Human Capital Investments and Family Size in Italy: Estimates using Twin Births' Instrument

Michela Ponzo [a] Vincenzo Scoppa [b, c *]

a) Department of Political and Social Sciences, University of Calabria

b) Department of Economics, Statistics and Finance, University of Calabria

c) Institute for the Study of Labor (IZA), Bonn

This version: 30 September 2021

Human capital investments at an early age appear crucial for individuals' outcomes but their determinants are not yet completely clear. One possible determinant is family size that might affect parental time and economic resources invested in children. This aspect is related to the children quantity/quality trade-off proposed by Becker that has been investigated only for a few countries because of data limitations. We investigate this issue for Italy – even in the absence of Census data relating family of origin with children educational outcomes – using many waves of the Survey on Household Income and Wealth of Bank of Italy and focusing on the educational achievement of 19-22 years old. We use twin births as an instrumental variable to identify exogenous variations in family size. In partial contrast with findings from other developed countries, we find a significant negative effect of family size on children's educational achievement. We show that these findings are robust to a number of checks and the effects are stronger for daughters, for low income families and when spacing between births is limited.

Keywords: Educational Outcomes; Family Size; Quantity-Quality Children Trade-off; Twin Births; Instrumental Variables.

JEL Classification Numbers: J13; J24; I21; C36.

[*] M. Ponzo, Department of Political and Social Sciences, University of Calabria. E-mail: michela.ponzo@unical.it . V. Scoppa, Department of Economics, Statistics and Finance "Giovanni Anania", University of Calabria. E-mail: v.scoppa@unical.it We would like to thank Guido de Blasio and Maria De Paola for useful comments and suggestions. We are also very grateful to the Research Unit of Bank of Italy for the help with the use of the data.

1. Introduction

Investments in human capital at an early age are key for the development of children and subsequent adults' outcomes (Heckman, 2006; Cunha and Heckman, 2007) but their determinants are not yet completely explored. Parental decisions on family size might affect significantly these investments: one additional child in the family could dilute parental time and economic resources devoted to each child, reducing their educational achievement.

This issue is related to the trade-off between quantity and quality of children proposed by Becker and coauthors (Becker, 1960; Becker and Lewis, 1973; Becker and Tomes, 1976): parents derive utility from both children's quantity and quality (more spending on each child) and the income elasticity for children's quality is assumed to be higher than the income elasticity for children's quantity. Quantity and quality are related through the household's budget constraint: if children's quality increases, increasing quantity (more children) becomes more expensive. On the other hand, if quantity increases, quality (which is assumed to be the same for each child) becomes costlier and will be reduced.

To examine empirically this question researchers have to tackle two big challenges. First of all, data availability is an issue: it is very hard to link data on family of origin and siblings with data on their educational achievements, labor market outcomes and so on. While it is common to observe the educational achievement of individuals when they have left home, matching these data with detailed information on the family of origin is rarely possible with the existing datasets. Secondly, endogeneity problems plague the estimations: family size is not exogenously determined but is related to many other observable and unobservable parental characteristics that can also affect children educational achievements, producing estimation biases.

The empirical evidence on the quantity-quality trade-off is quite ambiguous, with some works finding a null effect for some countries (Black, Devereux and Salvanes (2005) for Norway; Angrist, Lavy and Schlosser (2010) for Israel) while others finding negative effects of family size on some outcomes (Rosenzweig and Zhang (2009) for China; Åslund and Grönqvist (2010) for Sweden; Grawe (2008) for UK, among others).

The relationship between family size and children's education has not been investigated for Italy. Italy is a country with a surprising low share of individuals with tertiary education: in 2018 only 19% got a tertiary education among 25-64 years old, while the OECD average is 37% (OECD at a glance, 2019). At the same time, Italy recorded high fertility rates until mid-70s. However, tertiary education is increasing in the last decade: currently 28% among 25-34 years old have attained a tertiary education. At the same time, maybe not by coincidence, the Italian fertility rate has fallen at historical low levels since Nineties (in 2018 to 1.3 children per woman, one of the lowest in the OECD countries, whose average is 1.6).

We investigate the relationship between family size and children's educational achievements for Italy and tackle the problem of linking family of origin and children's outcomes using many waves of the Survey on Household Income and Wealth (SHIW) of Bank of Italy, and focusing on the educational achievement of 19-22 years old that still live in their family of origin, using the academic generalist secondary school track (Lyceum) as a predictor for University enrollment. To avoid endogeneity problems, we adopt an Instrumental Variable estimation strategy using multiple births as an instrument to identify exogenous variations in family size.

In this way, we are able to investigate the existence of a trade-off between children's quantity and quality for Italy, even in the absence of Census data relating family of origin with children's education. In partial contrast with findings from other developed countries, we find a strong negative effect of family size on children's educational achievement.

We analyze if these effects are heterogeneous according to family characteristics and we find that the impact of a larger family has a stronger impact for daughters, in the South, for low income families and when spacing (the difference in years among the birth of first and the second child/children) is low.

We also run a number of robustness checks changing the age range we consider in our sample and we find very similar effects. We also use an alternative dataset, the Italian Health Conditions Survey, which reports directly university attendance of individuals and we confirm our main findings. Furthermore, we carry out a Monte Carlo simulation to investigate whether a small sample selection bias due to individuals leaving earlier their home might determine our results and we show that in fact our coefficient of interest tends to be biased towards zero, suggesting that we are estimating a lower bound.

The paper is organized as follows. Section 2 discusses the related literature. In Section 3 we describe our data and how we build our sample. In Section 4 we investigate the relationship between family size and educational achievement using an OLS estimator. In Section 5 we present our Instrumental Variables estimates and in Section 6 we investigate if the effects are heterogeneous according to individual and household characteristics. Section 7 is devoted to robustness checks. In Section 8 we deal with the problem of sample selection through a Monte Carlo simulation. Section 9 offers some concluding remarks.

2. Related Literature

An early empirical evidence tried to test the quantity/quality trade-off between family size and child quality, generally supporting a negative impact of family size (see Schultz, 2005, for a review). For example, Hanushek (1992), Björklund et al., (2004), and Holmlund, (1988 show that children's educational achievements are negatively correlated with family size. However, only few of these findings can be considered as causal since family size is typically affected by endogeneity problems.

In general, the empirical investigation of the trade-off between child quantity and quality in a family is complicated by the simultaneous determination of both variables: family size and child investments are jointly chosen by parents and, hence, they are both affected by unobservable parental preferences and household characteristics.

To address this endogeneity problem, economists have used different natural experiments to exploit exogenous variations in family size: twin births, siblings' sex-composition and fertility shocks, but no clear results emerge from the empirical literature. [1],[2]

Rosenzweig and Wolpin (1980), in a pioneering study that uses multiple births as an instrument in an IV identification strategy, with a small sample of about 1,600 children from India, find that family size – as affected by the birth of twins – has a negative impact on children's (aged 5-14) educational attainments.

Using data from the whole population of Norway, Black, Devereux and Salvanes (2005) also use twins as an exogenous source of variation to estimate family size effects on children's education and adult earnings. They find that the effect of family size shrinks to almost zero after controlling for birth order (though the standard errors are very large), and that there is monotonic decline in educational attainment by birth order.

A similar study by Angrist, Lavy and Schlosser (2010) that uses both twin births and siblings' gender composition as instrumental variables finds no evidence for a quantity-quality trade-off of children in Israel, even considering higher-fertility populations.

Åslund and Grönqvist (2010) find no effect of family size on long-term educational attainment in Sweden, but find a negative impact of family size on school grades among children of more vulnerable families, such as those with large sibships and low-educated parents.

