Books, 2004b).
Downey D. B.. *“When Bigger is not Better: Family Size. Parental Resources. and Children’s Educational Performance”*. American Sociological Review. Vol. 60. No. 5. (Oct.1995). pp.746-761
Duru-Bellat M.. Kieffer A.. 2000. “La démocratisation de l’enseignement en France: polémiques autour d’une question d’actualité”. Population (French Edition). 55e Année. No. 1. (Jan. – Feb... 2000). pp. 51-79.
Easterlin R. A.. “Birth and fortune. The impact of numbers on personal welfare.”. The University of Chicago Press. 1987.
Gary-Bobo R.J.. Prieto A.. Picard N.. “Birth-order and sibship sex-composition effects in the study of education and earnings”. 2006.
Goux, Dominique, and Eric Maurin. 2005. “The Effect of Overcrowded Housing on Children’s Performance at School.” *Journal of Public Economics* 89(5-6): 797-819.
Heer D. M.. “Effects of Sibling Number on Child Outcome”. Annual Review of Sociologo. Vol. 11 (1985). pp. 27-47.
Hosmer. Lemeshow. “Applied logistic regression”. Wiley. 1984
Impicciatore R.. Dalla Zuanna G.P.*.. A difficult social mobility: the education of children of southern parents emigrated to central and north Italy*.
Istat. “Famiglia. soggetti sociali e condizione dell’infanzia. La mobilità sociale. indagine multiscopo sulle famiglie.1998”. Informazioni. 2000
Knodel J.. Havanon N.. Sittitrai W.. *Family size and the education of children in the context of rapid fertility decline*. Population and development review. 1990. vol. 16. No.1
Knodel J..Wongsith M.. *Family Size and Children’s Education in Thailand: Evidence from a National Sample.* Demography. vol.28. No.1. 1991
Maurin, E. and J. Moschion (2006), *The Social Multiplier and Labour Market Participation of Mothers,* IZA Discussion Paper No. 2513, Bonn
Merllié D., Monso O., *La destinée sociale varie avec le nombre de frères et sœurs*. France, portrait social – édition 2007
Meurs D.. Pailhé A.. Simon P.. “Persistance des inégalités entre générations liées à l'immigration : l'accès à l'emploi des immigrés et de leurs descendants en France”. 2006. Population (French Edition). 61e Année. No. 5-6. (Sep. – Dec.. 2006). pp. 763-802.
Piketty T.: « The Impact of divorce on school performance : evidence from France, 1968-2002 », CEPR discussion paper series no 4146, 2003
Toulemon L.. “La transformation des fratries au cours du XX^{e} siècle”. Colloque « Frères. soeurs. jumeaux : passé et présent des fratries ». 16^{e} entretiens du Centre Jacques Cartier. Institut des sciences de l'Homme de Lyon. 1er et 2 décembre 2003
Zarca B.. “L’héritage de l’indépendence professionnelle selon les lignées. le sexe et le rang dans la fratrie”. Population (French Edition). 48e Année. No. 2. (Mar. Apr.. 1993a). pp. 275-306.
Zarca B.. “L’héritage de l’indépendence professionnelle: un ou plusieurs élus au sein de la fratrie?”. Population (French Edition). 48e Année. No. 4. (Jul. Aug.. 1993b). pp. 1015-1042.
Zarca B.. “L’héritage et la mobilité sociale au sein de la fratrie. I. L’héritage et la mobilité sociale différentielle des frères”. Population (French Edition). 50e Année. No. 2.(Mar. Apr.. 1995a). pp. 331-356.
Zarca B.. “L’héritage et la mobilité sociale au sein de la fratrie. II. L’héritage et la mobilité sociale différentielle des sœurs”. Population (French Edition). 50e Année. No. 4/5.(Jul. - Oct.. 1995a). pp. 1137-1154.

Appendix: Instrumental variable construction and validity

1. The construction of the instrumental variable

Practically, the instrumental variable "same sex of the two eldest children" has been created as an indicator of whether the first 2 children of a family of 2 or 3 children were of the same sex.

The construction of the instrument is quite straightforward in the situation of families composed of 2 children. Indeed, the possible combinations of sibling's compositions by sex and birth order are 8 (=2x2²), 4 for each considered birth order. Specifically, the information on siblings is collected at the individual level and for each person data on the number of brothers and sisters as well as on the birth order is available. Moreover, since the individual who has been interviewed can either be the first or the second-born, 2 different cases are possible for each composition in this case. Comparing for instance case 3_1st with case 4_2nd, the group is always composed of a girl and a boy, but in case 3 the information is available for the firstborn girl, while in case 4 the information regards the second-born boy.

Table 2: Possible siblings sex compositions and instrumental variable value among 2-children families[3]

case	Firstborns (1st)	Z	Second-borns (2nd)	Z
1	**G**G	1	G**G**	1
2	**B**B	1	B**B**	1
3	**G**B	0	B**G**	0
4	**B**G	0	G**B**	0

Concerning the 3 children families, the number of different combinations becomes 24 (3x2³), 8 for each birth order.

