Discussion Paper Series

2009 – 3

Department of Economics
Royal Holloway College
University of London
Egham TW20 0EX

Separate Effects of Sibling Gender and Family Size on Educational Achievements - Methods and First Evidence from Population Birth Registry

Yen-Chien Chen, Stacey H. Chen, and Jin-Tan Liu*

February 9, 2009

Abstract

Son-preferring parents tend to continue to have babies until a son's birth. After deciding the set of children, the parents with resource constraints may divert family sources from daughters to a son. Thus, the presence of a son, relative to a daughter, have 2 distinct effects on his sister's educational outcomes: the *direct effect* while holding constant family size and the *indirect effect* through decreasing family size. Previous estimates of the direct effect take family size as an *exogenous* and *predetermined* covariate, and assume the indirect effect to be captured by the main effect of family size. However, family

*Y.C. Chen: Ph.D. candidate of National Taiwan University; S.H. Chen: Lecturer at the Department of Economics in Royal Holloway University of London; J.T. Liu: Professor of National Taiwan University and NBER Research Associate. Josh Angrist, Analia Schlosser, Michael Greenstone, Helen Hsi, Jessica Pan and seminar participants at the 2007 Conference of the Society of Labor Economists and the seminar at the Institute of Economics of Academia Sinica provided helpful comments. We thank the Ministry of the Interior Affairs and the Ministry of Education Affairs of Taiwan for providing administrative population data of birth registry and college entrance examinations. We thank the National Science Council of Taiwan (J.T. Liu and Y.C. Chen), and National Health Research Institute (NHRI-EX96-9622PI) for financial support.

size is endogenous and dependent on the sex composition of early-born siblings. We show that *even if child gender and family size are both exogenous*, use of an instrument for family size is required to isolate the direct effect from the main effects of family size. Using a large and unique administrative data from Taiwan, we demonstrate how Instrumental-Variable Methods resolve both problems of endogeneity and causal dependence of an important covariate (family size) on treatment status (sibling sex). Furthermore, we minimize the incident of sex-selective abortion by restricting our birth data on cohorts prior to abortion legalization and prior to prevalent practice of prenatal sex determination. Using the occurrence of twining to instrument for family size conditional on birthweights, our IV estimates show a strong direct effect of a male sibling, relative to a female, on women's college attainment, if the women were born in the earliest year of our data, 1978. After 1978, both effects of sibling gender and family size are almost zero.

JEL Classification: I20, J13, J16, J24, O10, R20 **Keywords**: Sibling sex composition, family size, intrafamily allocation of resources; quantity-quality trade-off; education

Economists and policymakers have long been interested in how family environment affects children's educational achievements. Despite years of studies, evidence on the components of the home production function for children's human capital is still limited. Many empirical challenges arise because of data limitations, such as lack of credible measures for completed family size (the number of siblings), sibling sex composition, or the level of education. Even if data limitations can be overcome, empirical analysis is subject to problems of multiple endogenous factors in human capital formation. This problem can be prevalent if the endogenous factors are highly correlated.

Sibling sex composition and family size are two of the most important determinants in human capital formation. Conditional on family size, son-preferring parents may divert resources to sons from daughters, if they have time or financial constrains (Behrman, Pollak and Taubman 1982, Thomas 1990). Thus, daughters may have less education in the presence of a brother, conditional on family size. This relationship between sibling gender and child outcomes is referred to as "sibling rivalry" in the previous literature (e.g., Parish and Willis 1993; Butcher and Case 1994; Kaestner 1997; Garg and Morduch 1998; Morduch 2000). However, parents may be less motivated to have another child after having a son. If there is a trade-off between children's quality and quantity, as suggested in Becker (1960), Becker and Lewis (1973), Blake (1981), Hanushek (1992), Powell and Steelman (1993), the son's presence may lead to a reduction in family size, which potentially offer a better family environment for his female siblings. .

It is challenging to separate the causal effects of sibling gender and family size on children's education. First, if ultrasound and sex-selective abortion technology are available to son-preferring parents, they may select child gender of their favor and thus child gender can be endogenous. Second, even if child gender is assigned randomly, family size causally depends on the sex composition of earlier-born siblings. In particular, family size decreases with the presence of a son. As a result, the effect of having a male sibling is confounded by the effect of decreasing family size. Given a common approach which controls for family size to estimate sibling gender effect on educational outcome, some fraction of sibling gender effect is "*indirect*," which is mediated through its impact on family size, and some fraction of sibling gender effect is "*direct*," which is operated through a pathway that does not involve family size. Therefore, the common approach can identify neither the direct sibling-gender effect nor the family size effect. Third, birth order can affect children's educational outcome in a manner that confounds the observed effect of family size on the outcome. Black, Devereux and Salvanes (2005) has shown that omitting birth order in the formation function of human capital can overstate the effect of family size. These issues have long imposed difficulties to interpret previous results regarding sibling rivalry in an economy where sex bias is prevalent.