Caceres-Delpiano (2006) investigates the impact of the number of children on investment on some inputs in children's human capital and other outcomes for the US and finds evidence of a trade-off only for some outcomes. He finds that families facing an exogenous change in family size due to multiple births re-allocate resources consistent with Becker's model in that an additional younger sibling reduces the likelihood that older siblings attend a private school and increases the likelihood that their parent's divorce. On the other hand, he finds little evidence that an exogenous change in family size affects educational achievement such as grade retention or the highest grade completed.

Grawe (2008) finds evidence of a trade-off between family size and several child outcomes for Britain. In particular, he split the sample on the basis of father's earnings and finds strong negative family size effects on children's education even for the richest family (in the top quintile of the income distribution), suggesting that the trade-off is not a matter of financial resources but rather a problem of time constraints.[3]

All in all, the quantity/quality trade-off seems to be absent in developed countries, like Norway or Sweden, where there are both a well-functioning public education system and generous support for childbearing and childcare. In these contexts, families receive public support and they can protect children's

[1] A frequently used alternative instrument, instead of twin births, consists of instrumenting the number of children with the gender mix of children born in the family. Generally, parents prefer to have offspring of both genders (Angrist et al., 2010; Becker et al., 2010; Fitzsimons and Malde, 2014), and so those having the first two children of the same sex are more likely to have an additional child. Finally, other instruments that have been proposed include infertility (Bougma et al., 2015) and miscarriage (Hotz et al., 1997; Miller, 2009).

[2] A parallel literature has evaluated the effect of fertility on parental outcomes, such as female labour supply. See, for example, Bronars and Grogger (1994) and Angrist and Evans (1998).

[3] Rosenzweig and Zhang (2009) using data from China find negative effects of family size on children's education. Similar results are shown by Ponczek and Souza (2012) for Brazil.

quality. In contrast, studies from developing countries have often found evidence in support of the quantity-quality trade-off (Rosenzweig and Zhang, 2009).

3. The Data

This section describes the data and the criteria we have followed to build the sample. The data source for our empirical analysis is the Survey of Household Income and Wealth (SHIW) that is conducted every two years by the Bank of Italy on a representative sample of about 8,000 Italian households.[4] The SHIW collects detailed information on the demographic and social characteristics of all the individuals in a household, such as age, gender, marital status, education, region of residence, and on their working activity (earnings, employment status, type of occupation, industry, work experience, and so on).

In our main analysis, we pool together 11 waves of the SHIW conducted from 1995 to 2016.[5]

To select the sample, we adopt a number of criteria. We use only children from 19 to 22 years old since they have almost certainly completed secondary education (and therefore we are able to observe the type of secondary school they have attended) and it is very likely that they are still at home with their parents: in this way we have information on the family of origin and on our proxy for educational achievement.[6] Since these subjects typically have not yet completed their educational path, as a proxy for educational achievement we use the type of High School attended by individuals.

The Italian secondary school system can be described as tripartite, with an academic generalist track ("Lyceum"), a technically oriented education (Technical schools) and a more labor market orientated track (Vocational or Professional schools). Track selection is a relevant factor for individual future career since the type of secondary school strongly affects university attendance. Lyceum is considered the most prestigious secondary educational track and provides an in-depth, general knowledge aimed at preparing students for university. In contrast, Technical and Vocational schools offer an education oriented toward more practical subjects, enabling the students to start searching for a job as soon as they have completed their studies.

Preliminarily, to document to what extent Lyceum is related to University enrolment we use data on ISTAT 2015 Survey of High School Graduates (sampling individuals graduated in 2011). Using about 26,000 observations and estimating a Linear Probability Model we find that having attended a Lyceum increases the probability of going to College of 52 percentage points on average, increasing the probability of College enrolment from 40% to 92%, (*t*-stat=106.7), see column (1) of Table A1 in the Appendix. In column (2), controlling for Female, Immigrant, and dummies of Region of Residence, we find very similar results.

[4] SHIW data are freely available at www.bancaditalia.it. These data have been used, among others, by Andini et al. (2018), Jappelli and Pistaferri (2020) and Guiso et al. (2004). We refer to the Appendix of the latter work that contains detailed information about the dataset.
[5] We use the following waves: 1995, 1998, 2000, 2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016. Unfortunately, information on the exact age of children are not available for previous waves.
[6] We deal with possible sample selection biases in Section 8.

With respect to Vocational schools (attended by about one third of students) attending a Lyceum increases the probability of going to College of more than 70 p.p.[7]

Furthermore, using the whole sample of SHIW with individuals of any age equal or above 26 (167,000 obs.) we also estimate with OLS the probability of graduating from University in relationship to the choice of a Lyceum as High School. In column 1 of Table A2 of the Appendix, we find that having attended a Lyceum increases the probability of graduating of 40 p.p. (in the first basic specification from 7.4% to 47.8%; *t*-stat=87.1). We obtain very similar results (+37-38 p.p.) when we control for gender, a quadratic function of year of birth (or using, in alternative, birth of year dummies), and geographical areas in columns (2)-(4) of Table A2.[8]

We then turn our attention to our sample of 19-22 years old living with their parents and we try to verify how representative this sample is of the whole population aged 19-22. To investigate this issue, we verify in the complete SHIW dataset at which age people are likely to leave home to form a new family. We build a dummy *Own Family* equal to one if an individual *i* lives alone or he/she is the head of the family or his/her partner; *Own Family* is instead equal to zero if an individual *i* is a son/daughter in a family. We regress *Own Family* on a dummy for each age level. Results are reported in Table A3 in the Appendix: we show that the probability of leaving home is about 3 percentage points higher for individuals aged 19-22 with respect to 16-18 years old (the reference category). Therefore, we conclude that only a very small fraction of individuals is leaving home at the age we consider, while almost 97% remains in the family of origin.

From Table A3 we also notice that Lyceum attracts a negative coefficient: students who have attended a Lyceum are about 9 p.p. less likely to leave home. So, in our main analysis in which we use Lyceum as a dependent variable a sample selection is at work. In Section 8 we will show that this sample selection imparts a bias towards zero to our coefficient of interest.

Since the SHIW dataset does not include an explicit identifier for twins, we define twins as children who were born in the same calendar year in the same family.[9] We only include in the sample the households that report that no other children have left home and live outside the family (since we do not know the age and educational attainments of the latter).

[7] In the same Survey of High School Graduates, we also find that Lyceum is related to other important educational outcomes: one additional point of Final Grade in the Lower Middle School increases the probability of attending a Lyceum of about 20 p.p. (*t*-stat=70.0); a student retained in a grade has a lower probability of attending a Lyceum of 15 p.p. (*t*-stat=–25.6); having attended a Lyceum increases by 12 the number of credits gained at university.

[8] In line with these findings, Agarwal, Brunello and Rocco (2019) document – using a different dataset (PLUS-ISFOL) – the huge differences in the probability of attaining a College Degree according to the type of High School attended by individuals.

[9] The exact date of birth is not publicly available in the SHIW dataset. However, the Research Unit of Bank of Italy (to whom we are very grateful) kindly agrees to check if our list of "twins" have the same date of birth and signaled to us only the cases in which the date of birth was different. In these (very few) cases, we consider as brothers – instead of twins – the involved children.

