Do Sibship Size and Birth Order Matter to Child Education?Evidence from Vietnam

Nguyen, Cuong and Pham, Nguyet and Westbrook, Daniel

10 June 2007

Do Sibship Size and Birth Order Matter to Child Education? Evidence from Vietnam

Nguyen Viet Cuong

Pham Minh Nguyet

Daniel Westbrook

Abstract

The main measures the impact of sibship size and birth order on educational performance of children between 6 and 18 years old using regression and matching methods. Data used in this study are from Vietnam Household Living Standard Survey 2006. It is found that the number of siblings has a negative effect child school enrollment. An increase of one sibling leads to a decrease of around 2 percentage points in the percentage of school enrolment for children. Schooled children who have fewer siblings are also more likely to have better educational performance than school children who have more siblings. Having one additional brother or sister reduces the percentage of pupils with "excellent" educational performance as well as the percentage of pupils with "excellent and best" educational performance by around 2 percentage points. The birth order also matters to child education. Estimates from regressions and matching methods show that children with a high birth order tend to have higher rates of school enrollment than children with a low birth order.

JEL Classification: J13, J1, I2,

Keywords: Sibship size, birth order, education, household surveys, Vietnam.

1. Introduction

Educational development is an important policy goal in all countries. Economists have long been discussing the determinants of education attainment. Two household variables which receive increasing attention of researchers as well as policy makers are sibship size and birth order.[1] Economic theories suggest that there is a negative correlation between the number of children and the educational achievement of children (e.g. Becker, 1960; Becker and Lewis, 1973; Becker and Tomes, 1976). Parents will spend less time and other resources for each child if the number of children increases. Economists as well as sociologists agree to this negative relationship between sibship size and individual achievement (Kessler, 1991).

However, there is less agreement about the relationship between birth order and educational achievement. Some theories suggest that earlier-born children have better educational performance, while other theories predict that earlier-born children have worse performance. Arguments in favor of higher achievements by older siblings are that older siblings can receive more care from parents since when they are small their parents have fewer children to care for. Mothers can also have better health when they deliver earlier-born children (Booth and Kee, 2005). On the contrary, later-born children can have some advantages so that they can have better educational achievements than earlier-born children. They can receive care from both parents and older siblings. When having later-born children, parents can have more child rearing experience and perhaps higher income compared to the time of having earlier-born children. Thus the sign of the effect of birth order on educational performance cannot be known a priori (Kessler, 1991; Booth and Kee, 2005).

Empirical findings on the impact of number of siblings on educational attainment are not always consistent. Although most studies show a negative impact of sibship size (e.g., Belmont and Marolla, 1973, Blake, 1981; Booth and Kee, 2005; Rosenzweig and Wolpin, 1980; Lee, 2008; Rosenzweig and Zhang 2006), several studies do not find a significant effect (e.g., Black et al, 2005; Angist, 2005). Empirical studies also report different conclusions on the effect of birth order on educational achievement. For example, Belmont and Marolla (1973), Blake (1981) and Hauser and Kuo (1998) did not

[1] Sibship size is the number of siblings of the household

find significant effects of birth order, while Booth and Kee (2005), Kantarevic and Mechoulan (2005) and Black (2005) found a negative impact of birth order on educational achievements.

Vietnam has a population of 86 million and some parts of the country are the most densely populated areas in Asia. Rapid population growth can hamper national development efforts, destroy the environment and create pressure on living conditions such as housing, health care, education, employment opportunities, etc. The proportion of households having more than two children has increased in recent years (Hong Hai, 2009). The government of Vietnam has issued population policies to encourage families to have no more than two children. At the same time, the government has implemented a large number of education programs and policies to promote education of children. If having more children has adverse impacts on child education, then reducing the number of children can improve the quality of child education.

The effects of sibship size and birth order on child educational performance in Vietnam are unknown. Thus, this paper is expected to make an empirical contribution by estimating those effects and examining their magnitudes and statistical significance. The effects will be disaggregated by child age, genders and urban/rural locations. Information from the paper can be useful for policy makers and researchers in designing policies and programs on population and education.

The impact evaluation methods are instrumental-variables regression and propensity score matching. Although instrumental-variables regression is widely used in measuring the effects of sibship size and birth order, propensity score matching is rarely applied. The propensity score matching method provides an alternative by which the robustness of the instrumental variables results can be checked.

The main objective of the paper is to measure the impacts of sibship size and birth order on educational performance of children between 6 and 18 years old. The estimation controls for characteristics of individuals and households. This paper is structured into seven sections. The second section introduces data sources used in this paper. The third section reviews briefly the literature. The fourth section presents descriptive data on sibship size, birth order and child education in Vietnam. The fifth and sixth sections present estimation methods and empirical results. Finally, the seventh section concludes.

2. Review of the literature on sibship size, birth order and education

2.1. Economic theories

The theory of the trade-off between child quality and quantity was developed by Gary Becker and his associates (Becker, 1960; Becker and Lewis, 1973; Becker and Tomes, 1976). According to this theory, the smaller the number of children parents has, the better child quality is. Parents will invest less time and money in each child if they have more children. The quantity quality theory also is known in sociology as the “Resource Dilution Hypothesis”. The resource Dilution Hypothesis (RDH) is a general hypothesis of the relationship between family resources, parental resource allocation, and children’s outcomes. The RDH argues that the increasing dilution of parent’s resources including economic, social, time, etc., is the reason why children with many siblings obtain less education than children with fewer siblings (Anastasi 1956; Blake, 1981, 1985, 1989; Downey, 1995, 2001). The RHD argument is based on the observation that all types of parental resources (economic, time, social) are limited. Thus, when the size of a family increases, the amount of parental resources available for each child is decreased and as the result the child outcome is lower.

The other important theory used to predict effect of birth order and family size on child education is the Confluence Model (CM). The CM argues that the creation of an inferior intellectual environment in large families with many children leads to lower child education attainment. According to Zajonc and Markus, 1975, “… larger families will be associated with lower intellectual levels because the larger the family, the larger is the proportion of individuals with low absolute intelligence”. Parents have higher intellectual development than children, then when a new child is arrived with lower intellectual development the total family intellectual level decreases. The CM explains child educational attainment through the impacts of sibship size, birth order and birth spacing. According to CM, the first child has an advantage over the second child because he/she is born into higher intellectual environment. Similarly, the third child has less advantage than the second child and first child. This model predicts a negative correlation between birth order and children education attainment.[2]

[2] Notice that the argument is not about intellectual ability or intelligence. It is about development intellectual skill which accumulate over time

2.2. Empirical studies

Most empirical find that sibship size and individual achievement are negative correlated. However, there is less agreement about the relationship between birth order and educational achievement. Belmont and Marolla (1973) find negative relationships between family size and intellectual performance and between the birth order and IQ. Blake (1981), also finds a negative correlation between sibship size and educational attainment, but doesn't find a systematic difference in educational performance between earlier born and later born children.

However, the studies of Belmont and Marolla (1973), Blake (1981) and Hanushek (1992) do not consider the possible endogeneity of sibship size. Parents who do not pay much attention to child education might be more likely to have more children. More recent studies using instrumental variables approach to measure the impact of sibship size to children education attainment. Black, Devereux and Salvanes (2005) use multiple births as an instrument for sibship size to investigate the effect of number children on children educational performance. They find a negative relationship between family size and educational attainment, but when they use twin births as an instrument and include birth order dummies, they find no significant negative effect of the sibship size on educational attainment and a significant negative effect of birth order on educational attainment. Dalton and Glauber (2005) use the sex mix of the first two children as an instrumental variable, they find small significant negative effect of the number of siblings.

Recent studies from developing countries showed that the negative effect of sibship size to educational performance is "neither universal nor inevitable, particularly in developing countries, but depends on demographic, socio economic, and political factor external to the family, which influence both the availability of resources to the family and their internal allocation within the family in ways that affect children's education" (Lu and Treiman, 2005). For China case, Lu and Treiman (2005) argue that the affect of sibship size on educational performance disappears when schooling expands or becomes relatively less expensive. Moreover, sibship size has a little impact on the education of boy. Lu and Treiman (2005) emphasize that the external factors to the family, specially government policies which affect the cost of schooling can influence the impact of sibship size to educational outcome of children. If the government policies are oriented to social equality, educational disadvantages will be eliminated for children who have many siblings or live in rural area.

The relationship between sibship size and educational attainment in Vietnam was studied in Truong et al. (1998). The study used the 1994 Vietnam Inter-censal Demographic Survey (VNICDS). The study found that an increase in family size will lead to reduction in the probability of school attendance. The study also examined the difference of effect of the family size on girls and boys and found that the family size has larger effects (absolute value) on girls than on boys, especially girls in a family size with six members and above. However, Truong et al. (1998) did not find a significant difference in the effects of the family size between rural and urban areas.

3. Data sources

The research relies on data from the Vietnam Living Standard Survey (VHLSS) conducted in 2006. The 2006 VHLSS was conducted by the General Statistics Office of Vietnam (GSO). The survey contains detailed information on characteristics of individuals, households and communes. Commune data can be linked to individual and household data.

The Individual and household data include basic demographics, employment and labor force participation, education, health, income, expenditure, housing, fixed assets and durable goods, the participation of households in socioeconomic programs. The 2006 VHLSS collected especially detailed information on education of individuals including educational attainment, educational record in the schooling year 2005-2006 and the number of grades repeated in primary, secondary and high schools.

The commune questionnaires collect information on commune characteristics that affect local living standards. More specifically, commune data includes data on demographics and the general situation of the commune, general economic conditions and aid programs, non-farm employment, agricultural production, local infrastructure and transportation, education, health, and social affairs.

