HOW FAMILY SIZE AFFECTS CHILDREN'S EDUCATIONAL OUTCOMES: EVIDENCE FOR RUSSIA

by

Ievgeniia Iarmosh

A thesis submitted in partial fulfillment of
the requirements for the degree of

MA in Economic Analysis

Kyiv School of Economics

2013

Thesis Supervisor: _____ Professor Tom Coupé

Approved by _____
Head of the KSE Defense Committee, Professor Wolfram Schrettl

_____

_____

_____

Date _____

Kyiv School of Economics

Abstract

HOW FAMILY SIZE AFFECTS CHILDREN'S EDUCATIONAL OUTCOMES: EVIDENCE FOR RUSSIA

by Ievgeniia Iarmosh

Thesis Supervisor: Professor Tom Coupé

This paper investigates the existence of the so-called quantity-quality tradeoff for children's educational outcomes for the case of Russia. Using parental time investments, regular school attendance and school grades as a dependent variable we find different results for Russia compared to those obtained in the majority of other studies: ceteris paribus each additional sibling decreases the time spent by parents with their children by about 23 minutes per week, does not affect the probability of attending a general school, decreases the probability of getting high grades and increases the probability of getting low grades. To address possible endogeneity of sibship size we use samesex dummy as an instrument: it is equal to one if the first two children in a family are of the same gender and zero otherwise. However, the instrument appears to be weak but we find that boy preferences instruments (two dummy variables for boys being the first and the second child in a family respectively and an interaction term for boys being the first two children in a family) do not show qualitatively different results. Therefore, the quantity-quality tradeoff seems present for children's educational outcomes in Russia.

TABLE OF CONTENTS

LIST OF FIGURES

Number *Page*

LIST OF TABLES

LIST OF TABLES – CONTINUED

Number *Page*

ACKNOWLEDGMENTS

First of all I wish to express my gratitude to Prof. Tom Coupé for his valuable comments, ideas and overall guidance throughout the whole process of writing my thesis. I am also grateful to all professors, who anyhow helped me with their comments and suggestions.

We thank the Russia Longitudinal Monitoring survey, RLMS-HSE, conducted by the National Research University Higher School of Economics and ZAO "Demoscope" together with Carolina Population Center, University of North Carolina at Chapel Hill and the Institute of Sociology RAS for making these data available.

I would like to thank all my relatives who supported me in one way or another. The last but not the least, I am extremely grateful to my dear husband for his enormous support, patience and for just being at my side during the logy days.

Chapter 1

INTRODUCTION

The educational process is an important part of human development. While there are many determinants of education, one factor still being controversial is family size (or, alternatively, as many researchers call it "sibship") size. Considering children's education in particular, family-related factors are often thought to be among the main determinants of educational outcomes. However, despite a lot of existing research works, the exact components of children's educational "production function" still remain unclear (Black, Devereux and Salvanes, 2005).

"Quality or quantity?" is one of the tradeoffs commonly faced by individuals while doing their decision making process. In particular, many economists believe that this tradeoff is also applicable to the above mentioned determinant of the children's education. In particular, they suppose that the number of children and their education constitutes a tradeoff between quality and quantity: the more the number of children, the less educated they are.

This thesis extends the existent research by studying the case of the Russian Federation. Nowadays Russia is experiencing a demographic crisis: population has been declining over several decades (World Population Review, 2012). Despite the fact that birth rates are not decreasing nowadays, high mortality rates (especially among males) still worsen the demographic situation. According to the estimates, if not stimulating birth rates these days, then in order to maintain the population size at the steady state the fertility rates have to be equal to 2.52 in 2015, 3.03 in 2020 and 3.41 births per one female in 2025 (Sinelnikov, 2010). Thus, the Russian government actively implements different policies aiming at increasing in fertility rates: one-time fixed payments are provided per each child

born. Also, in 2008 The Law on "Mother's Capital"[1] was approved, according to which families (and mothers in particular) are provided with additional assistance in different activities such as educational expenses, mortgage repayments, etc.

As can be seen from Figure A (Appendix A), the total fertility rates (births per woman) tend to increase in Russia since 1999. Since the results of the recent body of literature seem to vary from one country to another, it is not clear whether the number of siblings affects children's education in Russia or not. And if it turns out that sibship size indeed affects children's education, one could expect the average quantity and quality of education to decline as fertility rates grow (in case of Russia they are additionally stimulated by the government). Of course, governmental assistance relaxes the family budget constraints, which in its turn may increase the quality of children's education. However, it is not clear whether these funds will significantly affect the quality of education and which effect (i.e. positive effect of monetary assistance or negative effect of sibship size) will dominate if a negative relationship between sibship size and children's education is found.

It is also important to mention that the majority of research is done for developed countries, while Russia is a developing one. Also, Russian educational system differs from the one established in the countries, for which similar research was done. There are three stages of school education: 4 years of primary school, 5 years of secondary school and 2 years of high school, all of which are compulsory according to the Constitution of Russian Federation. Thus, the research results may differ for Russia.

The early research on the existence of the relationship between family size and children's education (Rozenzweig and Wolpin, 1980; Blake, 1981; Zajonc, 1976)

[1] http://www.materinski-capital.ru/

concludes that this relation is strong. There are many hypotheses explaining the nature of this relationship. The explanations often involve various economic, socio-psychological and genetic factors (Conley and Glauber, 2006). One hypothesis interprets the relationship in the way that each additional child in a family places additional financial constraints on their parents and thus fewer resources are invested in the education of each child. This so-called "resource dilution" hypothesis was developed by Anastasi (1956) and is supported by many early studies. Later, Steelman and Mercy (1980) add that this hypothesis is the strongest for poor people whose financial resources are more binding. Another hypothesis is called a "confluence model" according to which the negative relationship between sibship size and children's education is explained by psychological atmosphere within the family: the ones with many children or children born with relatively short time spaces have rather unfavorable intellectual climate since it is children but not adults who dominate in such families (Zajonc, 1976). However, as Conley and Glauber (2006) notice, according to this theory what really affects children's education is not the number of children, but the age distribution within each family.

However, later findings seem to be more mixed: while some researchers still support the hypothesis about the existence of negative correlation (Lee, 2008; Booth and Kee, 2009), others state that this relationship is spurious (Guo and WanWey, 1999; Conley and Glauber, 2006) and family size does not affect children's educational outcomes (Guo and VanWey, 1999; Black, Devereux and Salvanes, 2005; Angrist, Lavy and Schlosser, 2005; Conley and Glauber, 2006).

The theoretical model explaining quality-quantity tradeoff is a QQ (Quality-Quantity) model by Becker and Lewis (1973). As for the empirical model, OLS, probit and ordered probit models will be used for three dependent variables: parental time investments, general school attendance dummy (which is equal to

one if a child attends regular school) and school grades. In order to eliminate possible endogeneity we will use the samesex instrumental dummy variable (it is equal to one if the first two children are of the same gender).

The data used in the analysis are taken from the Russia Longitudinal Monitoring Survey. The sample size of about 3,500 individuals was obtained by complex merge of a Child, Adult and Household Surveys from the last wave of RLMS, which was conducted in 2011. Because of limitations of the instrumental variable, the sample size was reduced to about 1000 children, who has at least one sibling and who currently attends school.

The results reveal that there is a negative relationship between sibship size and children's educational outcomes. Other things being equal, each additional sibling reduces weekly parental time investments by about 23 minutes, increases the probability of getting low grades and decreases the probability of getting high grades. Also, sibship size does not affect the probability of attending general school. After implementing samesex instrument the effect of sibship size became insignificant. However, the first-stage results show that this instrument is weak. Thus, we cannot rely on the estimates, where sibship size was instrumented by the samesex dummy.

The thesis is organized in the following way: Chapter 2 presents the detailed literature review, Chapter 3 describes methodology, Chapter 4 discusses the dataset used in the analysis, Chapter 5 describes the results and Chapter 6 concludes the findings.

Chapter 2

LITERATURE REVIEW

The literature review consists of several parts. The first one includes a short discussion on the determinants of education. The second part discusses the papers in which the relationship between children's education and family composition is examined. Finally, the third one considers the papers which support the strength of the sibship gender instrumental variables (this instrument is needed in order to eliminate endogeneity of the family size, which will be discussed later in the methodology section).

There is substantial research conducted on the determinants of children's educational outcomes. The important determinants are parental income and parental level of education (Master, 1969; Bratti, 2002; Davis-Kean, 2005; Magnuson, 2007; Løken, 2010). For example, Stanley Master (1969) using 1960 Census data finds that children with low-educated or poor parents are 20% more likely to drop out from secondary school. Katrine V. Løken (2010) uses the Norwegian oil shock of 1970 as an instrument for studying the relationship between education of Norwegian children born in 1967-1969 and family income and confirms the positive relationship between family income and children's education. Davis-Kean (2005) examines how parental education and earnings affect children's education through parent's beliefs and attitudes and concludes that the observed factors are significant determinants in studying educational outcomes. Some researchers, in particular, study the effect of maternal education on children's outcomes (Magnuson, 2007; Behrman and Rosenzweig, 2002; Plug, 2002) and find out that the maternal level of education affects children's education much more than the paternal one (Behrman and Rosenzweig, 2002; Plug, 2002).

Many papers also emphasize the influence of various neighborhood and family characteristics on children's education (Case and Katz, 1991; Cherlin et al., 1991; Aaronson, 1998; Painter and Levine, 2000). Other researchers focus on various characteristics of schools and classrooms children study at. According to some findings (Angrist and Lavy, 1999; Case and Deaton, 1999) the larger the class size is, the worse level of education children have.

Unfortunately, I did not find any papers discussing the determinants of children's education which would be specific to Russia. The majority of research done for Russia discusses education in the context of future returns to the labor market, health and mortality (especially for males) outcomes, smoking and alcoholism.

Also the family size is a widely discussed and controversial determinant of children's educational outcomes. As can be seen, the majority of earlier papers report a negative correlation between the "quality and quantity" of children in a family while later ones often contradict those results.

Rozenzweig and Wolpin (1980) study the effect of family size on children's "quality" using data for India. They use twin births and sex of the first child as an instrument for family size and make a conclusion that higher fertility leads to lower child quality. However, there are only 25 observations for families with twins, thus the results cannot be considered as reliable due to the small sample of twins and impreciseness of estimates. Blake (1981), using the "dilution model" also reports a negative correlation between the number of siblings and chances to attend college. The results found by Zajonc (1976) are almost the same: the higher is the number of siblings, the lower are the NMSQT scores of students. Additionally he concludes that there is a penalty for children who are born later: their grades are significantly lower than grades of children who are among the first in their family. Zajonc (1976) explains this relationship by the adverse

psychological atmosphere and inferior intellectual climate within the family with a high number of children.

However, more recent papers challenge these conclusions. For example, Guo and VanWey (1999) use the NLSY (National Longitudinal Survey of Youth) dataset and change models meaning in order to re-estimate the dependence between quantity and quality of children. The authors say that "…change models allow us to control for such un-measured effects as family intellectual climate, family value system, and family genetic heritage" (Guo and VanWey, 1999). After controlling for sibship and year fixed effects they find no relationship between sibship size and quality of children's education.

Black, Devereux and Salvanes (2005) find no relationship between family size and children's educational attainment after controlling for the birth order. Using the data for the whole Norwegian population, the researchers find that family size becomes insignificant when they control for the birth order of each child. They also check the robustness of their results by using twin births as an instrument for the sibship size. The authors additionally run a robustness check using the same-sex dummy for the first two children in the family and find surprising results, according to which an increase in the family size corresponds to higher educational attainments of children. However, Black, Devereux and Salvanes (2005) decide to pay little heed to these results considering them as non-credible (since they expected to obtain a negative but not positive correlation between family size and children's education, which would be consistent with the quantity-quality model) and mention that this issue needs further research.