Table 1. Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max	Obs.
Lyceum	0.296	0.456	0	1	8198
# Children	2.273	0.845	1	6	8198
Twins in family	0.026	0.159	0	1	8198
Twin	0.017	0.129	0	1	8198
Female	0.468	0.499	0	1	8198
Age	20.480	1.113	19	22	8198
Birth Order	1.509	0.661	1	6	8198
Mother's Education	9.533	3.975	3	21	8178
Father's Education	9.805	4.009	3	21	8156
Mother's Age	27.816	5.051	16	46	8165
Father's Age	31.511	5.328	18	57	8140

Notes: 11 waves of the Survey of Household Income and Wealth (SHIW) dataset (1995-2016).

By focus on young adults of 19-22 years, families without children are excluded. Furthermore, in our analysis we use only married couples.[10]

Applying our selection criteria of individuals aged 19-22, in our complete sample we have 8,198 observations. About 29.6% of individuals have attended a Lyceum.[11] Family are composed on average of 2.27 children. About 1.8% of our sample are twins.[12] 47% are females, average age is 20.5. Mothers and fathers have acquired nearly 10 years of education[13] and their respective age (at the birth of their children) is 27.8 and 31.5.

The individuals in our sample are born from 1973 to 1997. We also use as controls the geographical areas: residents in the North-West or North-East constitute 38%, 19% live in the Centre and 43% live in the South and on the Islands.[14] Furthermore, in some specifications we take into account father's and mother's Employment Status, including in some specifications a dummy for each of the following categories: Blue-collar; White Collar or Teacher; Cadre; Manager; Member of the arts or professions; Entrepreneur; Self-employed; Family Firm; Unemployed or Non-in-the-Labor-Force.

[10] However, in the samples that we use in our main analyses with families with at least two births, only 17 children come from unmarried parents (0.49%). The results do not change at all if we consider also these observations.

[11] Consider that about 31% of individuals in the sample obtain less than High School. Lyceum are 43% of High School Graduates.

[12] This is in line with national statistics: according to ISTAT Historical Time Series ("Parti semplici e plurimi – Anni 1868-1998 – Serie Storiche – Istat") from 1973 to 1997 about 2.0% of children are twins (about 1% of births are twin births). This is similar to US data (see Caceres-Delpiano, 2006).

[13] Since the SHIW contains information only on the highest educational qualification obtained, we computed the number of years of education using the years of legal duration of the different educational grades, as follows: education is set at 0 for no educational qualification; 5 for elementary school; 8 for middle school; 11 for some high school; 13 for high school; 18 for university; 20 for a postgraduate qualification.

14 North-West includes the following regions: Piedmont, Valle d'Aosta, Lombardy, Liguria; North-East includes Veneto, Trentino Alto Adige, Friuli Venezia Giulia, Emilia Romagna; Centre includes Tuscany, Lazio, Marche, Umbria; South includes Abruzzi, Campania, Apulia, Molise, Basilicata, Calabria; Islands includes Sicily and Sardinia.

4. Family Size and Education: OLS Estimates

We first aim to show that the educational achievement of individuals is negatively related to family size (number of children) using a simple OLS estimator. This can be shown using different datasets.

First of all, we use the same Survey (SHIW) but focus on observations on the head of the family and his/her partner exploiting some questions asking them their respective number of brothers and sisters (still in live), their own educational attainments and the education of their parents. We use 7 surveys (from 1995 to 2008) in which these questions were asked (about 44,000 obs.). We build the variable *Family Size* as the number of children in the family of origin (*Family Size*=#Brothers+#Sisters+1) and we consider individuals between 26 to 55 years old (to reduce measurement errors deriving from siblings' premature deaths).

In this analysis we estimate the following equation:

$$Education_i = \beta_0 + \beta_1 Family\ Size_i + \boldsymbol{\beta_2 X_i} + \boldsymbol{\beta_3 P_i} + \varepsilon_i \qquad [1]$$

where we use as a dependent variable *Education*, the years of education attained; *Family Size* represents the number of children in the family of origin, X_i are individual control variables (gender, age, residence) while P_i are parental characteristics (father's and mother's years of education, age, occupations, etc.)

We estimate with OLS and our estimates are reported in Table 2. In the first three columns we use *Family Size* in linear form, while in columns (4)-(6) we include a dummy for each number of children.

In the first column, controlling only for gender and the year of birth, we find that an additional child in the family reduces years of education by 0.62 and the effect is highly statistically significant (*t*-stat=–49). In the following columns we control for father's and mother's years of education, 5 dummies of geographical area of birth and a quadratic function of birth year.[15] We find a quite relevant impact of –0.46 (*t*-stat=–36), slightly reduced in magnitude: this corresponds to 0.11 SD of the dependent variable for each additional child.

In columns (4)-(6) in which we use dummies for the number of children, we find that while in families with two children the years of education are not lower with respect to family with one child, a number of children greater than two strongly reduces children's educational attainments: in column (6) we find that in families with three children the years of children's education are reduced of 0.61, 4 children reduces education of 1.2, five children reduces of 1.6 and 6 children reduces of 2.1.

We have also used *College Degree* as an alternative dependent variable and we find a relevant and significant negative effect of family size on the probability of attaining a College Degree (1.4 percentage points less for each child) (estimates not reported to save space).

[15] Similar effects are found if we use a dummy for each year of birth instead of a quadratic function (not reported). Again, we find analogous results if we focus only on individuals aged below 40.

Table 2. OLS Estimates. The Impact of the Number of Children on Years of Education

	(1)	(2)	(3)	(4)	(5)	(6)
# Children	-0.620***	-0.463***	-0.464***			
	(0.013)	(0.013)	(0.013)			
# Children=2				0.208***	-0.053	-0.066
				(0.054)	(0.055)	(0.055)
# Children=3				-0.513***	-0.608***	-0.621***
				(0.057)	(0.057)	(0.057)
# Children=4				-1.500***	-1.227***	-1.255***
				(0.063)	(0.063)	(0.063)
# Children=5				-2.122***	-1.589***	-1.603***
				(0.072)	(0.073)	(0.073)
# Children=6				-2.828***	-2.187***	-2.181***
				(0.080)	(0.084)	(0.084)
Female	-0.293***	-0.363***	-0.363***	-0.300***	-0.365***	-0.364***
	(0.035)	(0.034)	(0.033)	(0.035)	(0.034)	(0.033)
Birth Year	0.078***	0.039***	15.740***	0.077***	0.039***	15.689***
	(0.002)	(0.002)	(0.773)	(0.002)	(0.002)	(0.773)
Father's Education		0.296***	0.294***		0.295***	0.292***
		(0.006)	(0.006)		(0.006)	(0.006)
Mother's Education		0.174***	0.179***		0.174***	0.179***
		(0.007)	(0.007)		(0.007)	(0.007)
Birth Year Squared			-0.004***			-0.004***
			(0.000)			(0.000)
Observations	44871	37366	37366	44871	37366	37366
Adjusted R^2	0.094	0.297	0.305	0.100	0.299	0.306

Notes: SHIW waves 1995-2008. Sample: individuals aged 26-55. OLS estimates. The dependent variable is *Education* (in years). Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

Turning to the data that we use in our main analysis (children aged 19-22 still in their family of origin) with a number of observations of about 8,000, we estimate a simple Linear Probability Model for the probability of attending a Lyceum (since in this age range education is not yet complete) in relationship with the number of children in the family and we report our results in Table 3. In the first specification we only control for the gender and age of individual *i*, in column (2) we control for mother's and father's education and in column (3) we also add mother's and father's age, geographical dummies, year of wave dummies. We find a small but significant negative coefficient.

In columns (4) and (5) instead of including the number of children in continuous form, we estimate the same specifications with a dummy for two children, another dummy for three children and so on. We again show that young adults with more siblings tend to attend a Lyceum with lower probability, although the negative effect decreases when several family controls are added.