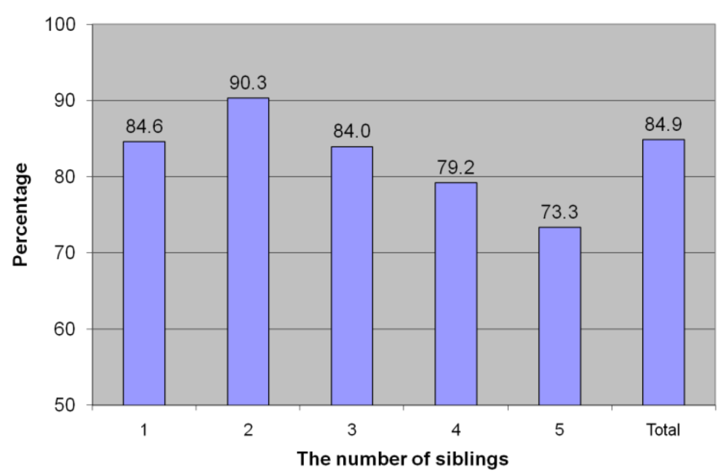
The 2006 VHLSS covers the 9,189 households. The number of individuals in the survey is 39,071. The number of households having children is 7,984. The number of siblings per household ranges from 1 to 11 in the 2006 VHLSS. The number of people between 6 and 18 year olds is 9,643. The large number of observations in the 2006 survey allow for analysis of the impacts of birth order and sibling size. Several other studies use smaller numbers of observations. For example, Booth and Lee (2005) and Lee (2008) used samples of 7,510 and 5,180 individuals, respectively.

The VHLSS is designed to be representative for rural and urban areas, and for 8 geographical regions. The number of communes sampled in the 2006 VHLSS is 2280.

4. **Sibship size, birth order and child education in Vietnam**

Vietnam experienced significant improvements in educational attainment during the 1990s and 2000s (World Bank, 2008). Economic growth and increasing investment in education from both the government and households are the main factors in educational development. However, Vietnam's population growth rate declined during the 1990s and early 2000s. The population growth rate decreased from 2.1 percent in 1989 to 1.26 percent in 2006 (VietNamNet Bridge, 2007). The fertility rate felt from 3.8 in 1989 to 2.33 in 1999 (VietNamNet Bridge, 2007). However, the estimated population growth rate has increased recently. The proportion of households having more than two children has also increased in recent years (Hong Hai, 2009).

Figure 1: School enrollment rate by sibship size

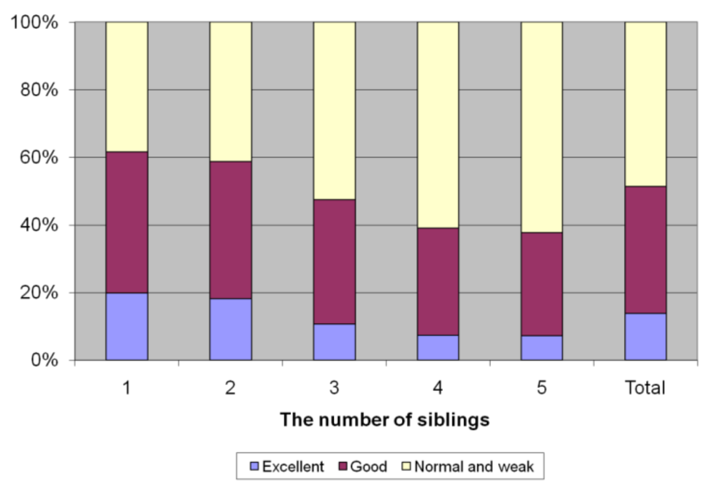


Source: Estimation from the 2006 VHLSS

Figure 1 graphs the relationship between school enrollment and sibship size in 2006 in Vietnam. It shows that the enrollment rate sharply significantly decreases when the number of siblings increases. In this study, we consider children of schooling ages, i.e., from 7 to 17 years old.[34]

As expected, educational performance is also better for pupils and students who have lower numbers of siblings (Figure 2). In the 2006 VHLSS, educational performance of pupils in the most recent academic year of pupils is measured by four outcomes: excellent, good, normal and weak. The percentage of pupils having educational performance "excenllent and good" is around 62 percent and 59 percent for pupils living in households who have one child and two children, respectively. On the contrary, for pupils in households with four children and above, this figure is around 38 percent.

Figure 2: Educational performance by sibship size

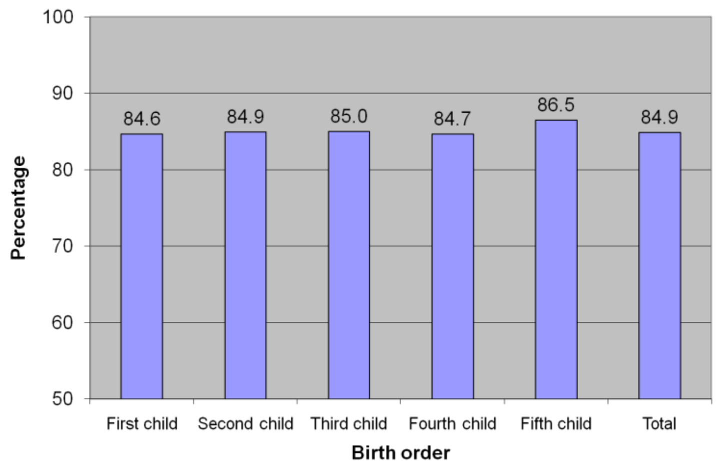


Source: Estimation from the 2006 VHLSS

[4] Although children start going to school from 6 years old, we do not consider 6 years old children in this study. Since the school year begins in September in Vietnam, 6 year old children who were surveyed before September 2006 are reported as "not attending schooling", while 6 year old children who were surveyed after September 2006 are reported as "attending schooling".

Figure 3 shows that the schooling rate is quite similar children of different birth order.

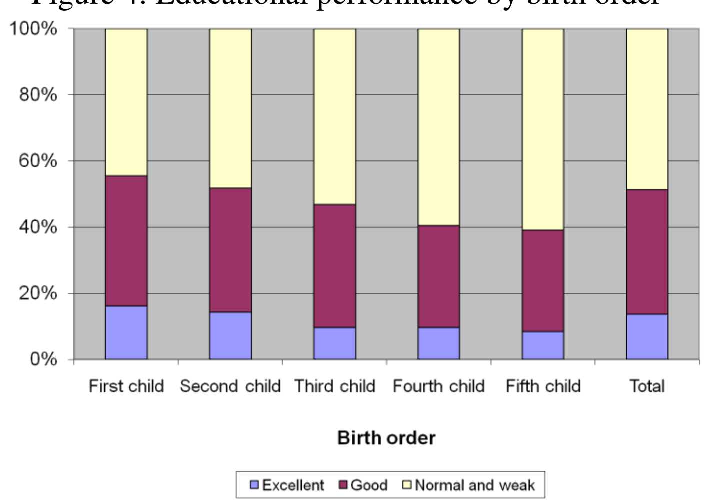
Figure 3: School enrollment rate by birth order



Source: Estimation from the 2006 VHLSS

Figure 4 shows that birth order and educational performance have a stronger relationship than birth order and school enrollment. Around 56 percent of first-born children and 52 percent of the second-born children have educational performance "excellent and good" while this "excellent and good" ratio among fourth-born and later-born children is 39 percent. The ratio of pupils having "excellent" educational results is much higher for children of lower birth order than for children of lower birth order.

Figure 4: Educational performance by birth order



Source: Estimation from the 2006 VHLSS

5. Methods to Measure the Effect of sibship size and Birth Order

5.1. Measuring the impact of sibship size

To measure the impact of the number of siblings on educational performance, we assume the follow function:

$$P(Y=1 \mid X, H) = G(\beta_0 + X\beta_1 + H\beta_2), \quad (1)$$

where *Y* is an indicator of educational performance, *H* is the number of siblings, and *X* is a vector control variables including individual and household characteristics which can affect educational performance. The impact of the sibship size is measured by β_2 under assumption that *H* is exogenous in equation (1).

The dependent variable *Y* is one of a set of dummy variables including school enrollment, "excellent" educational performance, and "excellent and good" educational performance in the most recent academic year. The *X* variables can be education and age of parents, household income, age and sex of children, and geographic variables (e.g., Black et al., 2005; Lee, 2008). Since the dependent variable is binary one, we will use a probit model to estimate equation (1).[5]

Since equation (1) is not linear, the partial effect of sibling size is not measured directly by β_2. For the probit model, the partial effect is calculated as follows:

$$PE = \frac{dP(Y=1 \mid X, H)}{dH} = \frac{dG(\beta_0 + X\beta_1 + H\beta_2)}{dH} = \beta_2 \phi(\beta_0 + X\beta_1 + H\beta_2), \quad (2)$$

where $\phi(z)$ is the standard normal density. The value of PE changes across X, and we can estimate the partial effect for the average unit in the population, i.e., at the average value of variables *X* and *H*.

The main problem in estimating β_2 is the endogeneity of the sibship size. Education of children and the number of siblings might be jointly determined. Parents who do not pay much attention to child education might be more likely to have more children. Omitted variables in equation (1) can affect the number of siblings, and as a result, the estimation of the impact of the sibship size can be biased. A standard method to

[5] We use the probit model instead of the logit model, because our software (STATA) allows for instrumental variables probit regressions but not instrumental variables logit regressions. Probit models are presented in most econometrics textbook such as Wooldridge (2002).

deal with endogeneity is instrumental-variables regression. In this paper, we will use instrumental-variables regressions to measure the impact of sibship size. Valid instrumental variables are those which affect sibship size but do not affect educational performance.