Almost at the same time, Conley and Glauber (2006) argue that using twins as an instrument has a severe drawback: the birth of twins can affect the family in some other way since it is a very unusual phenomenon. For example, some additional

budget constraints can be imposed on parents, who expected to have only one child, and thus the birth of twins can potentially correlate with educational attainment. The authors also comment and examine further the abnormal effect obtained as a result of using the same-sex instrument for the family size: the effect will not be the same for American children since educational systems and family support policies of Norway and the USA are different. Conley and Glauber (2006) use the 1990 five percent PUMS (Public Use Microdata Sample) dataset and include to their analysis only mothers who have at least 2 children. The dependent variables used by them are whether children attend private school and "the likelihood of being back to school". Eventually the authors conclude (controlling for the birth order) that there is a negative relationship between family size and the probability of private school attendance and a positive relationship between the former variable and the probability of being held back to school.

Conley and Glauber (2006) also argue that their results differ from the ones obtained by Guo and VanWey (1999), because the dependent variables used in their papers are different: while Guo and VanVey use score reports as a measure of education quality and achievements, Conley and Glauber use two different measures in order to estimate both investments in education (private school attendance is taken as a dependent variable) and educational achievements (i.e when the probability of being held back in school is taken as a dependent variable).

Angrist, Lavy and Schlosser (2005) perform a similar analysis using the Israeli Census data. In addition to using twin birth and mixed-sex sibship composition, they account for ethnic variations in preferences over specific child gender. The dependent variables are: wages, fertility and completed education. The authors find no evidence of a "quantity-quality tradeoff'. They additionally report that

according to their estimates girls who were born earlier (compared to their siblings) tend to get married at younger age.

According to the article of Julio Cáceres-Delpiano (2006), twin births reduce the probability of children born later to attend private school, mother's participation in a labor force and the probability of divorce. The effect of sibship size on children's grades is ambiguous and not clear. The results of budget reallocations done by parents after the increase in family size are consistent with Becker's Quantity and Quality model.

However, despite the fact that many recent research works confirm no relationship between family size and education of children, some recent findings still confirm the early results. Jungmin Lee (2008) focuses on son preference (according to which fertility timing and family size are affected by the gender of the first child). Using the Korean dataset, he finds out that the quality-quantity tradeoff is not as big as if cross-sectional analysis was performed instead. But larger quantity of children still negatively affects investments in children's educational attainments. The effect is especially strong for families where high fertility rates are dominating. Booth and Kee (2009) use British Household Panel Survey in their analysis. They construct a new birth order index (which is cleaned from the effect of family size) in order to compare time and educational investments, which each of the siblings receives, and they turn out to be decreasing for each subsequent child in a family. Thus, the authors conclude that birth order negatively effects educational outcomes and despite all the cleaning procedures and robustness checks, which they perform in the analysis, sibship size still negatively affects children's education.

The previously found negative relationship between family size and quality of children's education could possibly be spurious: there are many other factors that

affect children's achievements and do not depend on the sibship size. Since early articles used extensively OLS methodology of estimation, some uncontrolled factors (for example, parental IQ level) could lead to biased estimates (Guo and WanWey, 1999; Conley and Glauber, 2006). Joseph Price (2008), using the American Time Use Survey, concludes that ceteris paribus, every day the first-born child is paid 20 to 30 minutes of parental attention more than the second-born one. He claims that parent-quality time spent with a child becomes lower when children grow up. This result, in addition to the previous explanation, can account for the negative effect of birth order on sibship outcomes.

Now let us consider the research which studies the relationship between family size and sibship sex composition in order to strengthen the fact that this relationship is strong enough and will serve as a good base for endogeneity elimination in family size. Various papers provide empirical evidence on the relationship between genders of the first two children and family size. One of the first demographic research which studies interrelation between family size and sex composition of children within the family was done by Thomas (1951). Her sample consists of second year medical students from the University of Glasgow in 1947. Tomas finds out that genders of the first two children significantly affect the size of the family: families tend to be larger if first two children are of the same gender (compared to the variant when there are two children of different sexes in the family).

Subsequent research done for other countries and other samples (Freedman et al, 1960; Elder Jr. and Bowerman, 1963; Sheps, 1963) corroborates and expands the results obtained by Thomas. Later, Gray (1972) performs similar analysis on families of students who studied at Western Kentucky University in 1971. She concludes that if the first two children are of the same gender then the family is 6% more likely to have additional children. Then, Jacobsen, Møller and

Engholm (1999) publish the paper in which they take a large dataset consisting of 363 373 Danish families with about 613 900 children. They also discover "strong preference for a balanced composition of sexes in Danish families" (Jacobsen, Møller and Engholm, 1999). They consider families with two or more children and find out the same results as in the previous research. The authors also make additional conclusions of the lowest fertility in households where the first two children have the same gender while the third child has the opposite gender (e.g. if the first two children are boys and the third child is a girl or vice versa). Additionally, families with two boys being the first children turn out to be more fertile than families where the first two children are girls.

Our research conducted for Russia will enrich the existent literature in several ways, as it is not only extrapolation of results for another country. First of all, since the relation between educational outcomes and sibship size is not clear and unambiguous for all countries, the research will either support or contradict each of the two diametrically opposite conclusions obtained previously by the researchers. Also, the majority of research is conducted for developed countries while patterns for developing countries may differ. Additionally, the educational system of Russia differs a lot from the one of Norway, Israel or the USA: all children are obliged to study eleven years at the secondary school since September 2007, while for other countries this is not obligatory. Potentially it may influence studying incentives and our variables of interest such as parental time investment, school grades and some others.

Concluding and summarizing the literature review, it can be said that there are many determinants of children's educational outcomes. The number of siblings is a rather controversial determinant since the results obtained by all (or at least by almost all) research done so far are different. At the early stages of their research the authors found a negative relationship between the number of children and

quality/quantity of their education (Rozenzweig and Wolpin, 1980; Blake, 1981; Zajonc, 1976). However, more recent findings provide a mixed picture. While some of them claim that there is no relationship between sibship size and educational outcomes (Guo and VanWey, 1999; Black, Devereux and Salvanes, 2005; Angrist, Lavy and Schlosser, 2005; Conley and Glauber, 2006), others find this relationship ambiguous and unclear (Cáceres-Delpiano, 2006). Some researchers still continue to find that this relation is not spurious and does exist (Lee, 2008; Booth and Kee, 2009).

Unlike the ambiguity in the relation between educational outcomes and sibship size, all papers considered in the third part of the literature review corroborate the relation between sibship sex composition and final family size (as was previously mentioned, families with first two children of the same sex are more likely to expand than families with first two children having different sexes). The research papers were conducted for different countries such as the United Kingdom (Thomas, 1951), The United States of America (Freedman et al, 1960; Elder Jr. and Bowerman, 1963; Sheps, 1963; Gray, 1972) and Denmark (Jacobsen, Møller and Engholm, 1999). Thus, there is significant empirical background showing that the relationship between family size and sibship sex composition may serve as a good instrument for eliminating endogeneity in family size in case of Russia as well.

Chapter 3

METHODOLOGY

Theoretical Model

There is a theoretical three commodity QQ (Quality/Quantity) model developed by Becker and Lewis (1973). It is designed in such a way that allows to explain negative relationship between quality and quantity of children in a family. A short description of this model can be found in the article written by Black, Devereux and Salvanes (2005).

According to this model every family has its own preferences and it solves the following utility maximization problem:

$$max U(N, Q, S)\ s.t. I = P_N N + P_Q Q + P_S S + \Pi N Q \qquad (1)$$

Here I is the total family budget, N is the number of children, Q – quality of a child, S – the so-called composite commodity. Consequently, P_N, P_Q and P_S are prices for the number of children, their quality (per one child) and for the composite commodity. As can be seen, budget constraint is not linear because of the last term, which is comprised of quantity of children multiplied by Q, their quality, and Π, which represents "price-weighted sum of the cost minimizing levels of quality inputs required to increase the quality of one child by one unit" (Black, Devereux and Salvanes, 2005). Π in non-negative. However, it is assumed to be more than zero, otherwise (if it is equal to zero) the budget constraint becomes linear and QQ model becomes a simple Q model (the model without the interaction term ΠNQ, which brings non-linearity and a tradeoff to the budget constraint).

Then, Rozenzweig and Wolpin (1980) calculate the derivative of Q with respect to N (consequences of exogenous increment in quantity of children):

$$\frac{\partial Q}{\partial N} = \frac{\varphi_{12} - \Pi P_s^2 \lambda}{\varphi_{11}} \tag{2}$$

Here λ is positive (marginal utility of income), φ_{12} and φ_{11} are figures from the Hessian matrix for the model without interaction term. Since Hessian is negative definite, we obtain that denominator is strictly positive, while φ_{12} may be either positive (if Q and N are complements) or negative (if Q and N are substitutes). The second term in the numerator is strictly positive as well. Thus, the sign of the derivative depends only on whether Q and N are complements or substitutes. Obviously, it is not likely that Q and N are complements. Then, if the derivative of Q with respect to N tends to zero, then Π tends to zero as well and the non-linear interaction term ΠNQ disappears from the budget constraint. Hence, quantity-quality tradeoff will not be supported by the budget constraint and empirical analysis as well (Black, Devereux and Salvanes, 2005).

Empirical Model

The empirical model used in the analysis implies running regressions of the following form:

$$ED_i = \beta_0 + \beta_1 SIBSIZE_i + \beta_2 X_i + \epsilon_i \tag{3}$$

Here *SIBSIZE* actually represents the number of siblings in a household since only the number of siblings but not the number of parents fostering children (i.e. only mother/father or both of them) can affect education of a child due to our assumption. Presence of either mother or father or both, which potentially affects

children's quality as well, will be included later to the vector *X* of the explanatory variables.

X is a vector of determinants of child's education such as sex, nationality, dummy for a person born in a city versus a village, mother's and father's levels of education, overall parental income (it is assumed that maternal and paternal income do not differ a lot and make equal contribution to the child's education), two dummy variables which reflect the presence of a mother and a father in a family, birth order and others.

ED is a dependent variable reflecting child's quality of education. In this case three proxies of educational outcomes/educational investments will be used in the analysis.

The first one is the amount of time (minutes per week) parents spend with their children while doing different activities (e.g. read, do home assignments, learn some additional material, etc.). It's obtained by summing up all the variables indicated in the question 7.7.1[2] (surveys are publicly available and can be found by following the link in the footnote). For this dependent variable standard OLS regression will be run since we treat time spent with children as a continuous variable.

The second dependent variable is a dummy variable for the type of school a child attends (modified question 3.2 from the Child Survey, see footnote 2: dummy variable equals one for an answer falling into category 5 – regular school). It equals one if a child attends general school and zero if she studies in gymnasium, lyceum or other specialized educational institution. For this dependent variable, the probit regression model will be used.

[2] http://www.cpc.unc.edu/projects/rlms-hse/data/questionnaires/rtchild.pdf

The third dependent variable is represented by categorical school grades (question 3.4 of the Child Survey, see footnote 2). The order of grades is reversed so that the highest category now corresponds to the highest grade. That is, now the meaning of categories is the following: 1 – "Basically all the three and often the two"; 2 – "Basically all the three"; 3 – "Basically all the four and three"; 4 – "Basically all the four"; 5 – "Basically all the five and the four"; 6 – "Almost all grades are five". For this dependent variable ordered probit model will be used.

It is important to say that all the above mentioned regressions will be run with clustered standard errors across family units in order to control for unobservable factors inherent to each particular household.