Table 3. Linear Probability Model. The Impact of the Number of Children on the Probability of Attending a Lyceum

	(1)	(2)	(3)	(4)	(5)
# Children	-0.044***	-0.017***	-0.017***		
	(0.008)	(0.006)	(0.006)		
Two Children				-0.007	-0.010
				(0.018)	(0.018)
Three Children				-0.085***	-0.085***
				(0.021)	(0.020)
Four Children				-0.108***	-0.111***
				(0.033)	(0.033)
Five Children				-0.129*	-0.126*
				(0.068)	(0.067)
Six Children				-0.192**	-0.180**
				(0.078)	(0.075)
Female	0.147***	0.144***	0.144***		0.146***
	(0.012)	(0.010)	(0.010)		(0.012)
Age	0.005	0.010***	0.010***		0.005
	(0.004)	(0.003)	(0.003)		(0.004)
Father's Education		0.029***	0.029***		
		(0.002)	(0.002)		
Mother's Education		0.030***	0.030***		
		(0.002)	(0.002)		
Observations	8198	8136	8136	8198	8198
Adjusted R^2	0.032	0.251	0.252	0.008	0.033

Notes: Linear Probability Model. Sample: SHIW dataset, individuals aged 19-22. The dependent variable is Lyceum. Standard errors, corrected for heteroskedasticity and allowed for clustering at household level, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

In Figure 1 we report the effects (with Confidence Intervals) of the number of children on the probability of attending a Lyceum estimated in column (5) of Table 3: from the Figure emerges very clearly the negative impact of family size on children's educational achievements.

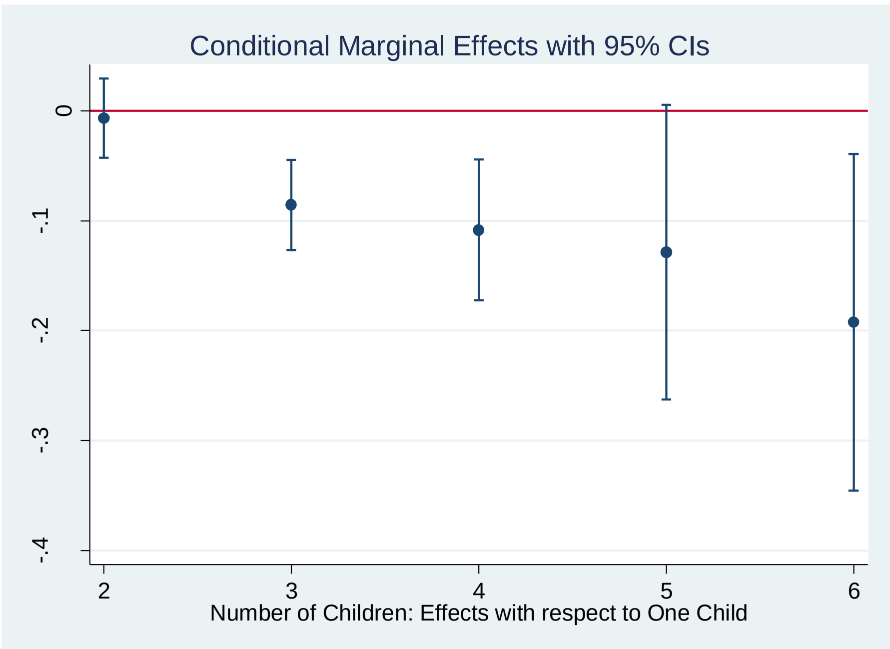

Figure 1. The Impact of the Number of Children on the Probability of Attending a Lyceum

5. *Instrumental Variables Estimates of the Impact of Family Size on Educational Achievement*

In the previous Section we have estimated an OLS model but, as explained above, OLS is likely to be biased since *Family Size* might be correlated with a number of observable or unobservable factors that can affect directly the educational achievement of children.

To overcome this problem in this Section – following the analyses of Black, Devereux and Salvanes (2005) and Angrist, Lavy and Schlosser (2010) – we adopt an Instrumental Variable approach in which *Family Size* is instrumented with the birth of twins. The idea is that the birth of twins represents an exogenous shock to the size of the family, uncorrelated – under certain conditions – with other determinants of educational achievement.

The equation we estimate is slightly different from equation [1]:

$$Lyceum_i = \beta_0 + \beta_1 Family\ Size_i + \beta_2 \boldsymbol{X}_i + \beta_3 \boldsymbol{P}_i + \varepsilon_i \qquad [2]$$

where the educational attainment is proxied with the dummy *Lyceum* (which in turn, as shown in Section 3, is highly correlated with College Degree attainment) and the number of children in the family, $Family\ Size_i$, is instrumented with *Twins* through our First Stage equation:

$$Family\ Size_i = \pi_0 + \pi_1 Twins_i + \pi_2 X_i + \pi_3 P_i + v_i \qquad [3]$$

Preliminarily, we estimate on the whole sample of families (with at least one child) using simply the number of children in the family as endogenous variable and the dummy *Twins in Family* as an instrument.[16]

As regards the First Stage (Table 4, Panel B), we find that twin births increase the average number of children of 0.70-0.95, according to specifications. The *F*-statistics for the null hypothesis that the coefficient on the instrument in the First Stage regression is zero take on values in the range 55-79, confirming that weak instruments (Staiger and Stock, 1997) are not a concern for our IV analysis.

In the second stage (Table 4, Panel A), we find that an additional child in the family reduces the probability of attending the Lyceum of about 9-11 percentage points (*t*-stats are typically around 3). In the first specification we control only for Female, Age and Mother's and Father's Education; in the second specification we include survey time dummies; in the third specification we also control for mother's and father's age; in the fourth specification we include dummies for mother's and father's employment status; in the fifth specification we also include geographical area dummies and city size dummies. The effect of family size on educational achievement is remarkably stable across specifications.

[16] An instrument based on twin births might be compromised by the proliferation of "In Vitro Fertilization" (IVF), a treatment for infertility, which tends to increase twin birth rate. However, assisted conception is a recent phenomenon: In Vitro Fertilization in Italy was very rare before 2000 and a Law governing medically assisted procreation was introduced in Italy only in 2004 (Law 40/2004). The Law has been rather restrictive in the use of IVF. The individuals in our sample are all born before 1997. Furthermore, the annual rate of multiple births in our sample (until 1997) is almost constant. Finally, in our main analysis we use twin births that occur at the second or third births while fertility treatments are often used in relation to the first birth.

Table 4. Two-Stage Least Squares Estimates. Educational Achievements and Family Size

	(1)	(2)	(3)	(4)	(5)
Panel A: Second Stage					
Family Size	-0.091***	-0.092***	-0.107***	-0.108***	-0.115***
	(0.031)	(0.030)	(0.039)	(0.039)	(0.040)
Female	0.143***	0.143***	0.142***	0.141***	0.140***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Age	0.010***	0.010***	0.012***	0.011***	0.011***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Mother's Education	0.029***	0.029***	0.028***	0.026***	0.026***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Father's Education	0.029***	0.029***	0.028***	0.023***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Birth Order			0.035	0.036	0.035
			(0.028)	(0.028)	(0.027)
Mother's Age			0.005***	0.005**	0.005***
			(0.002)	(0.002)	(0.002)
Father's Age			-0.000	-0.001	-0.001
			(0.002)	(0.002)	(0.002)
Survey Time dummies	NO	YES	YES	YES	YES
Father's and mother's employment dummies	NO	NO	NO	YES	YES
Geographical Areas and City Size dummies	NO	NO	NO	NO	YES
Observations	8136	8136	8107	8107	8107
Panel B: First Stage					
Twins in Family	0.943***	0.951***	0.732***	0.724***	0.703***
	(0.108)	(0.107)	(0.094)	(0.095)	(0.095)
F-stat	76.90	79.18	60.32	58.35	55.13
Adjusted R^2	0.044	0.046	0.276	0.287	0.315

Notes: The dependent variable is *Lyceum*. Standard errors, corrected for heteroskedasticity and allowed for clustering at household level, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

However, the estimates in Table 4 could still suffer from some residual bias since the instrument could be correlated with the error term in the first equation. In particular, twins could be characterized by some differences with respect to single birth children, the probability of having multiple births increases with the number of births, and the decision to have additional children after twins could be related to some other unobservable factors.