5.2. Measuring the impact of birth order

Impact parameters

Unlike sibship size, birth order is often treated as an exogenous variable in the educational performance equation once other observed variables such as sibship size and child age are controlled (e.g, Kessler, 1991; Ejrnæs and Portner, 2004; Kantarevic and Mechoulan, 2005; Haan, 2005). Instead of using linear regression, the paper will rely on the matching method to estimate the impact of birth order. Matching methods are often used for impact evaluation of treatments and interventions. There is a large amount of literature on matching methods of impact evaluation (e.g., see Rubin 1977, 1979, 1980; Rosenbaum and Rubin, 1983; Dehejia and Wahba, 1998). The main advantage of the matching method is that it does not impose assumptions on the functional form of the educational performance equation.

In previous studies on birth order, the birth order variable can be an index of birth order or dummy variables indicting the first child, the second child, and until the n^{th} child. In this paper, we will measure the effect of birth order by comparing the educational performance between children of different birth orders. We group children into three groups: (i) first-born children; (ii) second-born; (iii) third-born children and later-born children. Third-born children and later-born children are grouped together to increase the number of observations. There are only a few children which have birth order above three. Children from three and above birth orders are called third-born children.

The most popular parameter in the impact evaluation literature is Average Treatment Effect on the Treated (ATT) (see Heckman et al., 1999). In this study, we use the idea of ATT to define the impact of birth order. More specifically, we will measure the effect of being the first-born children relative to the second-born children and the third-born children:

$$ATT_{FS} = E(Y_F \mid F = 1) - E(Y_S \mid F = 1), \qquad (3)$$

$$ATT_{FT} = E(Y_F \mid F = 1) - E(Y_T \mid F = 1), \tag{4}$$

where Y_F, Y_S and Y_T are educational performance of first-born children, second-born children, and third-born children, respectively. *F* is a dummy variable denoting first-born children (*F* equals 1 if first-born children, 0 otherwise).

Similarly, we will examine the effect of being second-born children relative to third born children:

$$ATT_{ST} = E(Y_S \mid S = 1) - E(Y_T \mid S = 1), \tag{5}$$

where *S* is a dummy variable denoting second-born children (*S* equals 1 if second-born children, 0 otherwise).

Matching method

Since children of different birth orders can be different in observed characteristics such as age and parents' income, we will compare educational performance of similar children using the matching method. The matching method helps us identify the children of different birth orders who are nearly identical in every other way. For illustration, suppose that we estimate the effect of being first-born children as expressed by equation (3). The assumption to identify equation (3) is expressed as follows:

$$E(Y_S \mid F = 1, X) = E(Y_S \mid S = 1, X). \tag{6}$$

It implies that the outcomes for second born children (controlling for *X*) are good estimations for the counterfactuals for first born children (controlling for *X*). Then, the parameter ATT_{FS} conditional on *X* is identified:

$$ATT_{FS\_X} = E(Y_F \mid F = 1, X) - E(Y_S \mid S = 1, X) \tag{7}$$

To construct the comparison group, we find one or more second-born children whose have observed characteristics similar to the characteristics of the first-born children. To find second-born children who have similar characteristics as first-born children, we will use Mahalanobis matching (Rubin, 1979, 1980) and propensity score matching methods (Rosenbaum and Rubin, 1983). We also use different matching schemes including nearest neighbour matching, five nearest neighbor matching and kernel matching schemes.

6. The Impact of Birth Order and Sibship Size on Child Education

6.1. The impact of sibship size

We estimate the impact of sibship size on educational performance of children by estimating of equation (1). As mentioned above, the dependent variable *Y* is one of a set of dummy variables including school enrollment, "excellent" educational performance, and "excellent and good" educational performance in the most recent academic year. The list of independent variables and their basic statistics are presented in Table A.1 in the Appendix.

Table 1 presents estimates of the impact of sibship size in the probit regressions and the instrumental variables probit regressions. This table presents only the estimates of the sibship size coefficient. The full results are presented in the Table A.2 in Appendix. To examine the sensitivity of the estimates of sibship size to model specification, we use four models which mostly vary in the number of explanatory variables included. Model 1 uses only the sibship size as the explanatory variable. There are no control variables in this model. Model 2 uses the sibship size and strictly exogenous explanatory variables. Model 3 uses all the variables including education and age of parents, household income and birth order of children. Finally, we use the process of `stepwise deletion' to selection Model 4. In Model 4, only variables which are significant at 30 percent significance level are kept. We use the large significance level in determining which variables to drop so that we do not accidentally omit one that is relevant.

The upper panel of Table 20 presents the probit regressions. It shows that the impact estimates of sibling size are very similar in different models. It implies that the impact estimate of sibship size is not sensitive to the omission of control variables. Children with more siblings tend to have significantly lower school enrollment probability. In the largest model, model 4, the estimate of the coefficient of the number of siblings is -0.165. It implies that the partial effect for the average child in the population is around -0.021. In other words, having one additional sibling, the percentage of school enrolment for children is reduced by around 2.1 percentage points.

Table 1: The impact of sibship size on school enrollment

Explanatory variables	Model 1	Model 2	Model 3	Model 4
Probit regression				
The number of siblings	-0.146***	-0.137***	-0.164***	-0.165***
	[0.013]	[0.016]	[0.022]	[0.022]
Control variables				
Age and gender of children		Yes	Yes	Yes
Urban and regional variables			Yes	Yes
Age and education of parents			Yes	Yes
Per capita income				Yes
IV Probit regression				
The number of siblings	-0.209***	-0.341***	-0.334*	-0.323*
	[0.047]	[0.086]	[0.193]	[0.167]
Control variables				
Age and gender of children		Yes	Yes	Yes
Urban and regional variables			Yes	Yes
Age and education of parents			Yes	Yes
Per capita income				Yes

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Source: Estimation from the 2006 VHLSS

As mentioned, the sibship size can be endogenous in education equations. A standard and popular method to deal with endogeneity is instrumental-variables regression. A difficulty of this method is to find a valid instrument for sibship size, which is correlated with sibship size but not educational performance. Following the family economics literature, we tried to use the presence of twins, sex composition of children, and first child's sex as instruments for sibship size (e.g., see Goux and Eric, 2005; Rosenzweig and Kenneth, 1980; Black et al., 2005; Haan, 2005). Having twins can increases the number of siblings from a given fertility decision. Families who have the first two siblings with the same sex may be more likely to have third child, especially of the first two are girls. It is well known that Vietnamese prefer sons to daughters. If the first child is girl, they will tend to have the second child. It is often argued that these variables are absent in the educational attainment education. However, these instrumental variables are not functional in this study. They are not correlated with the number of sibship size, or the regressions with these instrumental variable yield very unusual estimates.

We create a variable which is equal to the average sibship size at the district level. For each child, we compute this variable by averaging the sibship size per district (excluding the child itself). Households in different areas can have different opinions on the number of children they should have. For example, parents in some rural areas tend to have more children, since they expect more care from children when older. In other words, different areas can have different culture which can affect the number of children per household. The condition on correlation between the number of siblings of each child and the average sibling size at the district level can be tested by running regression of the former variable on the later variable. Table A.5 in Appendix presents these regressions. It shows that the district average variable is strongly correlated with children's sibling size.

The exclusion condition that the district average sibship size is not correlated with unobserved variables in education equations cannot be tested without another valid instrument. It is expected that the district average sibship size does not affect educational performance of each child given other observed variables are controlled in the education equations. Living in a district which have high population growth would affect the number of children of a household but not influence the children's education.

Table 1 shows that the estimates yielded by instrumental variables probit regressions are also negative and statistically significant. The full regression results are presented in Tables in Appendix. In the largest model, model 4, the estimate of the coefficient of the number of siblings is -0.323. It implies that the partial effect for the average child in the population is around -0.038. The point estimates from instrumental variables probit regressions are smaller than the point estimates from probit regressions. If the instrument is valid, we can test the endogeneity of the sibship size. Stata reports the Wald test of exogeneity of the sibship size (in Tables in Appendix). In most models except Model 2, the hypothesis on exogeneity is not rejected. It means that if our instrument is valid, the variable of sibling size is exogenous in the equation of school enrolment.

Table 2 investigates the impact of sibship size on educational performance "Excellent". This table is estimated for pupils, i.e., schooling children. For both probit regression and instrumental variable probit regression, the estimates of the impact are negative. Except Models 3 and 4 in instrumental variable probit regression, all models yield statistically significant estimates. Having more siblings reduce the probability of having "excellent" education. For Model 4, probit regression and instrumental variable probit regression give the estimates of -0.063 and -0.155, respectively. The partial effects estimated for the average child is equal to -0.011 and -0.023, respectively. It means that one additional sibling reduce the percentage of having educational performance "excellent" by around 1.1 percentage points to 2.3 percentage points.