However, there is a widespread concern about possible endogeneity of family size in the recent literature. For example, the number of children in a family may be correlated with some unobservable variables like parental IQ level: children's educational outcomes are certainly affected by their mother's and father's intellectual abilities; also, parental IQ may be correlated with the number of children in a family, which creates endogeneity and bias of the obtained estimates. Thus, the following instrument will be used at the first stage in order to correct for potential endogeneity of a sibship size:

$$SIBSIZE_i = \pi_0 + \pi_1 SAMESEX_i + \pi_2 X_i + \varepsilon_i \qquad (4)$$

Here *SAMESEX* is the discussed above instrument for the *SIBSIZE* variable: it is equal to one if the first two children in a household are of the same sex and zero otherwise. It is obtained by the following formula (Angrist, Lavy and Schlosser, 2005):

$$SAMESEX = BOY1ST \cdot BOY2ND + (1 - BOY1ST)(1 - BOY2ND) \quad (5)$$

Here *BOY*1*ST* and *BOY*2*ND* are dummies for a boy being the first and the second child in a family, respectively. Intuitively, parents tend to diversify gender composition of their children (that is, they want to have both a boy and a girl). Therefore, they are more likely to have another child if their first two children are either two boys or two girls. Otherwise, if they have already two children who are a boy and a girl, then they are less likely to have additional child since gender composition of their children is already diverse.

There may be several econometric problems in our analysis. The first problem is a potential bias of estimates. It is rooted in the specifics of the dataset used and will be discussed in the data description section. The second problem is a weak instrument. Despite the fact that many studies use samesex dummy as an instrument for sibship size, for Russia it may not explain the number of children in a family by a proper degree. Provided that samesex dummy appears to be a weak instrument we will try using boy preference as another instrument: *BOY*1*ST*, *BOY*2*ND* and *TWOBOYS* (which equals a product of *BOY*1ST and *BOY*2*ND*). If boy preferences will be a weak instrument as well, the third problem, the endogeneity problem, arises: obtained estimates will be biased by unobservables not included into our analysis.

Chapter 4

DATA DESCRIPTION

The dataset used in the analysis is Russia Longitudinal Monitoring Survey of Higher School of Economics (RLMS-HSE) conducted by the National Research University Higher School of Economics and ZAO "Demoscope" together with Carolina Population Center, University of North Carolina at Chapel Hill and the Institute of Sociology RAS. RLMS consists of 20 waves; the first one was conducted in 1992 and the last one in 2011. The last 16 waves were conducted by Carolina Population Center at the University of North Carolina at Chapel Hill and by the Demoscope team in Russia.

There are several surveys in each round to be used in this thesis: a child survey that allows to trace schooling history; an adult survey, that gives the data for almost all control variables, and a household survey. The last one indicates the relationship between household members. Therefore, it is possible to "connect" children with their siblings and parents.

We use the last cross-section of 2011 only due to several reasons. First, the question regarding school progress (grades) is being asked since 2010 only. Thus, we have only two waves with data available for school grades and if we want to compare the results on the same data, we won't be able to analyze all but the last two phases of the survey. Second, the number of observations increased more than 1.5 times between 2010 and 2011 waves of a survey (around 2100 children in 2010 versus 3500 in 2011; 10,000 adults in 2010 versus 17,000 in 2011). Thus, even if our sample was appended by the wave of 2010 it is not likely that the final sample size would be significantly more than the sample of 2011 (since there is a lot of "newcomers" in 2011; also, there is a certain attrition rate from one year to

another). Third, it is not likely that the number of children changed a lot from 2010 to 2011. However, if we want to obtain reliable results we need to have enough variability in both dependent and explanatory variable of interest (that is, sibship size). Otherwise, it is not rational to construct a panel data for two years, which is the case for our analysis. Therefore, only the last wave of RLMS, which was conducted in 2011, will be taken for the analysis.

The merging process was applied as follows: first, the Household survey was analyzed, where connections mother-child and father-child were identified within each family. These connections were written in two separate files. After that, parents were connected to their children by merging two datasets with mothers and fathers being placed separately. After deletion of mother-grandmother and father-grandfather relationships from the dataset (since it's also the modification of mother-child and father-child relation) the additional information about parents and children (using Adult and Child surveys) were added to the dataset. Then, the data was cleaned, specific variables as birth order and same-sex IV were generated. After that the dataset was ready for the further analysis. If more extensive description of data merging procedure is needed, please, contact the author for the details.

According to the CPC information, the definition of a child in the survey is "a person who is under 19 for purposes of informed consent". However, the Child survey was conducted only for those children who are younger than 14 years and an adult in the household answers all the questions on behalf of the child. That is, Adult questionnaire is used for those children who are over 14.

Thus, the dataset has several drawbacks. First and the most important is that not all children of school age will be analyzed, but only those who are under 14. Since the age of 14 corresponds to a person studying at the 8^{th} class, children studying

at the 9th through 11th classes will not be taken into account. Therefore, we cannot extrapolate the obtained results onto all children of a school age as respondents are not randomly taken from the 7-18 years interval. The sample is randomly drawn from children of 7-14 years. Hence, we can conclude that the obtained results can be extrapolated onto Russian children studying at primary and secondary schools.

The second drawback is the following: one of the parents usually answers the Child survey questions on behalf of his/her child. However, she can give the wrong information since children may lie to their parents in order not to be punished, especially if their scores are very low. Or, alternatively, parents are inclined to overestimate their children's achievements. Thus, the reported scores will be biased upward and the amplitude of the misrepresentation will be increasing as a child's grades are getting lower.

Third, not all children studying in primary and secondary schools will be analyzed. Due to specifics of the merging procedures the children who live without both parents (for instance, the ones living with grandparents or an aunt/uncle) will not be included into the final sample.

Fourth, only those children who live in the household will be analyzed. For example, there might be other children born previously who grew up and moved to the other place. This will add some errors, especially for the same-sex IV construction process. Also, the number of children and birth order may be calculated incorrectly. It might be one of the reasons for weak instruments.

The list and full description of variables used in the final dataset is represented in the Appendix B. Father's and mother's highest educational degrees (variables *f_heducl* and *m_heducl*) were re-shaped such that the following categories were created: 0 – "Parent is absent"; 1 – "Primary or incomplete secondary school"; 2

– "Complete secondary school"; 3 – "Vocational training school or professional courses"; 4 – "Technical community college, various training schools (medical, art, etc.)"; 5 – "Institute, university, academy including master's degree"; 6 – "Post-graduate courses, academic degree".

The tabulation across the number of children in the family is represented in Table 1a (tabulation across individuals) and Table 1b (tabulation across households). Obviously, since the distribution is made across children in the merged sample, the minimum and the maximum number of children in the family are 1 and 7, respectively (that is, there are no families with zero children in the sample due to specifics of the merging procedure).

Table 1a. Number of children in the family (across individuals)

# of children	Freq.	Percent	Cum.
1	1,278	37.41	37.41
2	1,492	43.68	81.09
3	436	12.76	93.85
4	111	3.25	97.1
5	68	1.99	99.09
6	25	0.73	99.82
7	6	0.18	100
Total	3,416	100	

However, later our sample will be restricted to those children who come from families with two or more children because otherwise the same-sex IV could not be constructed just because there is no second child in the family. That is, we will have to drop 1,278 individuals (or 1,278 households, which is the same) from our sample.

Table 1b. Number of children in a family (across families)

# of children	Freq.	Percent	Cum.
1	1,278	50.16	50.16
2	983	38.58	88.74
3	212	8.32	97.06
4	47	1.84	98.9
5	21	0.82	99.73
6	6	0.24	99.96
7	1	0.04	100
Total	2,548	100	

Then, for the purposes of our analysis we restrict the sample to those children who currently attend school. Summary statistics for the main variables (across individuals) is represented in the Table 2a.

Table 2a. Summary statistics across individuals: Full sample

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
b_order	1,575	1.463	0.696	1	6
childnum	1,575	1.943	0.973	1	7
sibsize	1,575	0.943	0.973	0	6
female	1,575	0.504	0.500	0	1
gen_school	1,575	0.781	0.414	0	1
schpayments	1,562	113.9	720.5	0	16,000
timeinv	1,534	204.3	295.2	0	3,300
itprgres	1,503	4.364	1.108	1	6

The meaning of variables: *b_order* – birth order of a child; *childnum* - # of children; *sibsize* – sibship size; *female* – child's gender dummy; *gen_school* – equals 1 if a child attends general school; *schpayments* – monthly school payments (RUR); *timeinv* – parental time investments (min/week); *itprgres* – school grades.

Here we can note several points. There is almost equal distribution of children across gender in the dataset: 50.4% females versus 49.6% of males. 78.1% of children attend general school. The average amount of minutes spent by parents

with their children constitutes 204 minutes (or approximately 3.5 hours) per week. Average monthly payments for school constitute 113.9 rubles (3.5 USD). The average progress of children is 4.3, which means that the most frequent answers were "Basically all the five and the four" (category 5) and "Basically all the four" (category 4). The tabulation for the school progress variable can be found in the Appendix C (Table C1).

Table 2b – Summary Statistics Across Families: Full Sample

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
mother	1,412	0.986	0.118	0	1
father	1,412	0.771	0.421	0	1
m_age	1,392	35.38	5.569	24	56
f_age	1,088	37.60	6.411	19	65
childnum	1,412	1.827	0.864	1	7
sibsize	1,412	0.827	0.864	0	6
boy1st	1,412	0.530	0.499	0	1
sett	1,412	1.661	0.911	1	3
hthord	1,405	1.131	0.392	1	3
tincm_nt	1,389	41,212	31,693	0	300,000
m_heducl	1,412	3.663	1.245	0	6
f_heducl	1,412	2.691	1.764	0	6

The meaning of variables: *mother*/*father* – 1 if mother/father is present in a family; *m_age*/*f_age* – mother's/father's age, years; *childnum* - # of children; *sibsize* – sibship size; *boy1st* – 1 if the first child is a boy; *sett* – settlement type (1 – urban, 2 – PGT, 3 - rural); *hthord* – living conditions (1 – own residence, 2 – rented residence, 3 - dormitory); *tincm_nt* – nominal monthly income of the family (RUR); *m_heducl*/*f_heducl* – mother's/father's highest educational level.

Summary statistics for the full sample across families is represented in the Table 2b. Here it can be seen that on average mothers are younger than fathers (35 versus 37 years). Also, the number of families where mother is present is larger compared to the number of families with father being present (98.6% versus 77.1% of families, respectively). It can be seen from the means of *mother* and *father* dummy variables. It truly reflects the reality since in Russia children

usually live with their mothers rather than fathers after their parents' divorce. The average total monthly income of the family is 41,212 rubles (1,283 USD). Also it can be seen that on average mothers are more educated than fathers.

Table 3. Tabulation across the number of children within a family

Summary statistics: mean
by categories of: childnum (# of children)

childnum	1	2	3	4	≥ 5
timeinv	225.161	195.4339	214.9211	158.1818	104.1304
itprgres	4.460501	4.391796	4.198925	4.066667	3.804348
tincm_nt	36348.79	41597.63	51237.24	52857.22	47603.72
gen_school	0.768946	0.77285	0.812183	0.911111	0.791667
schpayments	130.381	93.83806	192.2755	12.72727	0
mother	0.977819	0.991936	1	0.933333	0.979167
m_age	35.28166	35.40108	34.80203	36.33333	35.38298
m_heducl	3.752311	3.721774	3.345178	2.555556	2.3125
father	0.676525	0.823925	0.857868	0.822222	0.708333
f_age	37.35792	37.66232	37.1716	40.24324	38.70588
f_heducl	2.377079	2.932796	2.77665	2.711111	1.6875
female	0.513863	0.489247	0.502538	0.466667	0.666667
sett	1.515712	1.680108	1.913706	2.666667	2.354167
samesex	.	0.489247	0.568528	0.577778	0.625

The meaning of variables: *timeinv* – parental time investments (min/week); *itprgres* – school grades; *tincm_nt* – nominal monthly income of the family (RUR); *gen_school* – equals 1 if a child attends general school; *schpayments* – monthly school payments (RUR); *mother/father* – 1 if mother/father is present in a family; *m_age/f_age* – mother's/father's age, years; *m_heducl/f_heducl* – mother's/father's highest educational level; *female* – child's gender dummy; *sett* – settlement type (1 – urban, 2 – PGT, 3 - rural); *samesex* – 1 if first two children are of the same gender.