In order to avoid any possible correlation between the instrument *Twins* and the error term in the first equation, following Black, Devereux and Salvanes (2005), we carefully define two specific samples and modify the definition of the instrument.

The first sample is defined for families with two or more births ("*At Least Two Births*") that at the second birth might or might not have had twins. We define the instrument *Twins (Second Birth)* equal to 1 if the second birth is a multiple birth and 0 if the second birth is a single child. In this way we are excluding families with only one child or families with twins at the first birth.

Comparing families with two or more births (in which therefore the exogenous variation in the number of children is given by whether or not multiple births occurred at the second birth) ensures that these families do not differ in terms of desired family size.

Furthermore, in this sample we consider only first born children, that is, children born before the second birth. We exclude later born children because the decision to have additional children might be correlated to other factors. More importantly, we exclude twins that tend to have different characteristics with respect to other children: in fact, twins are often born prematurely, have lower birthweight, tend to suffer of some health problems, and so on: these characteristics could affect directly their long-term outcomes.

Using these criteria our sample reduces to 3,479 observations. We report in Table 5 estimates of First and Second Stage of various specifications.

Considering the First Stage in Panel B, we find that the arrival of twins at the second birth increases the number of children in the family of about 0.65. This means that for most of the families the effective number of children has been increased by twins, while other families (approximately one third) planned to have more children also in the absence of twins. The *F*-statistics in the First Stage regression take on values in the range 158-269, well above the threshold of 10.

Table 5. Two-Stage Least Squares Estimates. The impact of the number of children on educational achievement. Sample: First born children in families with at least two births.

			Panel A: Second Stage		
	(1)	(2)	(3)	(4)	(5)
Family Size	-0.169**	-0.170**	-0.180**	-0.182**	-0.201**
	(0.077)	(0.076)	(0.075)	(0.074)	(0.085)
Female	0.136***	0.135***	0.136***	0.136***	0.135***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Age	0.016***	0.017***	0.016***	0.016***	0.016***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Mother's Education	0.030***	0.030***	0.029***	0.029***	0.030***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Father's Education	0.031***	0.031***	0.030***	0.025***	0.025***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mother's Age			0.006**	0.005*	0.006**
			(0.003)	(0.003)	(0.003)
Father's Age			-0.002	-0.002	-0.002
			(0.002)	(0.002)	(0.002)
Survey Time dummies	NO	YES	YES	YES	YES
Father’s and mother’s employment dummies	NO	NO	NO	YES	YES
Geographical Areas and City Size dummies	NO	NO	NO	NO	YES
Observations	3479	3479	3458	3458	3458
			Panel B: First Stage		
	(1)	(2)	(3)	(4)	(5)
Twins (Second Birth)	0.644***	0.653***	0.672***	0.672***	0.623***
	(0.040)	(0.041)	(0.041)	(0.048)	(0.050)
Observations	3479	3479	3458	3458	3458
Adjusted R^2	0.021	0.022	0.035	0.054	0.078
F-statistics	255.21	248.84	269.42	194.43	158.03

Notes: Sample of first-born child in families with two or more children; twins are excluded. The dependent variable is *Lyceum*. *Twins (Second Birth*) is set to 1 only if multiple births occurred at the second birth (and 0 otherwise). Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

Considering the Second Stage in Panel A of Table 5, in column (1) we control for gender, age, and parents' years of education. We find that an additional child in the family reduces of about 17 p.p. the probability to go to a Lyceum (*t*-stat=–2.20). In the second and third specification we include as additional controls survey time dummies and mother's and father's age. We find again that the probability of attending a Lyceum decreases of about 17 p.p. in larger families. In the fourth specification we include dummies for mother's and father's employment status and in the last specification we also include geographical area dummies and city size dummies. The effect of our interest does not change much when we include these additional control variables.

As regards the impact of control variables, we find that females attend a Lyceum with a higher probability of about 13 p.p.; this probability increases strongly – of about 3 p.p. – for each additional year of father's and mother's education and increases a little (+0.6 p.p.) when mothers are older.

The second sample we build along the lines of Black, Devereux and Salvanes (2005) is analogous to the first but consider families with at least three births that at the third birth may, or may not, have multiple births. *Twins (Third Birth)* is set to one if a multiple birth occurs at the third birth (and 0 otherwise). For this sample, for the reasons explained above, the estimates are run only for first and second born children, who are not twins, giving a final sample of about 1,900 observations.

We report the estimates in Table 6. In the First Stage (Panel B) we find that the birth of twins increases the number of children of about 0.96, much higher than in the previous Table, implying that rarely families desire to increase furtherly the number of children after twins at the third birth.

We estimate the same specifications of Table 5 and we find very similar results: the probability of attending a Lyceum for first and second born children decreases of about 20 p.p. when the effective number of children happens to be increased because of a multiple birth.

As regards the other variables, the effects are very similar to the ones we have found in the previous Table. In addition, in this sample we are able to control for birth order and we find that later born children tend to be slightly less educated, but the coefficient is imprecisely estimated.

All in all, in both samples we find clear negative effects of family size on educational attainment of children.[17] The uncovered impact of an additional child on the probability of attending a Lyceum is in the range 16-20 p.p. Considering that the effect of Lyceum on College Degree is estimated around 50%, one can infer that an additional child has an impact on the probability of attaining a Degree of about 8-10 percentage points.

[17] Black, Devereux and Salvanes (2005) use also the sample of families with at least 4 births, but in our case the number of observations is too low to make reliable estimates.

Table 6. Two-Stage Least Squares Estimates. Number of Children and Educational Achievement. Sample: First Two Children in Families with at Least Three Births.

	Panel A: Second stage				
	(1)	(2)	(3)	(4)	(5)
Family Size	-0.202**	-0.204**	-0.210**	-0.207**	-0.206**
	(0.096)	(0.093)	(0.088)	(0.085)	(0.085)
Female	0.154***	0.152***	0.153***	0.156***	0.154***
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
Age	0.012*	0.013*	0.013*	0.012*	0.013*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Mother's Education	0.029***	0.028***	0.027***	0.023***	0.025***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
Father's Education	0.026***	0.026***	0.026***	0.024***	0.024***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Birth Order			-0.021	-0.021	-0.025
			(0.020)	(0.020)	(0.020)
Mother's Age			0.010***	0.010***	0.009***
			(0.003)	(0.003)	(0.003)
Father's Age			-0.004	-0.005	-0.004
			(0.003)	(0.003)	(0.003)
Survey Time dummies	NO	YES	YES	YES	YES
Father's and mother's employment dummies	NO	NO	NO	YES	YES
Geographical Areas and City Size dummies	NO	NO	NO	NO	YES
Observations	1906	1906	1896	1896	1896
	Panel B: First Stage				
	(1)	(2)	(3)	(4)	(5)
Twins (Third Birth)	0.966***	0.955***	0.964***	0.973***	0.983***
	(0.147)	(0.138)	(0.140)	(0.132)	(0.122)
Observations	1906	1906	1896	1896	1896
Adjusted R^2	0.043	0.049	0.054	0.075	0.094
F-statistics	42.99	47.89	47.25	54.31	64.77

Notes: Sample of first and second born children in families with three or more children; twins are excluded. The dependent variable is *Lyceum*. *Twins (Third Birth)* is set to 1 only if multiple births occurred at the third birth (and 0 otherwise). Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

6. Heterogeneous Effects of Family Size

In this Section we investigate if the effects of family size on educational achievement are different according to a number of children's or parents' characteristics.