Table 2: The impact of sibship size on educational performance "Excellent"

Explanatory variables	Model 1	Model 2	Model 3	Model 4
Probit regression				
The number of siblings	-0.191***	-0.151***	-0.059*	-0.063**
	[0.021]	[0.022]	[0.032]	[0.030]
Control variables				
Age and gender of children		Yes	Yes	Yes
Urban and regional variables			Yes	Yes
Age and education of parents			Yes	Yes
Per capita income				Yes
IV Probit regression				
The number of siblings	-0.368***	-0.290***	-0.123	-0.155
	[0.053]	[0.099]	[0.199]	[0.141]
Control variables				
Age and gender of children		Yes	Yes	Yes
Urban and regional variables			Yes	Yes
Age and education of parents			Yes	Yes
Per capita income				Yes

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Source: Estimation from the 2006 VHLSS

Finally, Table 3 reports the impact of sibship size on the probability of achieving educational performance "Excellent and good". The impact estimates from all the models are negative. However, estimates from instrumental variable probit regression in Models 2 to 4 are not statistically significant. For the Model 4, the estimates of sibship size from probit regression and instrumental variable probit regression are approximately equal to -0.057 and -0.125, respectively. The estimate of -0.057 means that an increase of one sibling will decrease the percentage of having educational performance "excellent and good" by around 2.2 percentage points.

Table 3: The impact of sibship size on educational performance "Excellent and good"

Explanatory variables	Model 1	Model 2	Model 3	Model 4
Probit regression				
The number of siblings	-0.166***	-0.107***	-0.058***	-0.057***
	[0.014]	[0.015]	[0.022]	[0.015]
Control variables				
Age and gender of children		Yes	Yes	Yes
Urban and regional variables			Yes	Yes
Age and education of parents			Yes	Yes
Per capita income				Yes
IV Probit regression				
The number of siblings	-0.352***	-0.076	-0.146	-0.125
	[0.043]	[0.080]	[0.150]	[0.133]
Control variables				
Age and gender of children		Yes	Yes	Yes
Urban and regional variables			Yes	Yes
Age and education of parents			Yes	Yes
Per capita income				Yes

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Source: Estimation from the 2006 VHLSS

It should be noted that we also investigate whether the impacts of sibship size differ for different groups of children by age, gender, and urban/rural locations by including interactions between sibship size with age, gender and urbanity. However, these interaction variables are not statistically significant. We also run separate regressions for different age groups, boys and girls, children in urban as well as children in rural to test the heterogeneity in the impact of sibship size. Again, the difference in the estimated impact of sibship size between different child group is not statistically significant.

The regressions of school enrollment and educational performance also reveal characteristics associated with child school enrollment and educational performance (Tables in Appendix). Older children have lower school enrollment rates and lower educational performance. It is interesting that girls are better than boys in both school enrollment and educational performance. Kinh and Chinese children are more likely to have higher rates of school enrollment and better educational performance than ethnic minority children. There is no statistically significant difference in school enrollment between urban children and rural ones. However, once attending school, urban children have much better educational performance than rural ones. There are differences in child education across some regions. For example, North East and South Central Coast have the highest school enrolment rates. Although Mekong River Delta and South East have very higher income compared with other regions, these two regions have the lowest school enrolment rates.

Finally, as expected, parents with higher education degrees and higher income tend to have children who are better in school enrollment and educational performance.

6.2. The impact of birth order

The effect of birth order on school enrolment and educational performance is examined in regressions presented in Tables A.2, A.3 and A.4. It shows that birth order is positive and statistically significant in the equation of school enrollment. It means that children with a high birth order are more likely to be enrolled in school than children with a low birth order. The effect of birth order on educational performance is negative but not statistically significant. A problem with probit regressions is that we have to rely on assumptions on functional forms of education and standard normal distribution of error terms. To examine the robustness of the impact estimates of birth order, this section presents the impact of birth order using two matching methods: Mahalanobis matching and propensity score matching.

The first matching method is Mahalanobis matching which matches children based on the similarity of their characteristics. Children are grouped into three groups: first-born, second-born and third-born children children of birth order higher than three are also included in the third-born group). We will match first-born children with second-born children and third-born children to estimate the effect of being the first-born children relative to second-born children and to third-born children, respectively. Similarly, we will match second-born children with third-born children to estimate the effect of being second-born children relative to third-born children. The covariates (control variables) include sibship size, age and sex of children, education and age of parents, household income, urbanity and regional dummy variables. We also have two models which are different in the number of control variables. The so-called small model includes the number of siblings, age and sex of children, age of parents, urbanity and regional dummy variables, while the so-called large model includes all the control variables. The small model includes more exogenous variables. Comparison of small and large models is to examine the sensitivity of impact estimates to endogenous control variables.

In the propensity score matching method, the first step is to estimate propensity scores. Children are grouped into three groups: first-born, second-born and third-born children. We estimate the probability of being included in a group using a multinomial logit model and calculate the propensity score for matching using the approach of Lechner (2001). Table A.6 in the Appendix present the regressions to estimate the propensity scores. In this study, we use nearest neighbor matching, five nearest neighbor matching and kernel matching schemes. These matching schemes give very similar results. Thus in this study, we will present results from five nearest neighbor matching. We cannot reject equality of the means of the covariates between children groups and their matched groups for most of the covariates. We will use the results from the Mahalanobis matching for the interpretation. Results from propensity score matching are presented in Table A.7 in Appendix.

Table 4 presents the impact of birth order using Mahalanobis matching. Y_O denote the observed outcome, while Y_C denotes the estimated counterfactual of outcome. For example, in the large model, the observed school enrollment rate for first-born children is 84.64 percent, and the school enrollment rate of second-born children who are matched with the first-born children 87.12 percent. The impact of being "first child" relative to being "second child" for the first-born children is equal to -2.48 percentage points.

It shows that two models, small and large, give quite similar estimates. First-born children have lower schooling rates than matched second-born children and third-born children. Second-born children have lower rates of school enrolment than similar third-born children. This difference is statistically significant in the large model, but not in the small model. Similar findings are also found in the propensity score matching method (Table A.7 in Appendix). OLS regressions also show children with a high birth order tend to have higher enrollment rates that children with a low birth order. One possible explanation is that parents can pay more attention to their younger children when their older children grow. For example, suppose there are two families. The first family has two children: one is five years old and another is ten years old. The second family has two children: one is ten years old and another is fifteen years old. In this first family the parents have to care for their small kid and might pay less attention to their first-born child, while the parents in the second family can pay more attention to the second-born child since their the first-born child become more independent.

Yet, once attending school, first-born children are more likely to have better educational performance than second-born or third-born children. For the large model, the percentage of first-born children who have educational performance “excellent and good” is 3.4 percentage points higher than this percentage of similar second-born children and 9.6 percentage points higher than this percentage of similar third-born children. The percentage of first-born children with educational performance “excellent” is also 5.5 percentage points higher than third-born children who have similar characteristics as the first-born children. Similar findings are also found in the propensity score matching.

Table 4: The impact of birth order using Mahalanobis matching

Outcomes	**Small model**			**Large model**		
	Y_O	Y_C	ATT = $Y_O - Y_C$	Y_O	Y_C	ATT = $Y_O - Y_C$
The impact of being “first child” relative to being “second child”						
School enrollment rate	84.64***	86.02***	-1.38	84.64***	87.12***	-2.48**
	[0.65]	[1.15]	[1.32]	[0.65]	[0.99]	[1.19]
Performance: “excellent” rate	16.02***	14.79***	1.23	16.02***	14.73***	1.29
	[0.72]	[1.22]	[1.41]	[0.72]	[1.14]	[1.35]
Performance: “Excellent and good” rate	55.03***	49.68***	5.35***	55.03***	51.59***	3.44*
	[0.98]	[1.79]	[2.04]	[0.98]	[1.62]	[1.89]
The impact of being “first child” relative to being “third child”						
School enrollment rate	84.64***	85.65***	-1.00	84.64***	86.32***	-1.67
	[0.65]	[2.12]	[2.22]	[0.65]	[1.69]	[1.81]
Performance: “excellent” rate	16.02***	10.27***	5.75***	16.02***	10.50***	5.52***
	[0.72]	[1.83]	[1.97]	[0.72]	[1.60]	[1.75]
Performance: “Excellent and good” rate	55.03***	41.11***	13.92***	55.03***	45.46***	9.57***
	[0.98]	[3.15]	[3.29]	[0.98]	[2.65]	[2.82]
The impact of being “second child” relative to being “third child”						
School enrollment rate	84.94***	87.23***	-2.30	84.94***	88.37***	-3.43**
	[0.65]	[1.51]	[1.65]	[0.65]	[1.32]	[1.47]
Performance: “excellent” rate	14.25***	12.13***	2.12	14.25***	13.35***	0.90
	[0.69]	[1.32]	[1.50]	[0.69]	[1.30]	[1.47]
Performance: “Excellent and good” rate	51.48***	47.87***	3.61	51.48***	48.94***	2.54
	[0.99]	[2.28]	[2.49]	[0.99]	[2.13]	[2.35]

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Source: Estimation from the 2006 VHLSS

However, the estimates of the difference in educational performance between second-born children and matched third-born children are not statistically significant. In addition, the estimates of birth order on educational performance using regressions are not statistically significant. Thus the effects of birth order on educational performance should be interpreted with caution.

7. Conclusions

This paper aims to measure the impact of sibship size and birth order on child education. It is found that the number of siblings has a negative effect child school enrollment. However, the effect is not large. The partial effect for the average child in the population estimated from probit regression and instrumental probit regression is around -0.011 and -0.023, respectively. In other words, according to the probit regression and instrumental probit regression, an increase of one sibling leads to a decrease of around 1.1 to 2.3 percentage points in the percentage of school enrolment for children. Schooled children who have fewer siblings are also more likely to have better educational performance than school children who have more siblings. Having one additional brother or sister reduces the percentage of pupils with “excellent” educational performance as well as the percentage of pupils with “excellent and best” educational performance by around 2 percentage points.