Table 3 represents means of main variables tabulated by the number of children in a family. Here it can be clearly seen that as the number of children increases average time investments per each child decrease, school grades become worse and the probability of attending general school increases.

However, there may be other factors influencing educational outcomes, which change as the number of children changes. For example, we can note that on average parental levels of education are worse for families with more children (variables *m_heducl* and *f_heducl*). Or, that the share of individuals living in PGT and rural areas (variable *sett*) is bigger for households with more children. Therefore, while a simple look at the tabulated dataset suggests that the quantity-quality educational tradeoff should be present for Russian children, the situation may be more complicated than it seems at first glance.

For the subsequent analysis we need to restrict our sample to those children who are currently studying at school (which was done before) and come from the families with two or more children. This restriction enables us to use the samesex instrumental variable since for families with one child it simply does not exist. Summary statistics across individuals and families is represented in the Tables 4a and 4b, respectively.

Table 4a. Summary statistics across individuals: Restricted sample

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
b_order	1,034	1.706	0.752	1	6
childnum	1,034	2.436	0.857	2	7
sibsize	1,034	1.436	0.857	1	6
female	1,034	0.499	0.500	0	1
samesex	1,034	0.515	0.500	0	1
gen_school	1,034	0.787	0.409	0	1
schpayments	1,024	105.3	663.3	0	16,000
timeinv	1,006	193.3	291.6	0	3,300
itprgres	984	4.313	1.107	1	6

The meaning of variables: *b_order* – birth order of a child; *childnum* - # of children; *sibsize* – sibship size; *female* – child's gender dummy; *samesex* – 1 if first two children are of the same gender; *gen_school* – equals 1 if a child attends general school; *schpayments* – monthly school payments (RUR); *timeinv* – parental time investments (min/week); *itprgres* – school grades.

It can be seen from the tables that summary statistics has not changed a lot from dropping families with one child: the distribution of male and female children in the sample is almost equal (50.1% versus 49.9%, respectively); 78.7% of children attend general school; average monthly school payments constitute 105.3 rubles (3.28 USD); average weekly time investment is 193 minutes (roughly 3 hours); average school grades are between "almost all the four" and "almost all the four and the five". The distribution of grades for the restricted sample can be found in the Appendix C (Table C2). Visually it has not changed a lot as well.

Table 4b – Summary Statistics Across Families: Restricted Sample

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	mean	sd	min	max
mother	871	0.991	0.0955	0	1
father	871	0.829	0.377	0	1
m_age	863	35.43	5.288	24	56
f_age	722	37.72	6.259	19	65
childnum	871	2.341	0.722	2	7
sibsize	871	1.341	0.722	1	6
boy1st	871	0.558	0.497	0	1
boy2nd	871	0.485	0.500	0	1
samesex	871	0.506	0.500	0	1
sett	871	1.751	0.938	1	3
hthord	869	1.117	0.369	1	3
tincm_nt	854	44,258	34,809	24.49	300,000
m_heducl	871	3.607	1.265	0	6
f_heducl	871	2.885	1.665	0	6

The meaning of variables: *mother/father* – 1 if mother/father is present in a family; *m_age/f_age* – mother's/father's age, years; *childnum* - # of children; *sibsize* – sibship size; *boy1st* – 1 if the first child is a boy; *boy2nd* – 1 if the second child is a boy; *samesex* – 1 if first two children are of the same gender; *sett* – settlement type (1 – urban, 2 – PGT, 3 - rural); *hthord* – living conditions (1 – own residence, 2 – rented residence, 3 - dormitory); *tincm_nt* – nominal monthly income of the family (RUR); *m_heducl/f_heducl* – mother's/father's highest educational level.

According to the Table 4b mother is present in 99.1% of families while father is present in only 82.9% of families; mothers are on average 2 years younger and

more educated than fathers. Also the distribution of the samesex dummy is almost equal and the average monthly income of a household constitutes 44,258 rubles (1378 USD).

Chapter 5

RESULTS

We are to consider results for three dependent variables: parental time investments, general school attendance dummy and grades. As was previously mentioned, clustered standard errors across families were used in each regression in order to allow for fixed family-specific characteristics. It is also important to say that robustness check (running the same regressions with the same sample, available upon request) was made for all the results listed in this chapter. However, these results were qualitatively the same and that's why, the decision of keeping more observations and increasing preciseness of estimates was made.

First let's consider the results obtained while using the time spent by parents with their child as a dependent variable (expressed in minutes per week). The results are represented in the Table 5. Besides the sibship size variable (which is our main variable of interest), the following variables were included in the initial regression: presence of a mother and a father in a family (it is likely that full families devote more time to their children), parental age (as parents are getting older they spend less time with their children), birth order (e.g. if a child is the third one in the family then the birth order variable equals 3 for her; this variable is expected either to be insignificant or negatively affect the time spent with each child: parents are older when later-born children study, thus they are paid less attention), rural settlement type dummy (usually people living in rural areas pay less attention to studying), dummy for families living in a dormitory (usually children living in a dormitory are paid less attention by their parents), total income of a household (richer families are less time constrained and thus have more spare time, which they can spend with their children), general school attendance dummy (those parents who decided that their child should study at a

specialized school invest more time in their children), school payments (those parents who pay for school value education more and thus tend to invest more time in their children), parental educational levels (more educated parents usually spend more time with their children). School payments variable was deliberately excluded from the regression due to multicollinearity issues (R^2 rises almost twice and all variables become insignificant after including this variable into the regression) and was not used in all further specifications as well. The dummy variables for full family, mother's age, parental educational levels, birth order and rural settlement type dummies were insignificant and R^2 did not change a lot from excluding them, thus they were step-by-step excluded from the regression.

Let us consider the final results represented in the Table 5. Here columns 2, 4, and 6 represent the same specifications as in the columns 1, 3 and 5 respectively, but additionally the samesex instrumental variable is applied. Such a low R^2 of about 0.03 is common for the majority of research without regard to the one made by Black, Devereux and Salvanes (2005) since they had the exceptionally large sample size of 1.5 mln observations (with R^2 being about 0.19 after applying the twin birth instrument and 0.04 before).

According to the columns 1, 3 and 5 of the Table 5, other things being equal, an increase in the number of siblings by one person decreases time spent with a child by about 23 minutes per week, which is economically significant as the average time investment for our sample constitutes only 3 hours per week. Father's age negatively affects time investment too: ceteris paribus, an increase in father's age by 5 years decreases time spent with a child by about 25 minutes. Also, the increase in total income positively affects time investments, which is logical since parents are less budget constrained and thus they can devote more time to their children. General school attendance and living in the dormitory decrease average parental time investments, which is logical as well.

Table 5. Effect of sibship size on parental time investments

VARIABLES	(1) timeinv	(2) timeinv (w/ IV)	(3) timeinv	(4) timeinv (w/ IV)	(5) timeinv	(6) timeinv (w/ IV)
f_age	-4.817***	-1.227	-4.773***	-1.235	-4.74***	-1.225
	(1.623)	(5.463)	(1.622)	(5.451)	(1.600)	(4.976)
sibsize	-23.43**	-605.3	-23.10**	-602.2	-21.01*	-542.7
	(10.80)	(597.4)	(10.84)	(603.9)	(10.83)	(502.2)
tincm_nt	0.000474*	0.00209	0.000455	0.00208		
	(0.00027)	(0.0018)	(0.000280)	(0.00184)		
gen_school	-68.65**	18.78	-68.40**	18.39	-76.47**	-18.99
	(34.75)	(104.4)	(34.76)	(104.8)	(35.69)	(78.64)
dormitory			-145.0***	-37.27	-15.6***	-84.20
			(21.72)	(173.1)	(20.16)	(131.7)
Constant	446.4***	996.1	446.2***	993.1	470.3***	1,034*
	(73.06)	(615.8)	(73.01)	(621.2)	(73.72)	(587.1)
Observations	817	817	817	817	829	829
R-squared	0.026		0.028		0.026	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *f_age* – father's age, years; *sibsize* – sibship size; *tincm_nt* – nominal monthly income of the family (RUR); *gen_school* – 1 if a child attends general school; *dormitory* – 1 if a family lives in a dormitory.

We can see the same pattern in all three specifications of the Table 5: negative and significant effect of sibship size disappears and becomes insignificant after applying the instrument. However, as can be seen from the Appendix D (where the first-stage results for the columns 2, 4 and 6 are shown), the applied instrument is weak and there are several simple signs confirming this fact. First, the coefficient estimate of the samesex IV is insignificant in the first stage regression. Second, R^2 of the first stage is too low (0.025 at most). Third, as was advised by Staiger and Stock (1997), F-statistics of the first stage regression should not be less than 10. In our case it even does not exceed 2. Boy preferences were tested as an instrumental variable too, but these instruments did not bring any significant changes to the results. Also, according to the endogeneity test we

have to reject H_0 of all independent variables being exogenous for columns 2, 4 and 6 with p-values of 0.0246, 0.0275 and 0.0284, respectively. Therefore, in our case using IV is not appropriate since it is weak and, as it was precisely mentioned by Bound, Jaeger and Baker (1993, 1995), "cure can be worse than the disease". The same is true for other two dependent variables (general school attendance and grades). Hence, thereafter we will concentrate more on estimates obtained without using the instruments.

Now, let's consider the second dependent variable *gen_school* (dummy which equals one if a child attends regular school) for which probit regression was run. The marginal effects are represented in the Table 6.

Apart from the sibship size, the following variables were initially included into the regression: presence of a mother and a father in a family (it is more likely that full family has more funds and can pay for a specialized school), birth order (due to our assumption children who born later are more likely to attend general school; alternatively, the birth order effect can be insignificant), rural settlement type dummy (there are less specialized schools in rural areas), dummy for families living in a dormitory (families which live in dormitories have less funds to pay for a specialized school), total income (children from richer families are less likely to attend general school), parental educational levels (more educated parents usually value education more and thus their children are less likely to study in general school). Birth order, presence of a mother in a family, dormitory dummy and mother's educational level appear to be insignificant and almost nothing changes after excluding them from the model. Actually, mother's and father's educational levels are expected to be correlated as usually both a husband and a wife are equally (or almost equally) educated. Therefore, these variables were excluded from the model and the results presented in the Table 6 were obtained.