We estimate specification (5) of Table 5 on a number of subsamples and we use only the sample with at least two births to avoid to deal with too small subsamples. Results are reported in Table 7.

First of all, we consider the different impact on males and females in columns (1) and (2), respectively. We find that the effect of family size is very strong for females (–28 p.p.) while is small and not statistically significant for males (–8.4 p.p.). It seems that parents try to protect their male children from negative effects of family size, but do not care very much for females.

Considering the differences among geographical areas, in columns (3) and (4) we find a strong effect for Southern regions but very small effects for Center and North.

Then we consider parents' education, distinguishing between low educated parents (if the average years of education is 8 or below) (in column 5) or highly educated (if average years of education is higher than 9) (in column 6). We find that in terms of magnitude the effect does not differ according to the parental educational level.

Table 7. TSLS Estimates. Heterogeneous Effects of Family Size

	(1) Males	(2) Females	(3) North	(4) South	(5) Low ed.	(6) High ed.
Family Size	-0.084	-0.286***	-0.004	-0.307***	-0.177***	-0.166
	(0.124)	(0.073)	(0.152)	(0.071)	(0.036)	(0.224)
Observations	1845	1613	1822	1636	1751	1707

Notes: Sample of first-born child in families with two or more children; twins are excluded. The dependent variable is *Lyceum*. The instrumental variable is *Twins (Second Birth)*. Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

Furthermore, in Table 8 we investigate if a different impact is found according to household income, distinguishing between low income families (below the median income) and high income families (above the median). We find a strong effect for low income families (–27 p.p.) and a smaller effect (–16 p.p.) for high income families.

In column (3) and (4) of Table 8 we also consider spacing of the first born child with respect the second born child or children. Our estimates show that when spacing is below the median (4 years) the effect is strong (–27 p.p.), while the effect is much smaller (–13 p.p.) (although negative) when spacing is greater than 4.

These two findings – partially in line with the analysis of Grawe (2008) for the UK – suggest that the difficulties arising from a larger family size are deriving both from financial resources and from parental time constraints.

Table 8. TSLS Estimates. Heterogeneous Effects of Family Size

	(1) Low Inc.	(2) High Inc.	(3) Spacing<=4	(4) Spacing>4
Family Size	-0.274***	-0.160	-0.268**	-0.133
	(0.083)	(0.128)	(0.124)	(0.094)
Observations	1725	1733	1892	1566

Notes: Sample of first-born child in families with two or more children; twins are excluded. The dependent variable is *Lyceum*. The instrumental variable is *Twins (Second Birth)*. Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

7. Robustness Checks

In this Section we run a number of robustness checks to verify if our results are driven by some of the criteria that we adopted or by the definition of the dependent variable or by the dataset we use.

Our main analyses are based on a specific sample in terms of age (19-22) since we aim to include individuals old enough to have completed secondary school but young enough that they do not have chosen to leave home.

We now investigate if our results change when we slightly modify the age range. With the selection criteria adopted in Table 5 (families with at least two births) and the instrumental variable "Twins (Second Birth)", we use the following age range: 19-21, 19-23, 19-24, 19-25, 20-22, 20-23, 20-24. The respective estimates are reported in the columns of Table 9. We find very similar results: changing the age range does not change much the estimated effect that remains around 16-17 percentage points.

Table 9. TSLS Estimates. Alternative Samples in terms of Age

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	19-21	19-23	19-24	19-25	20-22	20-23	20-24
Family Size	-0.163**	-0.160**	-0.176**	-0.156**	-0.172**	-0.160**	-0.176**
	(0.079)	(0.067)	(0.069)	(0.068)	(0.077)	(0.067)	(0.069)
Observations	2581	4402	5321	6177	3480	4402	5321

Notes: Sample of first-born child in families with two or more children; twins are excluded. The dependent variable is *Lyceum*. The instrumental variable is *Twins (Second Birth)*. Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

As a further robustness check we use an alternative dataset, the Italian Health Conditions Survey that – in contrast to SHIW – contains information on current attendance of individuals at university or school. Specifically, the dataset we use is the latest available wave (conducted between 2012 and 2013) of the Survey "Italian Health Conditions and Use of Health Services" provided by the Italian National Statistical Office (ISTAT). This survey is conducted on a nationally representative sample of 49,811 households for a total of 119,073 individuals and collects a wide range of information on individual demographic and socio-economic characteristics (in addition to health conditions).

We are able to build the same variables of the previous analyses and in addition we exploit the question "Enrolled at School or University" in which individuals are asked if they are currently attending a school, a university, etc. We build the variable "College" equal to one if individual i has already attained a College Degree or he/she answered that is currently enrolled in a Degree Course at University (and zero otherwise).

We use the restricted sample of "Families with at least two births; only first born; twins excluded" using as an instrument "Twins (Second Birth)". The number of observations is about 2,700. Our findings reported in Table 10 are similar to previous analyses. First of all, we find that a multiple birth occurring at the second birth increases the number of children of about 0.8 (t-stats are always above 14) (Panel B).

As regards the Second Stage in Panel A we find that an additional child decreases the probability of attending university of siblings of about 18 percentage points (significant at the 10 percent level).

Table 10. Probability of Attending University and Family Size. IV Estimates with Data from "Italian Health Conditions and Use of Health Services"

	Panel A: Second Stage					
	(1)	(2)	(3)	(4)	(5)	(6)
Family Size	-0.205**	-0.185*	-0.183*	-0.175*	-0.173*	-0.185*
	(0.080)	(0.097)	(0.095)	(0.098)	(0.099)	(0.100)
Female	0.116***	0.126***	0.126***	0.126***	0.127***	0.129***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Age	0.038***	0.044***	0.044***	0.045***	0.045***	0.046***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Mother's Education		0.025***	0.026***	0.025***	0.024***	0.021***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Father's Education		0.034***	0.034***	0.032***	0.031***	0.029***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mother's Age				0.008**	0.007*	0.006
				(0.004)	(0.004)	(0.004)
Father's Age				0.001	0.001	0.004*
				(0.002)	(0.002)	(0.002)
Immigrant					-0.094**	-0.075*
					(0.044)	(0.044)
Observations	3179	2735	2735	2735	2735	2735
	Panel B: First Stage					
	(1)	(2)	(3)	(4)	(5)	(6)
Twins (Second Birth)	0.865***	0.807***	0.814***	0.790***	0.786***	0.785***
	(0.055)	(0.049)	(0.051)	(0.056)	(0.052)	(0.051)
Adjusted R^2	0.026	0.031	0.031	0.084	0.090	0.093

Notes: The dependent variable is *College*. Restricted Sample: families with at least two births; first born children; twins are excluded. Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

8. Dealing with the Sample Selection Bias

As shown in Section 3 and in Table A3 in the Appendix, a small fraction of individuals (around 3 percent) is leaving their family in the age range 19-22 and therefore they cannot be observed in our sample. We have also noticed that the probability of leaving home is inversely related to Lyceum, which is our dependent variable and this could lead to a sample selection bias.