The birth order also matters to child education. Estimates from regressions and matching methods show that children with a high birth order tend to have higher rates of school enrollment. Yet, once enrolled in school, children with a higher birth order have poorer educational performance. Descriptive analysis shows that around 56 percent of first-born children and 52 percent of the second-born children have educational performance “excellent and good”, while this “excellent and good” ratio among fourth-born and later-born children is 39 percent. The ratio of pupils having “excellent” educational results is much higher for children of lower birth order than for children of lower birth order. When other contemporaneous factors are controlled by the matching method, educational performance of children of higher birth order is still worse than educational performance of children of lower birth order. The percentage of first-born children who have educational performance “excellent and good” is 3.4 percentage points higher than this percentage of similar second-born children and 9.6 percentage points higher than this percentage of similar third-born children. The percentage of first-born children with educational performance “excellent” is 5.5 percentage points higher than third-born children who have similar characteristics as the first-born children.

Policy implications can be driven straightforwardly from the research findings. Having a large number of children can have unfavorable impacts on child education. Attention that parents pay to each child will decrease as the number of children increase. The recent increase in population growth can lead to reduction in child education. High population growth can reduce educational spending from the government as well as society for each child. Thus prevention of high population growth can be an effective measure to increase child education.

A long term measure to reduce population growth is to encourage families to have less than three children. In Vietnam, especially in rural areas, people tend to have a large number of children so that they can receive more care from their children when becoming old. If the social securities are developed, people will be more likely to have fewer children. In Vietnam, the thought “men better than women” remains popular, particularly in rural and mountainous areas. As a result, households tend to have more children until they have a boy. It is necessary to increase gender equality and women’s role, especially for ethnic minorities and people in rural, mountainous and remote areas. In addition, family planning methods should be also introduced and disseminated.

Although birth order can matter to child education, it is not clear how to propose policy implications, which are based on birth order, to improve child education. Perhaps, parents as well as teachers should be aware that children with a low birth order can be more likely to drop out of school.

Finally, the findings suggest the important role of economic growth on education and prevention of high population growth. Households with high income tend to have a small number of children and their children have better education than low-income households. Thus, stimulating economic growth is a long-run measure to improve education and reduce population growth in Vietnam.

References

Angrist, J. D., V. Lavy, and A. Schlosser (2005), "New Evidence on the Causal Link between the Quantity and Quality of Children", NBER Working Paper 11835.

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

Becker, G., and N. Tomes (1976), "Child Endowments and the Quantity and Quality of Children," *Journal of Political Economy*, 84, 398-419.

Becker, Gary S., (1960) "An Economic Analysis of Fertility," *Demographic an*d *Economic Change in Developed Countries*, Gary S. Becker, (ed.) Princeton, NJ: Princeton University Press.

Belmont, L., and Marolla, F. A. (1973), "Birth Order, Family Size, and Intelligence". *Science*, Vol 182 (4117), 1096-1101.

Black, Sandra, Paul J. Devereux, Kjell G. Salvanes (2005) "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education", *Quarterly Journal of Economics*, Vol. 120(2), pp. 669-700.

Blake, J. (1981), "Family Size and the Quality of Children", *Demography*, Vol 18, (4), 421-442.

Booth, Alison L. and Hiau Joo Kee (2005), "Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment", Discussion Paper No. 1713, The Institute for the Study of Labor (IZA), Germany.

Dehejia, Rajeev and Sadek Wahba. 1998. "Propensity Score Matching Methods for Non-Experimental Causal Studies." NBER Working Paper 6829, Cambridge, Mass.

Ejrnæs, M. and C. Portner (2004), "Birth Order and the Intrahousehold Allocation of Time and Education", Unpublished paper.

Goux, Dominique and Eric Maurin (2005) "The Effect of Overcrowded Housing on Children's Performance at School," *Journal of Public Economics*, Vol. 89, pp. 797-819.

Haan, Monique (2005), "Birth Order, Family Size and Educational Attainment", Tinbergen Institute Discussion Paper No. 111/3, Tinbergen Institute, the Netherlands.

Hauser, R.M., and Kuo, H.D. (1998), "Does the Gender Composition of Sibships Effect Women.s Educational Attainment?", *The Journal of Human Resources*, Vol 33, (3), 644-657.

Heckman, Jame, R. Lalonde and J. Smith. (1999). The Economics and Econometrics of Active Labor Market Programs. *Handbook of Labor Economics, Volume 3*, Ashenfelter, A. and *D*. Card, eds., Amsterdam: Elsevier Science.

Hong Hai (2009), "Tỷ lệ sinh con thứ 3 tiếp tục tăng", Dan Tri Newspaper, 02-07-2009. Available at: http://dantri.com.vn/c7/s7-334693/ty-le-sinh-con-thu-3-tiep-tuc-tang.htm

Kantarevic Jasmin and Stéphane Mechoulan (2005), "Birth Order, Educational Attainment and Earnings: An Investigation Using the PSID", IZA DP No. 1789,

Kessler, Daniel (1991), "Birth Order, Family Size, and Achievement: Family Structure and Wage Determination", *Journal of Labor Economics*, Vol. 9(4), pp. 413-425.

Lechner, M. (2001), "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", In: Lechner, M. and Pfeiffer, F. (Eds.). Econometric Evaluation of Labour Market Policies. Heidelberg, Physica-Verlag.

Lee, J. (2008), "Sibling Size and Investment in Children's Education: An Asian Instrument", *Journal of Population Economics*, Vol. 21(4), pp. 855-875.

Lu Y. and Treiman D. J. (2005), "The Effect of Family Size on Educational Attainment in China: Cohort Variations", Paper to be presented in Annual Meeting of the Population Association of America, Philadelphia, 31 March – 2 April, 2005.

Rosenbaum, Paul R., and Ronald Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41-55.

Rosenzweig Mark R., and Kenneth I. Wolpin (1980) "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, XLVIII, pp. 227-240.

Rosenzweig, M. R. and J. Zhang (2006), "Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's `One Child' Policy", Discussion Paper No. 2082, The Institute for the Study of Labor (IZA), Germany.

Rubin, Donald B. 1977. “Assignment to a Treatment Group on the Basis of a Covariate.” *Journal of Educational Statistics* 2(1): 1-26.

Rubin, Donald B. 1979. “Using Multivariate Sampling and Regression Adjustment to Control Bias in Observational Studies.” *Journal of the American Statistical Association* 74: 318–328.

Rubin, Donald B. 1980. “Bias Reduction Using Mahalanobis-Metric Matching.” *Biometrics* 36(2): 293–298.

Smith, J. and P. Todd. (2005), “Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators?”, *Journal of Econometrics*, 125(1–2), 305–353.

Truong, S. A., John Knodel, David Lam and Jed Friedman (1998), “Family size and childrent’s education in Vietnam”, *Demography*, Vol. 35(1), 57-70.

VietNamNet Bridge, 2007 “Vietnam’s Population Growth Under Control”, VietnamNet Newspaper, available at http://english.vietnamnet.vn/social/2007/07/714867/

Wooldridge, J. M. (2002), “Introductory Econometrics: A Modern Approach”.

Appendix

Table A.1: Variable description

Variables	Mean	Std. Dev.
The number of siblings	2.862	1.270
Age	12.79	2.99
Male (yes = 1)	0.509	0.500
Ethnic minorities (yes = 1)	0.166	0.372
Urban (yes = 1)	0.208	0.406
Red River Delta	0.191	0.393
North East	0.119	0.324
North West	0.035	0.183
North Central Coast	0.161	0.368
South Central Coast	0.095	0.293
Central Highlands	0.087	0.282
South East	0.142	0.349
Mekong River Delta	0.170	0.376
Per capita income (VND million)	7.119	8.154
Head with primary school degree	0.483	0.500
Head with lower-secondary school	0.312	0.463
Head with upper secondary school	0.089	0.284
Head with post-secondary school	0.116	0.320
Head living with spouse	0.079	0.269
Head’s spouse with primary school degree	0.509	0.500
Head’s spouse with lower-secondary school	0.266	0.442
Head’s spouse with upper secondary school	0.070	0.256
Head’s spouse with post-secondary school	0.075	0.264
Head’s gender (male = 1, female =0)	0.858	0.349
Head’s age	42.54	7.05
Spouse’s age	37.09	12.71