Table 6. Marginal effects of sibship size on the probability of general school attendance

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
father	-0.432*	-0.431*	-0.426*	-0.416**	-0.402**	-0.384**
	(0.255)	(0.239)	(0.222)	(0.205)	(0.191)	(0.183)
sibsize	-0.0369	-0.0393	-0.0413	-0.0429	-0.0440	-0.0445
	(0.0325)	(0.0366)	(0.0398)	(0.0417)	(0.0422)	(0.0399)
rural	0.212***	0.237***	0.261***	0.284***	0.305***	0.325***
	(0.0339)	(0.0406)	(0.0556)	(0.0718)	(0.0855)	(0.0943)
tincm_nt	-1.01e-06**	-1.07e-06**	-1.13e-06**	-1.17e-06**	-1.20e-06**	-1.22e-06**
	(4.5e-07)	(4.6e-07)	(4.8e-07)	(5.0e-07)	(5.1e-07)	(5.3e-07)
_If_heducl_2	0.237***	0.266***	0.296***	0.325***	0.354**	0.380**
	(0.0614)	(0.0754)	(0.0980)	(0.124)	(0.150)	(0.174)
_If_heducl_3	0.235***	0.264***	0.293***	0.322***	0.349**	0.375**
	(0.0634)	(0.0765)	(0.0979)	(0.123)	(0.148)	(0.171)
_If_heducl_4	0.244***	0.275***	0.306***	0.338***	0.368***	0.397**
	(0.0549)	(0.0669)	(0.0893)	(0.116)	(0.143)	(0.167)
_If_heducl_5	0.172	0.189	0.206	0.222	0.235	0.247
	(0.126)	(0.145)	(0.166)	(0.188)	(0.208)	(0.227)
Observations	972	972	972	972	972	972

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *father* – 1 if father is present in a family; *sibsize* – sibship size; *rural* – 1 for families living in rural areas; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_2* through *_If_heducl_5* – dummies for father's higher educational level: *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Columns of the Table 6 represent marginal effects computed for different number of siblings in a family while rows represent marginal effects for each control variable. Besides sibship size, general school attendance is controlled for presence of a father in a family, dummy for rural residence area, total income of a family and father's higher level of education. From the Table 6 it can be found

out that increasing in magnitude marginal effect of each additional sibling decreases the probability of attending regular school. However, this effect is statistically insignificant. First-stage results are reported in the Appendix E, where we can see that samesex instrument appeared to be weak as well. After applying the instrumental variable all marginal effects become insignificant (the results are represented in the Appendix F). According to the Wald test the p-value is 0.4679, which means that at the 5% significance level we cannot reject the null hypothesis of all independent variables being exogenous. Therefore, there is no need in using samesex instrument for this dependent variable.

Table 7. Marginal effects of sibship size on school performance

	(1)	(2)	(3)	(4)	(5)	(6)
GRADES	1 sibling	2 siblings	3 siblings	4 siblings	5 siblings	6 siblings
mainly 3 & 2	0.00527**	0.00750*	0.0104*	0.0140	0.0184	0.0236
	(0.00229)	(0.00386)	(0.00614)	(0.00921)	(0.0131)	(0.0176)
mainly 3	0.0187***	0.0223***	0.0257**	0.0285**	0.0304***	0.0310***
	(0.00610)	(0.00837)	(0.0104)	(0.0117)	(0.0118)	(0.0103)
mainly 4 & 3	0.0351***	0.0329***	0.0287***	0.0226***	0.0150**	0.00625
	(0.0123)	(0.0109)	(0.00768)	(0.00404)	(0.00630)	(0.0125)
mainly 4	0.00460	-0.00204	-0.00853	-0.0144	-0.0192	-0.0226*
	(0.00380)	(0.00360)	(0.00690)	(0.00997)	(0.0118)	(0.0121)
mainly 5 & 4	-0.046***	-0.047***	-0.046***	-0.044***	-0.040***	-0.035***
	(0.0158)	(0.0164)	(0.0149)	(0.0120)	(0.00823)	(0.00451)
mainly 5	-0.017***	-0.012***	-0.0094***	-0.0067***	-0.0046***	-0.0031**
	(0.00651)	(0.00390)	(0.00215)	(0.00132)	(0.00121)	(0.00124)

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results for school performance taken as a dependent variable are shown in the Table 7 (ordered probit regression was run). Appendix G (Tables G1-G6) represents marginal effects for each of the outcome for the school grades dependent variable. Actually, Table 7 is a summary of Appendix G showing marginal effects for the *sibsize* variable only. Here rows represent different

outcomes of grades and columns show marginal effects for different values of *sibsize* variable.

Initially the following variables (in addition to the sibship size) were used in the model: dummies for presence of a mother and a father in a family (there is a possibility that children from full families study better), birth order (children born later either tend to study worse or the birth order effect is insignificant), gender dummy (usually girls are more diligent and thus study better than boys), settlement type dummies (high grades may be got easier in rural areas since less attention is paid to education there), dummy for families living in a dormitory (children living in dormitories have less resources and opportunities to study in a quiet place because all family lives in one room and therefore children tend to study worse), total income (in general it should increase the probability of getting good grades and decrease the probability of getting low grades, but the effect can be uncertain due to many reasons such as laziness, corruption, etc.), parental highest educational levels (children who live in more educated environment tend to study harder and get higher grades). Birth order, mother's highest educational level, rural area and living in a dormitory dummies are insignificant and were step-by-step excluded from the model. Thus, only the dummy variable for father being present, gender dummy, total family income and father's highest educational level variables were kept in the model.

We can note from the Table 7 that marginal effect of sibship size is almost always significant. Ceteris paribus, the probability of receiving low or even failing grades increases by 0.5-3% with each additional sibling and this effect increases in magnitude as the number of siblings rises. Also, other things being equal additional sibling decreases the probability of receiving high grades by 0.3-5% and this effect is decreasing in the amplitude as the number of siblings increases.

Appendix H represents first-stage results, which show that the instrument is weak in this case as well. Here the first stage F-statistics equals 6.92, which is still less than 10 and R^2 is 0.0584. Appendix J (Tables J1-J6), like Appendix G, represents marginal effects for each school progress outcome (using samesex instrumental variable). Appendix I, like the Table 7, is a summary of all marginal effects of sibship size represented in the Appendix J. Appendix I reveals that patterns of marginal effects remain the same while almost all of the marginal effects become insignificant after using the instrument. According to the Appendix H Wald test of exogeneity has a p-value of 0.642, which means that we cannot reject the null hypothesis of all independent variables being exogenous at the 5% significance level. Therefore, there is no need in using the samesex instrument (non-instrumented results are better) in this case.

Chapter 6

CONCLUSIONS

There are a lot of factors influencing children's educational outcomes. Sibship size is a rather controversial one since different studies provide us with mixed results. While a lot of research is conducted for developed western countries, this one tries to shed more light on the problem by examining Russia, which is a developing country significantly differing in its educational system.

Three dependent variables were used in the research: time investment of parents (that is, the number of minutes per week spent with their children while performing various activities), regular school attendance dummy and categorical school grades. In order to correct for possible endogeneity of a sibship size, the samesex instrumental variable was applied: it is equal to one if the first two children are of the same gender and zero otherwise.

Using RLMS-HSE dataset, Russia appeared to show the results different from those obtained for the majority of developed countries: despite the fact that negative influence of sibship size on school performance, parental time investments and the probability of attending general school disappeared after instrumenting sibship size with same-sex IV, all indicators showed that the instrument appeared to be weak. Boy preferences were tested as an instrument as well, but they did not show qualitatively different results. Thus, the instrumented results cannot be deemed as reliable estimates and it may be better to use the ones obtained by simple OLS, probit and ordered probit models. However, non-instrumented estimates may still suffer from the endogeneity bias. While better datasets and instruments are not available, it is the case when it would be more reasonable not to fix the problem.

According to the non-instrumented estimates each additional sibling decreases the average parental time investment by roughly 23 minutes per week. Also, ceteris paribus, sibship size does not affect the probability of attending general school, decreases the probability of getting high grades by 0.3-5% and increases the probability of getting low grades by 0.5-3%.

While the endogeneity problem is still not solved in some cases, searching for strong instruments is one of the main prospective developments of the thesis. Therefore, so far we can make a conclusion that the quantity-quality tradeoff in children's educational outcomes is present in Russia: the more the number of children in a family, the less time they are devoted and the more failing grades they get.

WORKS CITED

Aaronson, D. 1998. Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources*: 915-946.

Anastasi, A. 1956. Intelligence and family size. *Psychological Bulletin* 53(3): 187-209.

Angrist, J. D., and V. Lavy. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics* 114(2): 533-575.

Angrist, J. D., V. Lavy, and A. Schlosser. 2005. New evidence on the causal link between the quantity and quality of children (No. w11835). National Bureau of Economic Research.

Becker, G. S., and H. G. Lewis. 1973. On the Interaction between the Quantity and Quality of Children. *The Journal of Political Economy* 81(2): 279-288.

Behrman, J. R., and M. R. Rosenzweig. 2002. Does increasing women's schooling raise the schooling of the next generation? *American Economic Review* 92(1): 323-334.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2005. The More The Merrier? The Effect Of Family Size And Birth Order On Children's Education. *Quarterly Journal of Economics*, *120*(2, May): 669-700.

Blake, J. 1981. Family size and the quality of children. *Demography* 18(4): 421-442.

Booth, A. L., and H. J. Kee. 2009. Birth order matters: the effect of family size and birth order on educational attainment. *Journal of Population Economics* 22(2): 367-397.

Bound, John, David A. Jaeger, and Regina Baker. 1993. The Cure Can Be Worse than the Disease: A Cautionary Tale Regarding Instrumental Variables. NBER Technical Working Paper No. 137.

Bound, John, David A. Jaeger, and Regina Baker. 1995. Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak. *Journal of the American Statistical Association* 90(430): 443-450.

Bratti, M. 2002. Parents' Current Income, Long-Term Family Characteristics and Children's Educational Outcomes: Evidence from the 1970 British Cohort Study. *The University of Warwick Economic Research Paper*: 658.

Cáceres-Delpiano, J. 2006. The impacts of family size on investment in child quality. *Journal of Human Resources* 41(4): 738-754.

Case, A., and A. Deaton. 1999. School inputs and educational outcomes in South Africa. *The Quarterly Journal of Economics* 114(3): 1047-1084.

Case, A. C., and L. F. Katz. (1991). The company you keep: The effects of family and neighborhood on disadvantaged youths. National Bureau of Economic Research, Working Paper No. w3705.

Cherlin, A. J., F. F. Furstenberg, P. L. Chase-Lansdale, and K. E Kiernan. 1991. Longitudinal studies of effects of divorce on children in Great Britain and the United States. *Science* 252(5011): 1386-1389.

Conley, Dalton, and Rebecca Glauber. 2006. Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility. *Journal of Human Resources,* University of Wisconsin Press, Vol. 41(4).

Davis-Kean, P. E. 2005. The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment. *Journal of Family Psychology* 19(2): 294-304.

Elder Jr, G. H., and C. E. Bowerman. 1963. Family structure and child-rearing patterns: The effect of family size and sex composition. *American Sociological Review*, 891-905.

Freedman, D. S., R. Freedman, and P. K. Whelpton. 1960. Size of family and preference for children of each sex. *American Journal of Sociology*: 141-146.

Gray, E. 1972. Influence of sex of first two children on family size. *Journal of Heredity* 63(2): 91-92.

Guo, G., and L. K. VanWey. 1999. Sibship size and intellectual development: Is the relationship causal? *American Sociological Review*: 169-187.

Jacobsen, R., H. Møller, and G. Engholm. 1999. Fertility rates in Denmark in relation to the sexes of preceding children in the family. *Human Reproduction* 14(4): 1127-1130.

Lee, J. 2008. Sibling size and investment in children's education: An Asian instrument. *Journal of Population Economics* 21(4): 855-875.

Løken, Katrine V. 2010. Family income and children's education: Using the Norwegian oil boom as a natural experiment. *Labour Economics* 17(1): 118-129.

Magnuson, K. 2007. Maternal education and children's academic achievement during middle childhood. *Developmental Psychology* 43(6): 1497-1512.

Masters, S. H. 1969. The effect of family income on children's education: Some findings on inequality of opportunity. *Journal of Human Resources* 4(2): 158-175.

Painter, G., and D. I. Levine. 2000. Family structure and youths' outcomes: Which correlations are causal? *Journal of Human Resources* 35(3): 524-549.

Plug, E. 2002. How Do Parents Raise the Educational Attainment of Future Generations? *IZA Discussion Paper* 652: 1-13.

Price, J. 2008. Parent-Child Quality Time Does Birth Order Matter? *Journal of Human Resources* 43(1): 240-265.

Rosenzweig, M. R., and K. I. Wolpin. 1980. Testing the quantity-quality fertility model: the use of twins as a natural experiment. *Econometrica: journal of the Econometric Society*: 227-240.

Sheps, M. C. 1963. Effects on family size and sex ratio of preferences regarding the sex of children. *Population Studies* 17(1): 66-72.