In this Section we carry out a simple Monte Carlo simulation to investigate whether the small sample selection bias due to individuals leaving their home with a reduced probability if they attended a Lyceum could drive our results.

In our Monte Carlo Simulation we set the sample size to *N*=1,000. We randomly generate the number of children in a family and – following the theoretical trade-off between quantity and quality and the evidence shown above – we determine the probability of attending a Lyceum as an inverse function of the number of children:

$$Lyceum = I(0.5 - 0.20 * Family\ Size + \ \varepsilon > 0) \quad [4]$$

where $I(.)$ is an indicator function equal to one if the condition is true and 0 otherwise; ε is a random variable with a Normal distribution and zero mean.

After generating, as described, the variables *Family Size* and *Lyceum* we estimate by OLS the following simple regression on the whole sample:

$$Lyceum = \beta_0 + \beta_1 Family\ Size + u \qquad [5]$$

where β_1 represents the true effect, without sample selection bias. We run 10,000 replications. We find that, on average, $\beta_{1\text{True}}$=–0.184, SD=0.014. The 95% confidence interval is (–0.238; –0.125).

Subsequently, we assume that the probability of exiting from the sample is negatively related to Lyceum (exaggerating the magnitude of the effect to make more evident its consequences):

$$Exit = I(0.1 - 0.2 * Lyceum + \nu > 0) \qquad [6]$$

where ν is a random variable with a Normal distribution and zero mean.

Then, we estimate equation [5] on the sample affected by selection, that is, only on individuals who did not exit, running again 10,000 replications. We find that on average $\beta_{1\text{Biased}}$=–0.143; SD=0.018, with a 95% confidence interval equal to (–0.216; –0.082). Clearly, it emerges that the magnitude of the estimated coefficient is biased towards zero when sample selection is at work.

The histogram with the distributions of the estimated coefficients generated in our simulations is plotted in Figure 2. In blue the estimated coefficients on the whole sample, in red the estimated coefficients affected by a sample selection bias. The distribution of coefficients with sample selection is clearly shifted towards zero. Our Monte Carlo simulation suggests that since with real data we are estimating on a selected sample, our coefficient of interest tends to be biased towards zero, and thus we are estimating a sort of lower bound.

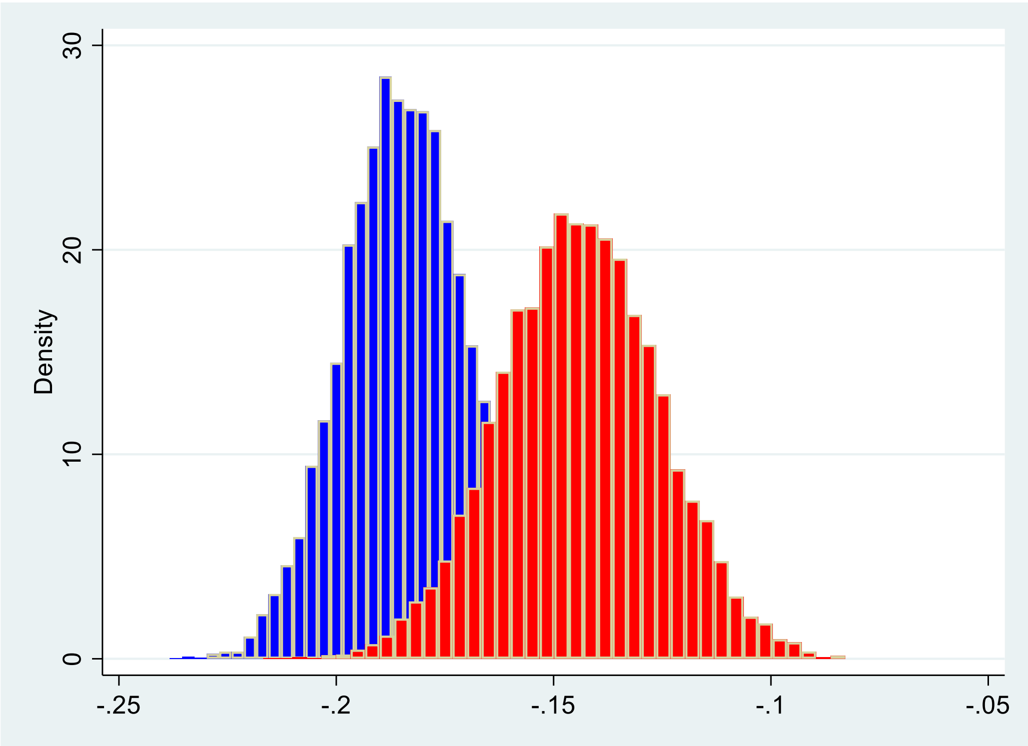

Figure 2. Monte Carlo Simulation: Coefficients from the whole sample in blue; Coefficients from a selected sample in red.

9. Concluding Remarks

Early investments in human capital have been shown to be crucial for children's outcome but could be negatively affected by family size, in that a higher number of siblings tends to dilute parental time and economic resources invested for each child.

The related children quantity/quality trade-off proposed by Becker has been investigated only for a few countries mainly for the difficulty in linking data on family of origin with children educational achievement.

In this paper we have tackled this problem using many waves of the Survey on Household Income and Wealth of Bank of Italy, and focusing on the educational achievement of children aged 19-22 years old that still live in their family of origin. Since individuals at this age have not yet completed their education, we have used the academic generalist secondary school track (Lyceum) as a measure of educational achievement, showing preliminarily with several data sources that Lyceum is a very strong predictor of University attainment.

Since family size cannot be considered exogenous but it is related to observable and unobservable family characteristics, to overcome endogeneity problems we have used multiple births as an instrumental variable affecting the number of children in a family.

Following the literature we have carefully defined the sample used to avoid any correlation between the instrument and the error term: we have used only samples with the possibility of multiple births at the second or third birth, excluding twins from the analysis and considering only children born before the considered birth level (only first born and only first and second born children, respectively).

In partial contrast with findings from other developed countries, we find a strong negative effect of family size on children's educational achievement. An additional child in the family reduces, for children born earlier, the probability of attending a Lyceum of about 16-20 percentage points. Given the relationship between Lyceum and College, the probability of attaining a College Degree is presumably reduced of about 8-10 percentage points when an additional child is born in a family.

In our analysis we have also shown that our findings are robust to a number of checks and the effects are stronger for daughters, for low income families and when spacing between births is limited.

An important determinant of the negative impact of family size on children's education is likely related to the fact that Italy is a country with a poor system of family assistance for families with children and childcare services are heavily undersupplied. According to ISTAT, only 13% of children aged 0-2 obtained a slot in the public child care system in 2013 (see Del Boca, Pronzato, Sorrenti, 2016).

References

Agarwal, L., Brunello, G., and Rocco, L. (2019). The pathways to college, *IZA Discussion Paper*.

Andini, M., Ciani, E., de Blasio, G., D'Ignazio, A., and Salvestrini, V. (2018). Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior & Organization*, 156, 86-102.

Angrist, J., and Evans, W., "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 1998, 450-477.

Angrist, J., Lavy, V., and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, *28*(4), 773-824.

Åslund, O., and Grönqvist, H. (2010). Family size and child outcomes: Is there really no trade-off?. *Labour Economics*, 17(1), 130-139.

Becker, G. (1960). "An Economic Analysis of Fertility." in *Demographic and Economic Change in Developed Countries,* edited by G.S. Becker. Princeton, NJ: Princeton University Press.

Becker, G. and H.G. Lewis. 1973. "On the Interaction Between the Quantity and Quality of Children." *Journal of Political Economy* 81(2):S279–S288.