Table A.2: Probit regression of school enrolment

Explanatory variables	Probit regression				IV Probit regression			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
The number of siblings	-0.146***	-0.137***	-0.164***	-0.165***	-0.209***	-0.341***	-0.334*	-0.323*
	[0.013]	[0.016]	[0.022]	[0.022]	[0.047]	[0.086]	[0.193]	[0.167]
Age		-0.249***	-0.250***	-0.250***		-0.237***	-0.232***	-0.234***
		[0.010]	[0.012]	[0.012]		[0.013]	[0.026]	[0.022]
Male (yes = 1)		-0.154***	-0.156***	-0.157***		-0.191***	-0.187***	-0.185***
		[0.040]	[0.042]	[0.042]		[0.042]	[0.053]	[0.050]
Ethnic minorities (yes = 1)		-0.473***	-0.228***	-0.233***		-0.304***	-0.164	-0.170*
		[0.065]	[0.071]	[0.071]		[0.102]	[0.104]	[0.099]
Urban (yes = 1)		0.280***	0.043			0.183**	0.027	
		[0.058]	[0.063]			[0.072]	[0.065]	
Red River Delta	Omitted							
North East		0.233***	0.324***	0.343***		0.197**	0.295***	0.297***
		[0.088]	[0.094]	[0.086]		[0.088]	[0.100]	[0.100]
North West		0.001	0.152	0.173		0.013	0.138	0.14
		[0.111]	[0.117]	[0.111]		[0.108]	[0.116]	[0.115]
North Central Coast		-0.075	-0.029			0.035	0.016	
		[0.075]	[0.079]			[0.087]	[0.094]	
South Central Coast		-0.048	0.288***	0.308***		0.045	0.303***	0.299***
		[0.083]	[0.091]	[0.082]		[0.089]	[0.091]	[0.082]
Central Highlands		-0.079	0.102	0.127		0.141	0.197	0.191*
		[0.089]	[0.095]	[0.086]		[0.127]	[0.141]	[0.108]
South East		-0.385***	-0.157*	-0.132*		-0.267***	-0.123	-0.126*
		[0.076]	[0.086]	[0.076]		[0.092]	[0.095]	[0.076]
Mekong River Delta		-0.646***	-0.310***	-0.293***		-0.578***	-0.330***	-0.333***
		[0.066]	[0.076]	[0.065]		[0.074]	[0.078]	[0.076]
Birth order			0.156***	0.156***			0.299*	0.288**
			[0.029]	[0.028]			[0.162]	[0.140]
Head with primary school degree	Omitted							
Head with lower-secondary school			0.352***	0.352***			0.330***	0.331***
			[0.056]	[0.056]			[0.062]	[0.061]
Head with upper secondary school			0.635***	0.637***			0.611***	0.614***
			[0.111]	[0.111]			[0.116]	[0.114]
Head with post-secondary school			0.873***	0.882***			0.830***	0.836***
			[0.111]	[0.111]			[0.124]	[0.123]
Head living with spouse			-0.679**	-0.735***			-0.845**	-0.967***
			[0.288]	[0.201]			[0.344]	[0.311]

Explanatory variables	Probit regression				IV Probit regression			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Head's spouse with primary school degree	Omitted							
Head's spouse with lower-secondary school			0.251***	0.253***			0.210***	0.212***
			[0.064]	[0.064]			[0.080]	[0.077]
Head's spouse with upper secondary school			0.513***	0.517***			0.467***	0.471***
			[0.133]	[0.132]			[0.144]	[0.142]
Head's spouse with post-secondary school			0.557***	0.565***			0.508***	0.516***
			[0.138]	[0.137]			[0.150]	[0.148]
Head's gender (male = 1, female =0)			0.158*	0.147*			0.182**	0.164**
			[0.081]	[0.078]			[0.084]	[0.080]
Head's age			-0.001				-0.004	
			[0.005]				[0.006]	
Spouse's age			-0.014**	-0.015***			-0.017**	-0.019***
			[0.006]	[0.004]			[0.007]	[0.006]
Per capita income			0.032***	0.033***			0.030***	0.031***
			[0.007]	[0.007]			[0.008]	[0.008]
Constant	1.467***	5.191***	4.790***	4.782***	1.642***	5.520***	4.984***	4.976***
	[0.045]	[0.166]	[0.214]	[0.211]	[0.132]	[0.177]	[0.269]	[0.264]
Number of observations	8381	8381	8381	8381	8381	8381	8381	8381
R-squared	0.02	0.22	0.28	0.28				
Wald test of exogeneity					1.89	5.24	1.12	0.99
P-value of Wald test					0.169	0.022	0.289	0.320

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%
Source: Estimation from the 2006 VHLSS.

Table A.3. 1: Probit regression of educational performance "Excellent"

Explanatory variables	Probit regression				IV Probit regression			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
The number of siblings	-0.191***	-0.151***	-0.059*	-0.063**	-0.368***	-0.290***	-0.123	-0.155
	[0.021]	[0.022]	[0.032]	[0.030]	[0.053]	[0.099]	[0.199]	[0.141]
Age		-0.086***	-0.104***	-0.103***		-0.084***	-0.099***	-0.099***
		[0.007]	[0.009]	[0.008]		[0.008]	[0.018]	[0.010]
Male (yes = 1)		-0.345***	-0.357***	-0.358***		-0.367***	-0.368***	-0.374***
		[0.043]	[0.045]	[0.045]		[0.045]	[0.056]	[0.050]
Ethnic minorities (yes = 1)		-0.593***	-0.433***	-0.450***		-0.488***	-0.412***	-0.413***
		[0.094]	[0.095]	[0.089]		[0.122]	[0.116]	[0.106]
Urban (yes = 1)		0.451***	0.161***	0.157***		0.387***	0.156***	0.151***
		[0.049]	[0.055]	[0.054]		[0.067]	[0.057]	[0.055]
Red River Delta	Omitted							
North East		-0.135*	-0.157*	-0.130*		-0.155*	-0.165*	-0.153*
		[0.081]	[0.085]	[0.077]		[0.081]	[0.088]	[0.085]
North West		-0.09	-0.066			-0.093	-0.071	
		[0.154]	[0.161]			[0.152]	[0.162]	
North Central Coast		-0.255***	-0.218***	-0.194***		-0.188**	-0.203**	-0.186**
		[0.079]	[0.080]	[0.075]		[0.090]	[0.089]	[0.075]
South Central Coast		0.188**	0.337***	0.364***		0.246***	0.344***	0.362***
		[0.073]	[0.077]	[0.070]		[0.081]	[0.079]	[0.071]
Central Highlands		-0.133	-0.091			0.008	-0.056	
		[0.090]	[0.091]			[0.135]	[0.139]	
South East		0.191***	0.260***	0.289***		0.253***	0.271***	0.291***
		[0.074]	[0.080]	[0.072]		[0.085]	[0.085]	[0.072]
Mekong River Delta		0.047	0.236***	0.260***		0.067	0.226***	0.237***
		[0.069]	[0.077]	[0.068]		[0.070]	[0.082]	[0.078]
Birth order			-0.039	-0.037			-0.014	-0.033
			[0.037]	[0.032]			[0.168]	[0.111]
Head with primary school degree	Omitted							
Head with lower-secondary school			0.215***	0.223***			0.210***	0.206***
			[0.062]	[0.058]			[0.065]	[0.063]
Head with upper secondary school			0.309***	0.315***			0.304***	0.300***
			[0.086]	[0.083]			[0.088]	[0.087]
Head with post-secondary school			0.651***	0.652***			0.640***	0.626***
			[0.076]	[0.073]			[0.085]	[0.085]
Head living with spouse			-0.085				-0.15	
			[0.291]				[0.349]	

Explanatory variables	Probit regression				IV Probit regression			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Head's spouse with primary school degree	Omitted							
Head's spouse with lower-secondary school			0.005				-0.005	
			[0.064]				[0.071]	
Head's spouse with upper secondary school			0.353***	0.357***			0.339***	0.345***
			[0.092]	[0.082]			[0.101]	[0.085]
Head's spouse with post-secondary school			0.381***	0.386***			0.367***	0.370***
			[0.089]	[0.078]			[0.100]	[0.082]
Head's gender (male = 1, female =0)			-0.095	-0.076			-0.089	-0.053
			[0.078]	[0.063]			[0.082]	[0.072]
Head's age			0.002				0.001	
			[0.006]				[0.006]	
Spouse's age			-0.001				-0.002	
			[0.006]				[0.007]	
Per capita income			0.012***	0.011***			0.011***	0.011***
			[0.003]	[0.003]			[0.003]	[0.003]
Constant	-0.592***	0.431***	0.147	0.123	-0.073	0.772***	0.243	0.193
	[0.057]	[0.115]	[0.188]	[0.132]	[0.163]	[0.266]	[0.349]	[0.171]
R-squared	0.02	0.11	0.16	0.16				
Number of observations	7100	7100	7100	7100	7100	7100	7100	7100
Wald test of exogeneity					11.31	1.87	0.27	0.10
P-value of Wald test					0.001	0.171	0.606	0.757

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%
Source: Estimation from the 2006 VHLSS.

Table A.4: Probit regression of educational performance "Excellent and good"

Explanatory variables	Probit regression				IV Probit regression			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
The number of siblings	-0.166***	-0.107***	-0.058***	-0.057***	-0.352***	-0.076	-0.146	-0.125
	[0.014]	[0.015]	[0.022]	[0.015]	[0.043]	[0.080]	[0.150]	[0.133]
Age		-0.053***	-0.062***	-0.062***		-0.053***	-0.077***	-0.074***
		[0.006]	[0.007]	[0.006]		[0.006]	[0.012]	[0.008]
Male (yes = 1)		-0.390***	-0.396***	-0.395***		-0.384***	-0.354***	-0.357***
		[0.033]	[0.034]	[0.034]		[0.037]	[0.048]	[0.045]
Ethnic minorities (yes = 1)		-0.684***	-0.543***	-0.541***		-0.705***	-0.602***	-0.600***
		[0.056]	[0.059]	[0.058]		[0.079]	[0.071]	[0.072]
Urban (yes = 1)		0.468***	0.250***	0.243***		0.481***	0.262***	0.253***
		[0.042]	[0.047]	[0.046]		[0.052]	[0.046]	[0.046]
Red River Delta	Omitted							
North East		-0.106*	-0.116*	-0.112**		-0.101	-0.086	-0.081
		[0.061]	[0.063]	[0.056]		[0.063]	[0.067]	[0.058]
North West		-0.403***	-0.385***	-0.382***		-0.401***	-0.363***	-0.364***
		[0.098]	[0.100]	[0.095]		[0.098]	[0.102]	[0.095]
North Central Coast		-0.289***	-0.263***	-0.258***		-0.302***	-0.304***	-0.294***
		[0.058]	[0.059]	[0.052]		[0.067]	[0.065]	[0.057]
South Central Coast		-0.291***	-0.155**	-0.151***		-0.304***	-0.174***	-0.169***
		[0.061]	[0.065]	[0.055]		[0.069]	[0.066]	[0.055]
Central Highlands		-0.380***	-0.341***	-0.336***		-0.411***	-0.446***	-0.431***
		[0.068]	[0.070]	[0.060]		[0.103]	[0.101]	[0.091]
South East		-0.101	-0.002			-0.115	-0.036	
		[0.061]	[0.067]			[0.071]	[0.071]	
Mekong River Delta		-0.184***	-0.013			-0.189***	0.018	
		[0.056]	[0.062]			[0.057]	[0.065]	
Birth order			-0.001				-0.169	-0.149
			[0.026]				[0.124]	[0.100]
Head with primary school degree	Omitted							
Head with lower-secondary school			0.185***	0.187***			0.200***	0.203***
			[0.045]	[0.044]			[0.045]	[0.044]
Head with upper secondary school			0.300***	0.300***			0.314***	0.316***
			[0.068]	[0.067]			[0.068]	[0.067]