Sinelnikov, Aleksandr. 2010. Mozhno li srazu ostanovit ubyl naselenija Rossii? *Demograficheskie Issledovanija* 7.

Staiger, Douglas, and James H. Stock. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica* 65: 557-586.

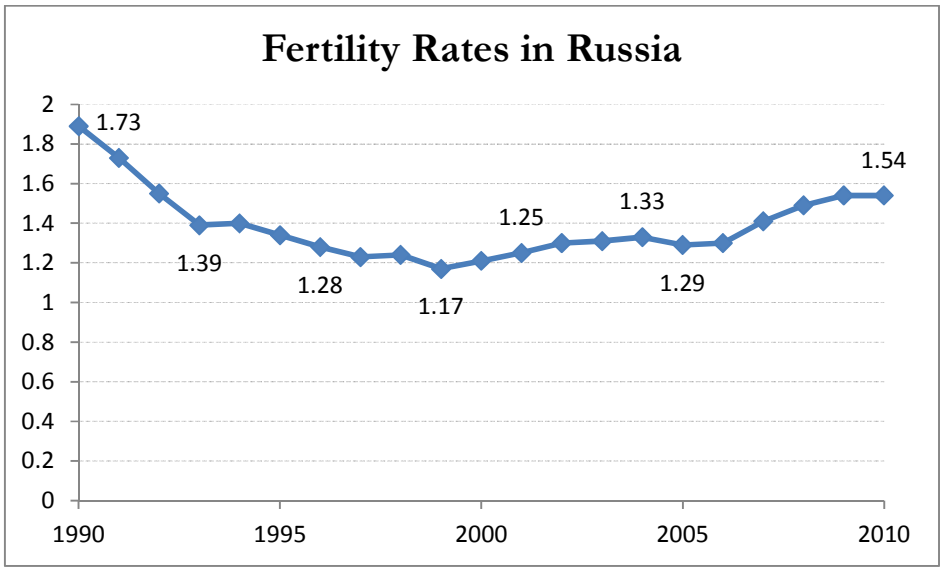
Steelman, L. C., and J. A. Mercy. 1980. Unconfounding the confluence model: A test of sibship size and birth-order effects on intelligence. *American Sociological Review* 45(4): 571-582.

Thomas, Margaret H. 1951. Sex Pattern and Size of Family. *British Medical Journal* 51(4): 733-734.

World Population Review. 2012. Population of Russia 2012. http://worldpopulationreview.com/population-of-russia-2012/ (Accessed on May 21, 2013).

Zajonc, R. B. 1976. Family configuration and intelligence: Variations in scholastic aptitude scores parallel trends in family size and the spacing of children. *Science.*

APPENDIX A



Source: World Bank Data, World Development Indicators

Figure A. Fertility rates in the Russian Federation, 1990-2010 years

APPENDIX B

Table B. Description of main variables used in the analysis

obs:	3,416
vars:	33
size:	386,008 (99.9% of memory free)
Var. Name	Variable Label
fam_pid	family ID
child_pid	child ID
moth_pid	mother ID
fath_pid	father ID
mother	dummy variable (1 if mother is present in the family)
father	dummy variable (1 if father is present in the family)
m_age	mother's age, years
f_age	father's age, years
b_order	birth order of a child
childnum	number of children in a family
sibsize	of siblings in a family (sibsize = childnum - 1)
female	dummy (1 if a child is a female, 0 if male)
boy1st	dummy (1 if the 1st child in a family is a boy, 0 otherwise)
boy2nd	dummy (1 if the 2nd child in a family is a boy, 0 otherwise)
samesex	dummy (1 if the first 2 children are of the same gender)
sett	settlement type: 1 – urban, 2 – PGT (urban settlement), 3 – rural
hthord	living conditions (1 - own residence; 2 - rented residence; 3 - dormitory; 7 - D/K; 8 - R)
tincm_nt	total income of a household (nominal)
itinschl	dummy (1 if a child currently attends school)
ittypesc	type of school did you study--a_c.k3.2
gen_school	a child attends general school (not gymnasium/lyceum)
schpayments	monthly school payments, RUR (required + non-required)
timeinv	time parents help in children's education, minutes per week
itprgres	grades/progress estimation --c.k3.4
m_heducl	mother: highest level of education, modified version
m_educ	mother's education, years
f_heducl	father: highest level of education, modified version
f_educ	father's education, years
Sorted by: fam_pid b_order	

APPENDIX C

Table C1. Distribution of grades (non-restricted sample)

progress estimation	Freq.	Percent	Cum.
Almost all 3 and often 2	14	0.93	0.93
Almost all 3	67	4.46	5.39
Almost all 4 and 3	276	18.36	23.75
Almost all 4	334	22.22	45.97
Almost all 5 and 4	625	41.58	87.56
Almost all 5	187	12.44	100
Total	1,503	100	

Table C2. Distribution of grades (restricted sample)

progress estimation	Freq.	Percent	Cum.
Almost all 3 and often 2	10	1.02	1.02
Almost all 3	48	4.88	5.89
Almost all 4 and 3	188	19.11	25
Almost all 4	220	22.36	47.36
Almost all 5 and 4	414	42.07	89.43
Almost all 5	104	10.57	100
Total	984	100	

APPENDIX D

First-stage regression summary statistics for Table 5.

Column 2: First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(1,695)	Prob > F
sibsize	0.0221	0.0173	0.0027	1.13848	0.2863

(F statistic adjusted for 696 clusters in fam_pid)

Column 4: First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(1,695)	Prob > F
sibsize	0.0225	0.0164	0.0026	1.1031	0.294

(F statistic adjusted for 696 clusters in fam_pid)

Column 6: First-stage regression summary statistics

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(1,706)	Prob > F
sibsize	0.0084	0.0036	0.0031	1.35081	0.2455

(F statistic adjusted for 707 clusters in fam_pid)

APPENDIX E

Table E. First-stage results for the general school attendance dependent variable

VARIABLES	(1) gen_school	(2) sibsize	(3) athrho	(4) lnsigma
sibsize	0.975			
	(0.699)			
father	-0.105	-0.521***		
	(1.258)	(0.121)		
rural	0.120	0.416***		
	(1.007)	(0.0814)		
tincm_nt	-5.15e-06***	3.26e-06***		
	(1.28e-06)	(9.82e-07)		
_If_heducl_2	0.0992	0.585***		
	(1.349)	(0.154)		
_If_heducl_3	0.290	0.381***		
	(1.241)	(0.103)		
_If_heducl_4	0.430	0.322***		
	(1.276)	(0.112)		
_If_heducl_5	0.118	0.262***		
	(0.803)	(0.0881)		
samesex		0.0495		
		(0.0719)		
Constant	-0.801	1.221***	-1.161	-0.234***
	(1.428)	(0.110)	(1.599)	(0.0724)
Observations	972	972	972	972

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *rural* – 1 for families living in rural areas; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_2* through *_If_heducl_5* – dummies for father's higher educational level: *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

APPENDIX F

Table F. Effect of sibship size on the probability of general school attendance (using samesex IV)

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	0.975	0.975	0.975	0.975	0.975	0.975
	(0.699)	(0.699)	(0.699)	(0.699)	(0.699)	(0.699)
father	-0.105	-0.105	-0.105	-0.105	-0.105	-0.105
	(1.258)	(1.258)	(1.258)	(1.258)	(1.258)	(1.258)
rural	0.120	0.120	0.120	0.120	0.120	0.120
	(1.007)	(1.007)	(1.007)	(1.007)	(1.007)	(1.007)
tincm_nt	-5.15e-06***	-5.15e-06***	-5.15e-06***	-5.15e-06***	-5.15e-06***	-5.15e-06***
	(1.28e-06)	(1.28e-06)	(1.28e-06)	(1.28e-06)	(1.28e-06)	(1.28e-06)
_If_heducl_2	0.0992	0.0992	0.0992	0.0992	0.0992	0.0992
	(1.349)	(1.349)	(1.349)	(1.349)	(1.349)	(1.349)
_If_heducl_3	0.290	0.290	0.290	0.290	0.290	0.290
	(1.241)	(1.241)	(1.241)	(1.241)	(1.241)	(1.241)
_If_heducl_4	0.430	0.430	0.430	0.430	0.430	0.430
	(1.276)	(1.276)	(1.276)	(1.276)	(1.276)	(1.276)
_If_heducl_5	0.118	0.118	0.118	0.118	0.118	0.118
	(0.803)	(0.803)	(0.803)	(0.803)	(0.803)	(0.803)
Observations	972	972	972	972	972	972

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *father* – 1 if father is present in a family; *sibsize* – sibship size; *rural* – 1 for families living in rural areas; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_2* through *_If_heducl_5* – dummies for father's higher educational level: *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

APPENDIX G

Table G1. Marginal effects of sibship size for the outcome 1 (almost all the 3 and the 2)

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	0.00527**	0.00750*	0.0104*	0.0140	0.0184	0.0236
	(0.00229)	(0.00386)	(0.00614)	(0.00921)	(0.0131)	(0.0176)
father	-0.238	-0.286	-0.336	-0.387	-0.436	-0.480
	(0.234)	(0.257)	(0.277)	(0.290)	(0.296)	(0.294)
female	-0.010***	-0.014***	-0.021***	-0.030**	-0.041**	-0.0547*
	(0.00330)	(0.00492)	(0.00792)	(0.0127)	(0.0194)	(0.0281)
tincm_nt	-8.3e-08*	-1.19e-07	-1.66e-07	-2.24e-07	-2.94e-07	-3.76e-07
	(4.96e-08)	(7.3e-08)	(1.0e-07)	(1.5e-07)	(2.1e-07)	(2.7e-07)
_If_heducl_1	0.364	0.421	0.478	0.530*	0.577**	0.616**
	(0.285)	(0.295)	(0.298)	(0.294)	(0.283)	(0.264)
_If_heducl_2	0.130	0.164	0.202	0.242	0.284	0.324
	(0.166)	(0.194)	(0.222)	(0.248)	(0.270)	(0.286)
_If_heducl_3	0.0123***	0.0185***	0.0272***	0.0390**	0.0546**	0.0745*
	(0.00414)	(0.00627)	(0.0104)	(0.0171)	(0.0272)	(0.0411)
_If_heducl_4	0.156	0.194	0.235	0.279	0.323	0.366
	(0.184)	(0.212)	(0.239)	(0.262)	(0.281)	(0.292)
_If_heducl_5	0.0924	0.119	0.149	0.183	0.218	0.254
	(0.133)	(0.160)	(0.190)	(0.218)	(0.245)	(0.268)
Observations	969	969	969	969	969	969

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table G2. Marginal effects of sibship size for the outcome 2 (almost all the 3)

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	0.0187***	0.0223***	0.0257**	0.0285**	0.0304***	0.0310***
	(0.00610)	(0.00837)	(0.0104)	(0.0117)	(0.0118)	(0.0103)
father	-0.245***	-0.233***	-0.210***	-0.176**	-0.131	-0.0793
	(0.0650)	(0.0419)	(0.0443)	(0.0728)	(0.108)	(0.141)
female	-0.045***	-0.057***	-0.069***	-0.081***	-0.092***	-0.102***
	(0.00904)	(0.0109)	(0.0141)	(0.0175)	(0.0200)	(0.0207)
tincm_nt	-2.9e-07*	-3.5e-07*	-4.0e-07*	-4.5e-07*	-4.8e-07*	-4.9e-07*
	(1.6e-07)	(1.9e-07)	(2.3e-07)	(2.5e-07)	(2.7e-07)	(2.7e-07)
_If_heducl_1	0.254***	0.224***	0.184*	0.135	0.0782	0.0176
	(0.0459)	(0.0712)	(0.100)	(0.127)	(0.151)	(0.168)
_If_heducl_2	0.197	0.200**	0.195***	0.180***	0.154***	0.119
	(0.121)	(0.100)	(0.0751)	(0.0521)	(0.0512)	(0.0782)
_If_heducl_3	0.0638***	0.0828***	0.105***	0.128***	0.153***	0.177***
	(0.0130)	(0.0162)	(0.0227)	(0.0318)	(0.0420)	(0.0520)
_If_heducl_4	0.213**	0.213**	0.203***	0.183***	0.152**	0.112
	(0.108)	(0.0837)	(0.0577)	(0.0439)	(0.0617)	(0.0960)
_If_heducl_5	0.164	0.172	0.173*	0.166**	0.149***	0.123**
	(0.134)	(0.121)	(0.102)	(0.0781)	(0.0569)	(0.0571)
Observations	969	969	969	969	969	969