Table 3: Possible siblings sex compositions and instrumental variable value among 3-children families[4]

case	Firstborns (1st)	Z	Second-borns (2nd)	Z	Third-borns (3rd)	Z
1	**G**GG	1	G**G**G	1	GG**G**	1
2	**B**BB	1	B**B**B	1	BB**B**	1
3	**G**BB	0	B**G**B	0	BB**G**	1
4	**B**GG	0	G**B**G	0	GG**B**	1
5	**G**BG	0	B**G**G	0	BG**G**	0
6	**G**GB	1	G**G**B	1	GB**G**	0
7	**B**GB	0	G**B**B	0	GB**B**	0
8	**B**BG	1	B**B**G	1	BG**B**	0

[3] Z=instrumental variable: same sex of the two oldest children; G=girl; B=boy
[4] Z=instrumental variable: same sex of the two oldest children; G=girl; B=boy

In this case, some combinations are impossible to be distinguished. Looking for instance at cases 5_1st and 6_1st the situation is the following: the reference individual is a firstborn girl, who has a brother and a sister. Since the information on the birth order of next children is missed, the case 5_1st in which she has a brother right after her and then a sister (i.e. the 2 oldest children are of mixed sex) cannot be distinguished from the case 6_1st in which she has first a sister and then a brother (i.e. the 2 oldest children are of the same sex). The same issue remains valid when the observed individual is a firstborn boy with one brother and one sister (7_1st and 8_1st) as well as when he/she is the second-born with a brother and a sister (cases 5/6/7/8_2nd). The situation is different when the third-born is observed: even if the 4 cases are not completely identifiable, they all take the value zero with regards to the instrumental variable. Nonetheless, French data enable to identify one of the three groups that are counted together and namely that regarding the second-born. Indeed, the French questionnaire does not ask to the respondent directly for his/her birth order, but for the number of brothers or sisters born before him/her. Thus, if the individual is the second born and he/she has 1 brother and 1 sister a distinction between the case when the instrument is zero or one is possible (i.e. cases 5, 6, 7 and 8 are all identifiable for the second-born).

However, this does not help in distinguishing the "mixed" category among firstborns (cases 5/6/7/8_1st) neither to solve the problem for the Italian data. Thus, the solution here proposed assumes the followings:

(1) As far as Italian data are concerned, the distribution of each birth order among families with 3 children of the same sex can be assumed as equal to the distribution of each birth order among families in which the first 2 children are of the same sex and the third is different. In other words, the hypothesis is that the total number of GGGs (no matter the birth order) equals the sum of GGBs (cases 1_1st + 1_2nd + 1_3rd = **6_1st** + **6_2nd** + 4_3rd). Since the three GGG's cases are countable, it is easy to obtain each GGG's birth order proportion on the sum of GGGs. Further, knowing the GGB_3rd number (case 4_3rd), only two cases remain undefined. Thus, in order to establish the proportion of the unknown GGB_1st and GGB_2nd, the assumption which assigns the (known) proportion of GGG_1st to that (unknown) of GGB_1st (and the one of GGG_2nd to that of GGB_2nd) does not seem to be too dangerous. Similarly, the same holds true when boys are examined.

(2) For French dataset, the solution becomes even easier to find. Indeed, maintaining the identity between the frequency of the same sex 3-children families and the number of families with same sex of first 2 children, the unknown cases collapse into 1. In other words, assuming that 1_1st + 1_2nd + 1_3rd = **6_1st** + 6_2nd + 4_3rd, only case 6_1st has to be estimated and it can be simply derived from the algebraic sum of the known terms. Likewise, the same procedure can be applied to boys' cases.

Henceforth, the combinations for which the 2 oldest siblings of the same sex instrument takes the value 1 can be randomly imputed in each sibling's group, following the distribution obtained in the previous stages.

2. Statistical validity of the instrument

In this section, the proof of the instrument validity under a firm statistical point of view will be described. In practice, once one wants to assess the causal effect of the family size on educational outcomes, the use of a *dummy endogenous variable model* is required. Here the Agrist et al.'s approach is replicated (Angrist et al.,1996). If for each individual *i,* Y_i is the observed educational outcome, D_i is the observed "treatment" (i.e. belonging to a 2- or 3-children family) and Z_i is the (hypothesized) instruments as it has been above defined, the model specification would be the following:

$$Y_i = \beta_0 + \beta_1 \cdot D_i + \varepsilon_i \qquad (1)$$

$$D_i^* = \alpha_0 + \alpha_1 \cdot Z_i + \upsilon_i \qquad (2)$$

$$D_i = \begin{cases} 1 \;\; if \;\; D_i^* > 0 \\ 0 \;\; if \;\; D_i^* \leq 0 \end{cases} \qquad (3)$$

In this model β_1 (i.e. the parameter relative to the treatment) simply represents the causal effect of D on Y. In order to **correctly** identify β_1 and thus to obtain an unbiased estimation of the causal effect, three main assumptions must hold true:

1) *The covariance between the correct instrument and the instrumented covariate must differ from 0*. In principle, this is straightforward: the parity progression to a third child obviously increases the family size. This condition is verified empirically looking at the coefficient of the first stage regression: for both Italian and French data α_1 turns out to be different from 0.

2) *Any effect of Z on Y must be through an effect of Z on D*. This means, together with the absence of Z in Equation (1), that Z_i is uncorrelated with error terms ε_i and υ_i (orthogonal error process):

$$E[Z_i \cdot \varepsilon_i] = 0, \quad E[Z_i \cdot \upsilon_i] = 0$$

3) *Theoretical justification of the instrument choice*. This instrument exploits the widely observed phenomenon of parental preferences for a mixed siblings-sex composition. In particular, parents of same-sex siblings are significantly more likely to carry on having an additional child. Because sex mix is almost randomly assigned, a dummy for whether the sex of the first two children matches provides a plausible instrument for further childbearing among women with at least two children (Angrist, Evans, 1998). If the coefficients related to the family size will change to positive after the inclusion of the instrument in the model, it will mean that the educative advantage of being reared in a small family is not related to the family size in itself but to some other unobserved characteristics which distinguish this kind of children from those who were reared in a bigger family.