Explanatory variables	Probit regression				IV Probit regression			
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Head with post-secondary school			0.588***	0.593***			0.614***	0.621***
			[0.066]	[0.065]			[0.068]	[0.068]
Head living with spouse			0.059				0.262	
			[0.225]				[0.265]	
Head's spouse with primary school degree	Omitted							
Head's spouse with lower-secondary school			0.065	0.069			0.100*	0.085*
			[0.046]	[0.043]			[0.052]	[0.044]
Head's spouse with upper secondary school			0.358***	0.360***			0.395***	0.378***
			[0.076]	[0.074]			[0.079]	[0.075]
Head's spouse with post-secondary school			0.472***	0.463***			0.512***	0.494***
			[0.084]	[0.080]			[0.086]	[0.083]
Head's gender (male = 1, female =0)			0.064				0.04	
			[0.066]				[0.067]	
Head's age			-0.005	-0.004			-0.002	
			[0.004]	[0.003]			[0.005]	
Spouse's age			0.001				0.004	
			[0.005]				[0.005]	
Per capita income			0.010***	0.010***			0.012***	0.012***
			[0.004]	[0.004]			[0.004]	[0.004]
Constant	0.484***	1.348***	1.091***	1.147***	1.000***	1.271***	0.772***	0.879***
	[0.041]	[0.092]	[0.145]	[0.116]	[0.121]	[0.220]	[0.281]	[0.163]
R-squared	0.02	0.09	0.12	0.13				
Number of observations	7100	7100	7100	7100	7100	7100	7100	7100
Wald test of exogeneity					18.37	0.16	1.64	1.87
P-value of Wald test					0.000	0.692	0.201	0.171

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%
Source: Estimation from the 2006 VHLSS.

Table A.5. First-stage regressions of the number of siblings

Explanatory variables	In IV-Probit regressions of school enrolment				In IV-Probit regressions of educational performance: drop children not attending school				
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4 regression performance "Excellent"	Model 4 regression performance "Excellent and good"
The average number of siblings in district	0.677***	0.423***	0.217***	0.245***	0.726***	0.465***	0.242***	0.322***	0.272***
	[0.041]	[0.038]	[0.024]	[0.024]	[0.039]	[0.041]	[0.024]	[0.023]	[0.025]
Age		0.022***	0.088***	0.085***		0.011**	0.075***	0.043***	0.042***
		[0.005]	[0.004]	[0.004]		[0.005]	[0.004]	[0.003]	[0.003]
Male (yes = 1)		-0.200***	-0.186***	-0.185***		-0.187***	-0.182***	-0.180***	-0.188***
		[0.027]	[0.019]	[0.019]		[0.028]	[0.019]	[0.019]	[0.019]
Ethnic minorities (yes = 1)		0.621***	0.311***	0.322***		0.575***	0.266***	0.292***	0.301***
		[0.050]	[0.036]	[0.035]		[0.051]	[0.037]	[0.034]	[0.037]
Urban (yes = 1)		-0.367***	-0.073***			-0.350***	-0.056**	-0.039	-0.079***
		[0.032]	[0.024]			[0.033]	[0.024]	[0.025]	[0.024]
Red River Delta	Omitted								
North East		-0.155***	-0.155***	-0.231***		-0.167***	-0.145***	-0.206***	-0.116***
		[0.042]	[0.032]	[0.031]		[0.043]	[0.032]	[0.030]	[0.029]
North West		-0.056	-0.114**	-0.207***		-0.143*	-0.132**		-0.065
		[0.073]	[0.058]	[0.056]		[0.076]	[0.057]		[0.054]
North Central Coast		0.360***	0.184***			0.289***	0.144***	0.026	0.133***
		[0.048]	[0.034]			[0.047]	[0.033]	[0.033]	[0.032]
South Central Coast		0.319***	0.065*	-0.038		0.267***	0.042	-0.054	0.052*
		[0.048]	[0.034]	[0.033]		[0.050]	[0.035]	[0.033]	[0.031]
Central Highlands		0.690***	0.389***	0.268***		0.645***	0.364***		0.382***
		[0.063]	[0.046]	[0.045]		[0.066]	[0.048]		[0.046]
South East		0.372***	0.130***	0.017		0.310***	0.099***	0.002	
		[0.047]	[0.037]	[0.036]		[0.050]	[0.039]	[0.036]	
Mekong River Delta		0.183***	-0.131***	-0.217***		0.110**	-0.151***	-0.199***	
		[0.041]	[0.031]	[0.030]		[0.045]	[0.032]	[0.030]	
Birth order			0.827***	0.822***			0.821***	0.748***	0.742***
			[0.011]	[0.011]			[0.012]	[0.011]	[0.011]
Head with primary school degree	Omitted								
Head with lower-secondary school			-0.087***	-0.087***			-0.069***	-0.140***	-0.073***
			[0.025]	[0.025]			[0.025]	[0.025]	[0.026]
Head with upper secondary school			-0.095***	-0.102***			-0.072**	-0.133***	-0.082**
			[0.034]	[0.034]			[0.034]	[0.034]	[0.035]

Explanatory variables	In IV-Probit regressions of school enrolment				In IV-Probit regressions of educational performance: drop children not attending school				
	Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4 regression performance "Excellent"	Model 4 regression performance "Excellent and good"
Head with post-secondary school			-0.177***	-0.205***			-0.147***	-0.242***	-0.194***
			[0.032]	[0.031]			[0.032]	[0.031]	[0.032]
Head living with spouse			-1.050***	-1.476***			-1.010***		
			[0.126]	[0.090]			[0.129]		
Head's spouse with primary school degree	Omitted								
Head's spouse with lower-secondary school			-0.202***	-0.212***			-0.161***		-0.081***
			[0.027]	[0.028]			[0.027]		[0.025]
Head's spouse with upper secondary school			-0.219***	-0.235***			-0.189***	-0.103***	-0.114***
			[0.038]	[0.038]			[0.037]	[0.033]	[0.036]
Head's spouse with post-secondary school			-0.226***	-0.246***			-0.205***	-0.149***	-0.207***
			[0.036]	[0.036]			[0.035]	[0.031]	[0.033]
Head's gender (male = 1, female =0)			0.138***	0.107***			0.103***	0.236***	
			[0.032]	[0.030]			[0.033]	[0.025]	
Head's age			-0.012***				-0.011***		
			[0.003]				[0.003]		
Spouse's age			-0.019***	-0.028***			-0.019***		
			[0.003]	[0.002]			[0.003]		
Per capita income			-0.008***	-0.009***			-0.006***	-0.007***	-0.007***
			[0.002]	[0.002]			[0.001]	[0.001]	[0.001]
Constant	1.279***	1.447***	0.943***	0.917***	1.103***	1.493***	0.994***	0.062	0.313***
	[0.093]	[0.100]	[0.091]	[0.092]	[0.087]	[0.106]	[0.093]	[0.078]	[0.074]
Number of observations	8382	8382	8382	8382	7100	7100	7100	7100	7100
R-squared	0.09	0.16	0.59	0.58	0.1	0.16	0.61	0.59	0.59

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%
Source: Estimation from the 2006 VHLSS.