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table G3. Marginal effects of sibship size for the outcome 3 (almost all the 3 and the 4)

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	0.0351***	0.0329***	0.0287***	0.0226***	0.0150**	0.00625
	(0.0123)	(0.0109)	(0.00768)	(0.00404)	(0.00630)	(0.0125)
father	-0.0585	0.00760	0.0732	0.134	0.188	0.231
	(0.137)	(0.152)	(0.162)	(0.165)	(0.160)	(0.147)
female	-0.121***	-0.123***	-0.120***	-0.111***	-0.094***	-0.0723*
	(0.0167)	(0.0173)	(0.0173)	(0.0192)	(0.0268)	(0.0396)
tincm_nt	-5.6e-07*	-5.2e-07*	-4.5e-07*	-3.60e-07	-2.38e-07	-9.95e-08
	(3.1e-07)	(3.0e-07)	(2.6e-07)	(2.2e-07)	(2.0e-07)	(2.2e-07)
_If_heducl_1	-0.0153	-0.0849	-0.150	-0.208	-0.255**	-0.290***
	(0.165)	(0.161)	(0.153)	(0.142)	(0.127)	(0.109)
_If_heducl_2	0.112*	0.0606	0.00499	-0.0510	-0.104	-0.150
	(0.0654)	(0.0981)	(0.128)	(0.150)	(0.164)	(0.168)
_If_heducl_3	0.219***	0.241***	0.256***	0.262***	0.257**	0.240*
	(0.0511)	(0.0645)	(0.0804)	(0.0978)	(0.117)	(0.139)
_If_heducl_4	0.102	0.0453	-0.0141	-0.0726	-0.127	-0.173
	(0.0877)	(0.117)	(0.141)	(0.159)	(0.167)	(0.166)
_If_heducl_5	0.122***	0.0800	0.0326	-0.0170	-0.0654	-0.109
	(0.0260)	(0.0589)	(0.0952)	(0.127)	(0.151)	(0.164)
Observations	969	969	969	969	969	969

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table G4. Marginal effects of sibship size for the outcome 4 (almost all the 4)

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	0.00460	-0.00204	-0.00853	-0.0144	-0.0192	-0.0226*
	(0.00380)	(0.00360)	(0.00690)	(0.00997)	(0.0118)	(0.0121)
father	0.164*	0.185**	0.197***	0.200***	0.195***	0.182***
	(0.0963)	(0.0840)	(0.0709)	(0.0581)	(0.0479)	(0.0434)
female	-0.052***	-0.032***	-0.0108	0.0119	0.0338	0.0532*
	(0.0108)	(0.00986)	(0.0145)	(0.0213)	(0.0270)	(0.0297)
tincm_nt	-7.33e-08	3.25e-08	1.36e-07	2.29e-07	3.05e-07	3.60e-07
	(6.8e-08)	(5.4e-08)	(1.0e-07)	(1.6e-07)	(2.0e-07)	(2.2e-07)
_If_heducl_1	-0.205***	-0.219***	-0.225***	-0.222***	-0.211***	-0.195***
	(0.0701)	(0.0568)	(0.0452)	(0.0362)	(0.0328)	(0.0369)
_If_heducl_2	-0.107	-0.133	-0.151	-0.161*	-0.162**	-0.157**
	(0.113)	(0.108)	(0.0998)	(0.0885)	(0.0760)	(0.0649)
_If_heducl_3	0.171*	0.150	0.121	0.0843	0.0425	-0.00210
	(0.0913)	(0.102)	(0.111)	(0.119)	(0.123)	(0.125)
_If_heducl_4	-0.123	-0.148	-0.165*	-0.173**	-0.173**	-0.165***
	(0.110)	(0.103)	(0.0926)	(0.0801)	(0.0676)	(0.0575)
_If_heducl_5	-0.0784	-0.105	-0.124	-0.136	-0.141	-0.139*
	(0.110)	(0.112)	(0.108)	(0.101)	(0.0896)	(0.0777)
Observations	969	969	969	969	969	969

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table G5. Marginal effects of sibship size for the outcome 5 (almost all the 4 and the 5)

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	-0.046***	-0.047***	-0.046***	-0.044***	-0.040***	-0.035***
	(0.0158)	(0.0164)	(0.0149)	(0.0120)	(0.00823)	(0.00451)
father	0.327***	0.291***	0.252***	0.212***	0.174***	0.139**
	(0.0602)	(0.0476)	(0.0429)	(0.0455)	(0.0507)	(0.0547)
female	0.134***	0.153***	0.164***	0.167***	0.162***	0.152***
	(0.0202)	(0.0220)	(0.0226)	(0.0216)	(0.0218)	(0.0259)
tincm_nt	7.44e-07*	7.62e-07*	7.46e-07*	7.02e-07*	6.36e-07*	5.58e-07*
	(4.1e-07)	(4.2e-07)	(4.1e-07)	(3.8e-07)	(3.5e-07)	(3.1e-07)
_If_heducl_1	-0.348***	-0.306***	-0.262***	-0.219***	-0.178***	-0.142**
	(0.0360)	(0.0312)	(0.0353)	(0.0437)	(0.0513)	(0.0561)
_If_heducl_2	-0.284***	-0.258***	-0.227***	-0.194***	-0.161***	-0.130**
	(0.105)	(0.0852)	(0.0695)	(0.0601)	(0.0562)	(0.0553)
_If_heducl_3	-0.114	-0.188*	-0.251***	-0.300***	-0.335***	-0.353***
	(0.138)	(0.0989)	(0.0626)	(0.0474)	(0.0706)	(0.109)
_If_heducl_4	-0.298***	-0.269***	-0.235***	-0.200***	-0.166***	-0.133**
	(0.0911)	(0.0726)	(0.0597)	(0.0537)	(0.0532)	(0.0543)
_If_heducl_5	-0.254*	-0.233**	-0.208**	-0.179**	-0.150**	-0.122**
	(0.131)	(0.108)	(0.0883)	(0.0732)	(0.0634)	(0.0578)
Observations	969	969	969	969	969	969

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table G6. Marginal effects of sibship size for the outcome 6 (almost all the 5)

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	-0.017***	-0.012***	-0.009***	-0.006***	-0.004***	-0.003**
	(0.00651)	(0.00390)	(0.00215)	(0.00132)	(0.00121)	(0.00124)
father	0.0498***	0.0352***	0.0243***	0.0164**	0.0108*	0.00695
	(0.00954)	(0.00798)	(0.00777)	(0.00738)	(0.00654)	(0.00543)
female	0.0946***	0.0753***	0.0585***	0.0442***	0.0326**	0.0234*
	(0.0145)	(0.0122)	(0.0126)	(0.0135)	(0.0138)	(0.0132)
tincm_nt	2.70e-07*	2.04e-07*	1.50e-07*	1.07e-07	7.46e-08	5.06e-08
	(1.5e-07)	(1.1e-07)	(8.6e-08)	(6.6e-08)	(5.1e-08)	(4.0e-08)
_If_heducl_1	-0.050***	-0.035***	-0.024***	-0.0165**	-0.0109*	-0.00697
	(0.00958)	(0.00803)	(0.00783)	(0.00743)	(0.00658)	(0.00545)
_If_heducl_2	-0.048***	-0.034***	-0.023***	-0.0161**	-0.0106*	-0.00684
	(0.0104)	(0.00838)	(0.00782)	(0.00730)	(0.00644)	(0.00534)
_If_heducl_3	-0.352	-0.305	-0.258	-0.214	-0.173	-0.136
	(0.279)	(0.265)	(0.247)	(0.225)	(0.201)	(0.175)
_If_heducl_4	-0.048***	-0.034***	-0.024***	-0.0162**	-0.0107*	-0.00689
	(0.00994)	(0.00813)	(0.00775)	(0.00731)	(0.00647)	(0.00538)
_If_heducl_5	-0.046***	-0.033***	-0.023***	-0.0156**	-0.0104	-0.00669
	(0.0124)	(0.00935)	(0.00812)	(0.00730)	(0.00635)	(0.00525)
Observations	969	969	969	969	969	969

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

APPENDIX H

Table H. First-stage results for the school grades as a dependent variable

VARIABLES	(1) itprgres	(2) sibsize	(3) lnsig_2	(4) atanhrho_12
samesex		0.0775		
		(0.0739)		
father	1.156	-0.680***		
	(1.237)	(0.120)		
female	0.594***	0.118**		
	(0.0890)	(0.0602)		
tincm_nt	3.62e-06	2.72e-06***		
	(2.21e-6)	(9.98e-07)		
_If_heducl_1	-1.298	1.117***		
	(1.671)	(0.262)		
_If_heducl_2	-0.721	0.841***		
	(1.266)	(0.171)		
_If_heducl_3	-1.017	0.626***		
	(1.168)	(0.118)		
_If_heducl_4	-0.983	0.466***		
	(1.030)	(0.111)		
_If_heducl_5	-0.748	0.388***		
	(0.919)	(0.0866)		
sibsize	-0.586			
	(0.802)			
Constant		1.294***	-0.179**	0.381
		(0.116)	(0.0711)	(0.820)
Observations	1,015	1,015	1,015	1,015

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

APPENDIX I

Table I. Marginal effects of sibship size on school performance (using samesex IV)

GRADES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
mainly 3 & 2	0.0180	0.0572	0.129	0.206	0.233	0.187
	(0.0361)	(0.182)	(0.440)	(0.560)	(0.254)	(0.279)
mainly 3	0.0587	0.0979	0.0934***	0.0200	-0.0704	-0.104
	(0.0746)	(0.150)	(0.0247)	(0.399)	(0.482)	(0.0937)
mainly 4 & 3	0.120	0.0788***	-0.0249	-0.108	-0.113***	-0.0681
	(0.134)	(0.0163)	(0.311)	(0.302)	(0.0353)	(0.258)
mainly 4	0.0371	-0.0401	-0.0833	-0.071**	-0.0360	-0.0121
	(0.0794)	(0.103)	(0.119)	(0.0357)	(0.120)	(0.0855)
mainly 5&4	-0.140	-0.158	-0.104***	-0.0456	-0.0138	-0.00292
	(0.0950)	(0.170)	(0.00860)	(0.0982)	(0.0723)	(0.0271)
mainly 5	-0.0935	-0.0356	-0.00962	-0.00184	-0.00025	-2.41e-05
	(0.229)	(0.0486)	(0.00704)	(0.00676)	(0.00190)	(0.000299)

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

APPENDIX J

Table J1. Marginal effects of sibship size for the outcome 1 (almost all the 2 and 3) using samesex IV