Becker, G.S. and N. Tomes. 1976. "Child Endowments and the Quantity and Quality of Children." *Journal of Political Economy* 84(4): S143–S162.

Becker, S.O., Cinnirella, F. and Woessmann, L. (2010) The trade-off between fertility and education: Evidence from before the demographic transition. *Journal of Economic Growth* 15(3): 177–204.

Björklund, A., T. Eriksson, M. Jäntti, R. Oddbjørn and E. Österbacka (2004), "Family Structure and Labor Market Success: The Influence of Siblings and Birth Order on the Earnings of Young Adults in Norway, Finland, and Sweden", in *Generational Income Mobility in North America and Europe*, Miles Corak ed., Cambridge University Press.

Black, S., Devereux, P. and Salvanes, K. (2005). The more the merrier? The effect of family composition on children's education. *Quarterly Journal of Economics* 120:669–700.

Blake, J., (1989), *Family Size and Achievement*, Berkeley: University of California Press.

Bougma, M., LeGrand, T. and Kobian´e, J.-F. Fertility decline and child schooling in urban settings of Burkina Faso. *Demography* 52(1): 281–313.

Bronars, S., and Grogger, J., (1994) "The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment," *American Economic Review,* pp. 1141-1156.

Caceres-Delpiano, J. (2006). The impacts of family size on investment in child quality. *Journal of Human Resources* 41:722–37.

Conley, D. and Glauber, R. (2006) Parental educational investment and children's academic risk: Estimates of the impact of sibship size and birth order from exogenous variation in fertility. *The Journal of Human Resources* 41(4): 722–737.

Cunha, F., and J. Heckman. 2007. "The Technology of Skill Formation." *American Economic Review*, 97 (2): 31-47.

Del Boca, D., Pronzato, C., & Sorrenti, G. (2016). When rationing plays a role: Selection criteria in the Italian early childcare system. CESifo Economic Studies, 62(4), 752-775

Fitzsimons, E. and Malde, B. (2014) Empirically probing the quantity-quality model. *Journal of Population Economics* 27(1): 33–68.

Grawe, N. D. (2008). The quality–quantity trade-off in fertility across parent earnings levels: a test for credit market failure. Review of Economics of the Household, 6(1), 29-45.

Guiso, L., Sapienza, P., Zingales, L., 2004. The role of social capital in financial development, *American Economic Review* 94 (3), 526–556.

Hanushek, E., "The Trade-off between Child Quantity and Quality," *Journal of Political Economy* 100(1), 1992, pp. 84-117.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782), 1900-1902.

Holmlund, B. (1988), "Sibling Position and Achievement: The Case of Sweden", *Scandinavian Population Studies,* vol. 7, pp. 100–107.

Hotz, V.J., Mullin, C.H. and Sanders, S.G. (1997) Bounding causal effects using data from a contaminated natural experiment: Analysing the effects of teenage childbearing. *Review of Economic Studies* 64(4): 575–603.

Jappelli, T., and Pistaferri, L. (2020). "Reported MPC and Unobserved Heterogeneity." *American Economic Journal: Economic Policy*, 12 (4): 275-97.
Kumar, S. and Kugler, A. (2011) Testing the children quantity-quality trade-off in India. MPRA Paper 42487, University Library of Munich, Germany.
Lee, J. (2008) Sibling size and investment in children’s education: An Asian instrument. *Journal of Population Economics* 21(4): 855–875.
Leibowitz, A. (1974). Home investments in children. *Journal of Political Economy*, *82*(2, Part 2), S111-S131.
Li, H., Zhang, J., and Zhu, Y. (2008). The quantity-quality trade-off of children in a developing country: Identification using Chinese twins. *Demography*, 45(1), 223-243.
Miller, A.R. (2009) Motherhood delay and the human capital of the next generation. *American Economic Review* 99(2): 154–158. Effects of the joint custody law in Italy
OECD (2019), *Education at a Glance 2019*.
Ponczek, V., and Souza, A. P. (2012). New evidence of the causal effect of family size on child quality in a developing country. *Journal of Human Resources*, 47(1), 64-106.
Rosenzweig, M., and Wolpin, K. (1980). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica*, 48:227–40.
Rosenzweig, M., and Zhang. J, (2009). Do population control policies induce more human capital investment? Twins, birth weight, and China’s “one child” policy. *Review of Economic Studies* 76:1149–74.
Schultz, T. P., 2005, “Effects of Fertility Decline on Family Well-Being: Evaluation of Population Programs,” Draft for MacArthur Foundation Consultation meeting, Philadelphia, PA, 2005.

Appendix

Table A1. Type of High School and University Enrolment. Linear Probability Model

	(1)	(2)	(3)	(4)
Lyceum	0.521***	0.506***		
	(0.005)	(0.005)		
Scientific Lyceum			0.716***	0.707***
			(0.006)	(0.006)
Classic Lyceum			0.741***	0.728***
			(0.008)	(0.008)
Languages Lyceum			0.653***	0.625***
			(0.012)	(0.013)
Technical School			0.244***	0.247***
			(0.008)	(0.008)
Teaching Institute			0.544***	0.518***
			(0.008)	(0.009)
Artistic High School			0.208***	0.200***
			(0.011)	(0.011)
Reference Category:			Professional School	Professional School
Controls	NO	YES	NO	YES
Observations	26235	26205	26098	26074
Adjusted R^2	0.191	0.217	0.312	0.322

Notes: ISTAT 2015 Survey of High School Graduates. The dependent variable is *University Enrolment*, a dummy equal to one if an individual is currently enrolled at the university or he/she has a College Degree. In columns (2) and (4) we control for Female, Immigrant, and dummies of Region of Residence. Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

Table A2. The Probability of Attaining a College Degree and Lyceum. Linear Probability Model

	(1)	(2)	(3)	(4)
Lyceum	0.404***	0.379***	0.378***	0.377***
	(0.005)	(0.005)	(0.005)	(0.005)
Female		-0.007***	-0.007***	-0.007***
		(0.001)	(0.001)	(0.001)
Birth Year		-0.072***		
		(0.008)		
Birth Year Squared		0.000***		
		(0.000)		
Constant	0.074***	68.007***	0.007***	0.007***
	(0.001)	(8.221)	(0.001)	(0.001)
Observations	166980	166980	166980	166980
Adjusted R^2	0.116	0.130	0.131	0.132

Notes: SHIW Dataset. Sample: All individuals aged 26 or more. The dependent variable is College Degree. OLS estimates. Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

Table A3. The probability to leave home with respect to Age and Lyceum. Linear Probability Model. The dependent variable is *Own Family*

	(1)	(2)	(3)	(4)	(5)
Age 19-22	0.032***	0.032***	0.053***	0.054***	0.033***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)
Age 23-26	0.162***	0.163***	0.150***	0.150***	0.178***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Age 27-30	0.424***	0.423***	0.369***	0.368***	0.448***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)
Age 31-34	0.686***	0.684***	0.640***	0.638***	0.703***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)
Age 35-38	0.833***	0.828***	0.813***	0.811***	0.846***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)
Female		0.106***	0.112***	0.113***	0.109***
		(0.002)	(0.003)	(0.003)	(0.002)
Lyceum			-0.088***	-0.089***	
			(0.004)	(0.004)	
Observations	84511	84511	59060	59060	84511
Adjusted R^2	0.429	0.441	0.429	0.433	0.451

Notes: SHIW Dataset. Sample: all individuals aged 16 or more. The dependent variable is *Own Family*. Standard errors, corrected for heteroskedasticity, are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.