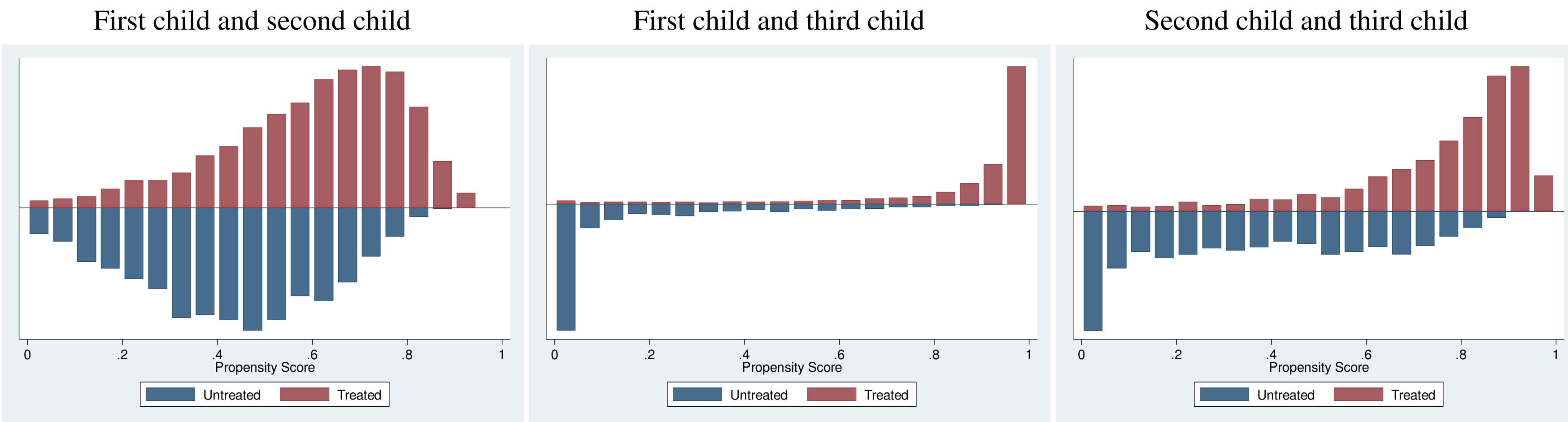
Table A.6. Multinomial logit regression of birth order

Explanatory variables	Small model			Large model		
	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
P(birth order = 2)						
The number of siblings	0.968	0.050	0.000	0.996	0.051	0.000
Age	-0.256	0.014	0.000	-0.261	0.014	0.000
Male (yes = 1)	0.141	0.060	0.019	0.143	0.060	0.018
Ethnic minorities (yes = 1)	-0.199	0.100	0.048	-0.104	0.104	0.317
Urban (yes = 1)	-0.052	0.078	0.499	-0.070	0.082	0.391
Red River Delta	Omitted					
North East	0.221	0.105	0.036	0.256	0.106	0.016
North West	0.181	0.166	0.275	0.252	0.168	0.134
North Central Coast	-0.202	0.108	0.061	-0.197	0.108	0.069
South Central Coast	-0.272	0.113	0.016	-0.163	0.116	0.163
Central Highlands	-0.368	0.128	0.004	-0.283	0.131	0.030
South East	-0.268	0.114	0.019	-0.156	0.119	0.193
Mekong River Delta	-0.092	0.096	0.337	0.075	0.107	0.485
Head with primary school degree	Omitted					
Head with lower-secondary school				0.232	0.082	0.004
Head with upper secondary school				0.052	0.121	0.665
Head with post-secondary school				0.056	0.116	0.630
Head living with spouse	3.867	0.436	0.000	3.990	0.443	0.000
Head with primary school degree	Omitted					
Head with lower-secondary school				0.149	0.087	0.085
Head with upper secondary school				0.190	0.127	0.134
Head with post-secondary school				0.059	0.136	0.664
Head's gender (male = 1, female =0)	0.120	0.112	0.284	0.116	0.115	0.312
Head's age	0.075	0.010	0.000	0.075	0.010	0.000
Spouse's age	0.090	0.010	0.000	0.091	0.010	0.000
Per capita income (VND million)				0.003	0.004	0.380
_cons	-5.731	0.293	0.000	-6.001	0.302	0.000
P(birth order = 3)						
The number of siblings	2.359	0.070	0.000	2.389	0.072	0.000
Age	-0.515	0.019	0.000	-0.527	0.019	0.000
Male (yes = 1)	0.463	0.086	0.000	0.470	0.086	0.000
Ethnic minorities (yes = 1)	-0.444	0.150	0.003	-0.300	0.153	0.051
Urban (yes = 1)	-0.112	0.110	0.310	-0.047	0.118	0.690
Red River Delta	Omitted					
North East	0.464	0.158	0.003	0.559	0.159	0.000
North West	0.343	0.261	0.188	0.492	0.262	0.061
North Central Coast	-0.222	0.154	0.149	-0.231	0.155	0.135
South Central Coast	-0.145	0.154	0.348	0.029	0.159	0.857
Central Highlands	-0.664	0.180	0.000	-0.507	0.183	0.006
South East	-0.377	0.157	0.016	-0.196	0.164	0.232
Mekong River Delta	0.067	0.140	0.633	0.323	0.152	0.034
Head with primary school degree	Omitted					
Head with lower-secondary school				0.377	0.114	0.001
Head with upper secondary school				0.044	0.171	0.799
Head with post-secondary school				-0.185	0.175	0.291

Explanatory variables	Small model			Large model		
	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
Head living with spouse	7.648	0.582	0.000	7.866	0.595	0.000
Head with primary school degree	Omitted					
Head with lower-secondary school				0.382	0.121	0.002
Head with upper secondary school				0.015	0.208	0.941
Head with post-secondary school				-0.370	0.217	0.088
Head's gender (male = 1, female =0)	0.149	0.151	0.325	0.071	0.157	0.652
Head's age	0.140	0.013	0.000	0.139	0.013	0.000
Spouse's age	0.171	0.013	0.000	0.175	0.013	0.000
Per capita income (VND million)				0.010	0.005	0.023
_cons	-13.671	0.425	0.000	-14.090	0.448	0.000
Pseudo R2	8382			8382		
Number of obs	0.307			0.311		

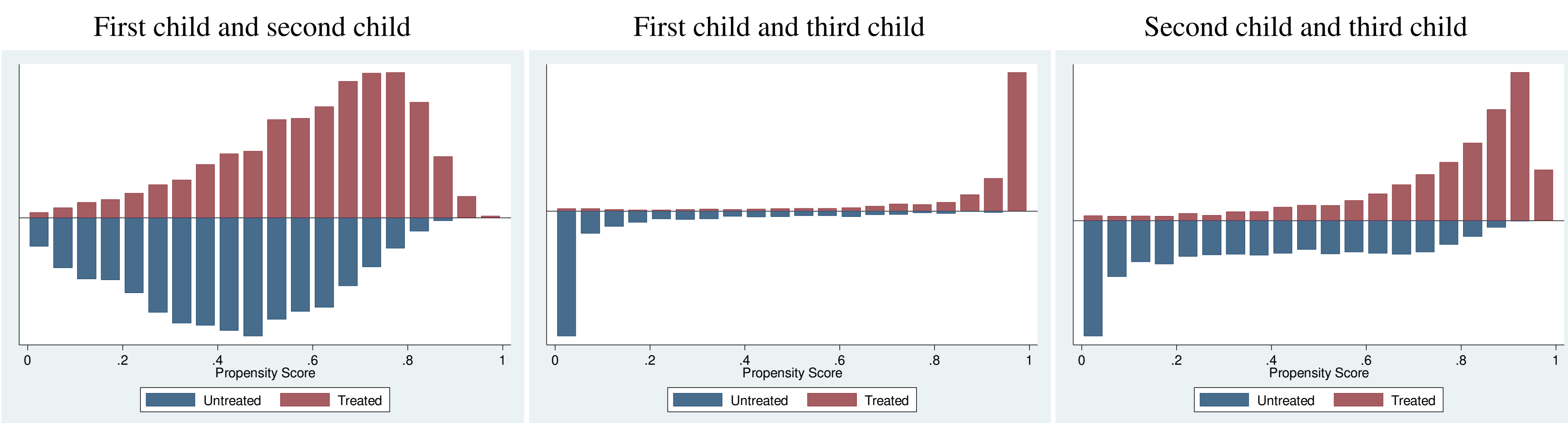
Estimates are corrected for sampling weights and cluster correlation.
Source: Estimation from the 2006 VHLSS.

Figure A.1. Propensity scores estimated from the small model



Source: Estimation from the 2002 VHLSS

Figure A.2: Propensity scores estimated from the large model



Source: Estimation from the 2002 VHLSS

Table A.7. The impact of birth order using propensity score matching

Outcomes	Small model			Large model		
	Y_O	Y_C	ATT = $Y_O - Y_C$	Y_O	Y_C	ATT = $Y_O - Y_C$
The impact of being "first child" relative to being "second child"						
School enrolment rate	84.65***	86.46	-1.81	84.65***	89.28***	-2.63*
	[0.69]	[2.21]	[2.15]	[0.69]	[1.92]	[1.51]
Performance: "excellent" rate	16.06***	14.21	1.85	16.06***	10.69***	5.37**
	[0.80]	[2.13]	[1.81]	[0.80]	[1.86]	[2.12]
Performance: "excellent and good" rate	55.08***	49.87	5.21**	55.08***	49.54***	5.54**
	[1.09]	[2.47]	[2.61]	[1.09]	[2.98]	[2.65]
The impact of being "first child" relative to being "third child"						
School enrolment rate	84.65***	87.13***	-2.48	84.65***	88.45***	-3.8*
	[0.69]	[2.23]	[2.01]	[0.69]	[17.72]	[2.42]
Performance: "excellent" rate	16.06***	8.37***	7.69**	16.06***	9.14**	6.92**
	[0.80]	[1.92]	[3.13]	[0.80]	[4.49]	[3.16]
Performance: "excellent and good" rate	55.08***	42.94***	12.14***	55.08***	40.89***	14.19**
	[1.09]	[3.45]	[4.01]	[1.09]	[13.25]	[6.15]
The impact of being "second child" relative to being "third child"						
School enrolment rate	84.97***	85.36***	-0.39	84.97***	86.06***	-1.09
	[0.69]	[6.60]	[2.64]	[0.69]	[9.61]	[1.85]
Performance: "excellent" rate	14.23***	8.64**	5.59*	14.23***	11.62***	2.61
	[0.79]	[3.94]	[3.25]	[0.79]	[4.06]	[2.12]
Performance: "excellent and good" rate	51.49***	45.38***	6.11	51.49***	47.09***	4.4*
	[1.17]	[9.49]	[4.95]	[1.17]	[8.75]	[2.35]

Robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

Source: Estimation from the 2006 VHLSS