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	0.0180	0.0572	0.129	0.206	0.233	0.187
	(0.0361)	(0.182)	(0.440)	(0.560)	(0.254)	(0.279)
father	-0.122	-0.254	-0.388	-0.436	-0.360	-0.218
	(0.228)	(0.260)	(0.289)	(0.463)	(0.936)	(1.123)
female	-0.00963	-0.0351	-0.0914	-0.171	-0.228	-0.218
	(0.00697)	(0.0715)	(0.227)	(0.346)	(0.152)	(0.356)
tincm_nt	-1.11e-07	-3.53e-07	-7.95e-07	-1.27e-06	-1.4e-6**	-1.16e-06
	(1.3e-07)	(8.1e-07)	(2.0e-06)	(2.3e-06)	(6.2e-07)	(2.7e-06)
_If_heducl_1	0.155	0.305	0.444	0.479	0.384	0.228
	(0.365)	(0.378)	(0.389)	(0.634)	(1.085)	(1.209)
_If_heducl_2	0.0496	0.123	0.218	0.279	0.257	0.170
	(0.134)	(0.204)	(0.282)	(0.411)	(0.711)	(0.883)
_If_heducl_3	0.0112	0.0431	0.120	0.243	0.358	0.385
	(0.00732)	(0.0847)	(0.293)	(0.497)	(0.307)	(0.645)
_If_heducl_4	0.0883	0.197	0.319	0.377	0.324	0.203
	(0.155)	(0.220)	(0.279)	(0.357)	(0.784)	(1.011)
_If_heducl_5	0.0529	0.129	0.228	0.289	0.264	0.174
	(0.102)	(0.182)	(0.270)	(0.301)	(0.602)	(0.835)
samesex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
Observations	1,015	1,015	1,015	1,015	1,015	1,015

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table J2. Marginal effects of sibship size for the outcome 2 (almost all the 3) using samesex IV

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	0.0587	0.0979	0.0934***	0.0200	-0.0704	-0.104
	(0.0746)	(0.150)	(0.0247)	(0.399)	(0.482)	(0.0937)
father	-0.178	-0.162	-0.0359	0.118	0.185	0.148
	(0.235)	(0.209)	(0.387)	(0.355)	(0.173)	(0.590)
female	-0.039***	-0.080**	-0.103**	-0.0626	0.0271	0.0944
	(0.00895)	(0.0393)	(0.0520)	(0.348)	(0.521)	(0.216)
tincm_nt	-3.6e-07*	-6.04e-07	-5.76e-07	-1.24e-07	4.35e-07	6.4e-07**
	(1.9e-07)	(4.5e-07)	(6.8e-07)	(2.5e-06)	(2.6e-06)	(2.6e-07)
_If_heducl_1	0.200	0.165	0.0175	-0.143	-0.203	-0.156
	(0.282)	(0.206)	(0.339)	(0.266)	(0.261)	(0.658)
_If_heducl_2	0.103	0.122	0.0635	-0.0454	-0.117	-0.110
	(0.237)	(0.229)	(0.301)	(0.312)	(0.128)	(0.426)
_If_heducl_3	0.0504**	0.112**	0.164	0.136	0.0110	-0.123
	(0.0223)	(0.0489)	(0.124)	(0.577)	(0.999)	(0.658)
_If_heducl_4	0.149	0.151	0.0528	-0.0876	-0.160	-0.136
	(0.204)	(0.194)	(0.375)	(0.396)	(0.119)	(0.507)
_If_heducl_5	0.108	0.125	0.0633	-0.0493	-0.122	-0.113
	(0.175)	(0.174)	(0.312)	(0.381)	(0.127)	(0.388)
samesex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
Observations	1,015	1,015	1,015	1,015	1,015	1,015

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table J3. Marginal effects of sibship size for the outcome 3 (almost all the 3 and 4) using samesex IV

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	0.120	0.0788***	-0.0249	-0.108	-0.113***	-0.0681
	(0.134)	(0.0163)	(0.311)	(0.302)	(0.0353)	(0.258)
father	-0.136***	0.0396	0.185	0.207	0.137	0.0611
	(0.0431)	(0.133)	(0.137)	(0.338)	(0.541)	(0.438)
female	-0.110**	-0.108	-0.0303	0.0721	0.118***	0.0934
	(0.0444)	(0.0883)	(0.285)	(0.314)	(0.0192)	(0.337)
tincm_nt	-7.4e-7**	-4.87e-07	1.54e-07	6.65e-07	6.95e-07	4.20e-07
	(3.3e-07)	(5.4e-07)	(1.7e-06)	(1.2e-06)	(8.8e-07)	(1.9e-06)
_If_heducl_1	0.128	-0.0665	-0.212	-0.222	-0.143	-0.0627
	(0.133)	(0.164)	(0.175)	(0.412)	(0.585)	(0.455)
_If_heducl_2	0.122	0.0207	-0.0977	-0.142	-0.106	-0.0513
	(0.144)	(0.0945)	(0.140)	(0.268)	(0.434)	(0.372)
_If_heducl_3	0.162	0.190	0.104	-0.0603	-0.178	-0.179
	(0.157)	(0.270)	(0.597)	(0.719)	(0.201)	(0.578)
_If_heducl_4	0.138***	-0.0106	-0.151	-0.184	-0.127	-0.0583
	(0.0381)	(0.138)	(0.152)	(0.270)	(0.486)	(0.413)
_If_heducl_5	0.125	0.0183	-0.103	-0.146	-0.109	-0.0522
	(0.0995)	(0.129)	(0.168)	(0.204)	(0.400)	(0.363)
samesex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
Observations	1,015	1,015	1,015	1,015	1,015	1,015

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table J4. Marginal effects of sibship size for the outcome 4 (almost all the 4) using samesex IV

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	0.0371	-0.0401	-0.0833	-0.070**	-0.0360	-0.0121
	(0.0794)	(0.103)	(0.119)	(0.0357)	(0.120)	(0.0855)
father	0.0660	0.145	0.133	0.0755	0.0291	0.00782
	(0.248)	(0.163)	(0.209)	(0.252)	(0.178)	(0.0769)
female	-0.065***	0.00117	0.0671	0.0816	0.0541	0.0231
	(0.0110)	(0.0686)	(0.0762)	(0.0762)	(0.194)	(0.164)
tincm_nt	-2.29e-07	2.48e-07	5.14e-07	4.35e-07	2.22e-07	7.47e-08
	(3.0e-07)	(4.2e-07)	(3.4e-07)	(6.3e-07)	(9.4e-07)	(5.9e-07)
_If_heducl_1	-0.0863	-0.162	-0.141	-0.0781	-0.0297	-0.00791
	(0.324)	(0.204)	(0.238)	(0.266)	(0.183)	(0.0780)
_If_heducl_2	-0.0127	-0.0860	-0.0960	-0.0607	-0.0251	-0.00706
	(0.166)	(0.172)	(0.200)	(0.221)	(0.158)	(0.0703)
_If_heducl_3	0.126	0.0419	-0.0755	-0.134	-0.112	-0.0584
	(0.167)	(0.248)	(0.229)	(0.106)	(0.420)	(0.435)
_If_heducl_4	-0.0424	-0.123	-0.120	-0.0711	-0.0280	-0.00763
	(0.193)	(0.141)	(0.184)	(0.233)	(0.170)	(0.0748)
_If_heducl_5	-0.0152	-0.0899	-0.0988	-0.0620	-0.0254	-0.00714
	(0.133)	(0.125)	(0.157)	(0.202)	(0.153)	(0.0697)
samesex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
Observations	1,015	1,015	1,015	1,015	1,015	1,015

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table J5. Marginal effects of sibship size for the outcome 5 (almost all the 4 and 5) using samesex IV

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	-0.140	-0.158	-0.104***	-0.0456	-0.0138	-0.00292
	(0.0950)	(0.170)	(0.00860)	(0.0982)	(0.0723)	(0.0271)
father	0.289	0.206	0.100	0.0345	0.00853	0.00152
	(0.251)	(0.188)	(0.216)	(0.152)	(0.0647)	(0.0177)
female	0.0896	0.159***	0.136	0.0746	0.0280	0.00738
	(0.138)	(0.0437)	(0.143)	(0.218)	(0.160)	(0.0697)
tincm_nt	8.6e-07**	9.7e-07**	6.44e-07	2.82e-07	8.51e-08	1.80e-08
	(3.9e-07)	(4.2e-07)	(6.4e-07)	(8.6e-07)	(5.2e-07)	(1.8e-07)
_If_heducl_1	-0.313	-0.216	-0.103	-0.0351	-0.00862	-0.00153
	(0.279)	(0.205)	(0.226)	(0.155)	(0.0655)	(0.0179)
_If_heducl_2	-0.193	-0.157	-0.0826	-0.0301	-0.00775	-0.00142
	(0.366)	(0.250)	(0.214)	(0.141)	(0.0603)	(0.0168)
_If_heducl_3	-0.0695	-0.235*	-0.253	-0.167	-0.0753	-0.0241
	(0.145)	(0.138)	(0.436)	(0.627)	(0.504)	(0.253)
_If_heducl_4	-0.255	-0.190	-0.0951	-0.0333	-0.00834	-0.00150
	(0.252)	(0.183)	(0.205)	(0.147)	(0.0632)	(0.0175)
_If_heducl_5	-0.200	-0.161	-0.0842	-0.0306	-0.00784	-0.00143
	(0.267)	(0.187)	(0.189)	(0.135)	(0.0594)	(0.0167)
samesex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
Observations	1,015	1,015	1,015	1,015	1,015	1,015

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.

Table J6. Marginal effects of sibship size for the outcome 6 (almost all the 5) using samesex IV

VARIABLES	(1) 1 sibling	(2) 2 siblings	(3) 3 siblings	(4) 4 siblings	(5) 5 siblings	(6) 6 siblings
sibsize	-0.0935	-0.0356	-0.00962	-0.00184	-0.000250	-2.41e-05
	(0.229)	(0.0486)	(0.00704)	(0.00676)	(0.00190)	(0.000299)
father	0.0819	0.0252***	0.00565	0.000918	0.000108	9.12e-06
	(0.0972)	(0.00743)	(0.0135)	(0.00505)	(0.00103)	(0.000132)
female	0.136	0.0629***	0.0209	0.00496	0.000843	0.000102
	(0.0975)	(0.0146)	(0.0400)	(0.0228)	(0.00690)	(0.00130)
tincm_nt	5.77e-07	2.20e-07	5.94e-08	1.14e-08	1.54e-09	1.49e-10
	(9.2e-07)	(1.3e-07)	(9.7e-08)	(5.2e-08)	(1.31e-08)	(1.96e-09)
_If_heducl_1	-0.0839	-0.025***	-0.00570	-0.00092	-0.000108	-9.13e-06
	(0.101)	(0.00729)	(0.0136)	(0.00508)	(0.00103)	(0.000132)
_If_heducl_2	-0.0689	-0.0223	-0.00518	-0.00086	-0.000103	-8.87e-06
	(0.0533)	(0.0158)	(0.0140)	(0.00494)	(0.00100)	(0.000129)
_If_heducl_3	-0.280	-0.152	-0.0599	-0.0171	-0.00356	-0.000533
	(0.318)	(0.306)	(0.228)	(0.113)	(0.0365)	(0.00794)
_If_heducl_4	-0.0782	-0.024***	-0.00554	-0.00090	-0.000107	-9.08e-06
	(0.0889)	(0.00835)	(0.0134)	(0.00500)	(0.00102)	(0.000131)
_If_heducl_5	-0.0701	-0.0226**	-0.00524	-0.00087	-0.000104	-8.90e-06
	(0.0730)	(0.0114)	(0.0131)	(0.00485)	(0.000995)	(0.000129)
samesex	0	0	0	0	0	0
	(0)	(0)	(0)	(0)	(0)	(0)
Observations	1,015	1,015	1,015	1,015	1,015	1,015

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The meaning of variables: *sibsize* – sibship size; *father* – 1 if father is present in a family; *female* – child gender dummy; *tincm_nt* – nominal monthly income of the family (RUR); *_If_heducl_1* through *_If_heducl_5* – dummies for father's higher educational level: *1* – primary or incomplete secondary school; *2* – complete secondary school; *3* – vocational training school/professional courses; *4* – technical community college, training schools; *5* – institute, university, master's degree.