In addition to these challenges, identification of the effect of family size needs to resolve the problem of endogenous family size. The observed trade-off between child quality and quantity can be driven primarily by unobserved family background or parental preference, not by the number of children, as shown by a long list of prior studies on quality-quantity trade-offs.[1]

By addressing all of these four issues, we aim to isolate the direct effect of sibling gender, from the effect of family size, on children's educational achievement. In Section 1 and Appendix 5, we show how the fact that family size is a function of sibling sex composition, alone, requires an instrument for family size, even if child gender and family size are both exogenous. This proposition that direct and indirect effects are not identified even in a randomized experiment in the absence of additional assumptions has been pointed out by statisticians and epidemiologists, e.g., Pearl (2000) and Robins and Greenland (1992), though not formally noted by the economics profession. We exploit plausibly exogenous changes in family size due to twin births to instrument for family size, following Black, Devereux and Salvanes (2005) and

[1] Important examples include Rosenzweig and Wolpin (1980), Caceres (2004), Black, Devereux and Salvanes (2005), Angrist, Lavy and Schlosser (2005), Conley and Glauber (2006), and Qian (2008).

Angrist, Lavy and Scholsser (2006). We take account of the effects associated with low birth weights of twins, as in Rosenzweig and Zhang (2006). Using the part of exogenous variation in family size as a replacement for endogenous family size, we show that the direct effect of sibling gender can be identified because the indirect effect transmitted trough family size has been removed.

Our data work aims to minimize the incident of sex-selective abortion. As introduced in Section 2, our analysis focuses on the earliest cohort of firstborn children in the data, who were born prior to the legalization of abortion and prior to the widespread of technology of sex-selective abortion. By doing so, we ensure the sex ratios of boys to girls among the firstborn sample is as balanced as those of the early 1970s when sex-testing methods were not yet available. Moreover, the second-born siblings of the firstborn population also exhibit arguably normal sex ratios, within the standard range of 1.05 to 1.07, as documented in demographics literature, by, e.g., James (1985) and Johansson and Nygren (1991).

Our contributions are made possibly by a unique data link, which matches two national administrative records: the College Entrance Tests and the Birth Registry records, of the entire Taiwan. Because the Birth Registry records covers over two decades, we are able to construct accurate measures of sibship size, birth order, sibling sex composition, and long-term educational outcomes of all children. Consequently, we overcome limitations of previous studies that used cross-sectional data or short panel surveys, where information about the number and sex composition of siblings could not be directly or fully observed.

Our efforts of resolving the previous identification issues make substantial differences, as shown in Section 3. Unlike the estimates emerged by using the commonly used methods, we find little evidence of rivalry effects of male siblings on women's college attainment, especially in recent years. This finding is particularly striking because Taiwan historically had stronger son preference than other areas and because this tradition remains extraordinarily strong in recent years. Mothers in Taiwan with two daughters are nearly 30 percentage points more likely to have a third child, than those with two sons. The same difference is less than 2 percentage points in both U.S. and Israel (Ben-Porath and Welch 1976; Angrist and Evans 1998; Angrist et al. 2006). Even with extraordinarily strong son preference in Taiwan, perhaps surprisingly, we little evidence of rivalry effects of male siblings on firstborn women's college attainment, except for the 1978-cohort.

Section 4 summeriness our interpretations. Given the strong preference for boys in Taiwan, the lack of evidence of the rivalry effects of male siblings on daughters' education may be not enough to rebuke the possibility of unequal treatments between daughter and son, but it sheds light on spillover effects of the presence of a son on his sisters. Recent studies have suggested that a son's birth is associated with changes in parental behaviors, such as increasing marital stability, increasing fathers' labor supply, and decreasing in maternal employment, all of which can offset the rivalry effects (if any). Additionally, we also find no evidence of family size effects on children's college enrollment, which echoes the recent studies on quality-quantity trade-offs (see footnote 1). Section 5 concludes.

1 Identifications

The evidence of sibling rivalry in literature is mixed. Butcher and Case (1994) suggested that a son's birth may increase daughters' education, probably because the son's masculine traits can help the sisters develop assertive attitudes toward greater success. In contrast, Garg and Morduch (1998) and Morduch (2000) in their pioneering work on sibling rivalry have noted that child health in certain African regions can be worsened by a shift from a scenario where all siblings are sisters to one where all are brothers. Kaestner (1997) similarly found that schooling levels of African Americans are negatively associated with the presence of a brother. Using Taiwanese data, Parish and Willis (1993) suggested that the rivalry effect of male siblings can be particularly strong on women's than on men's educational outcomes. Restricted by demanding data, these studies cannot cope with all the potential issues which have been pointed out in previous section. Our identifications strategies are described as follows.

1.1 Issue of Birth Order

We first address the issue of birth order. Because children with different birth orders have different endowments, the effect of younger siblings on children's education may be different from the effect of older siblings. Pooling children with different birth order together may contaminate the sibling gender effect[2]. Moreover, birth order has been proven as important factor to explain the variation in children's educational levels (e.g., Black et al. 2005). To control for birth order, our analysis focuses on the firstborn girl's (or boy's) educational outcome (Y), taking the subsequent sibling's male indicator ($B2$) and family size (N) as "treatments." By controlling for birth order, we test if the firstborn's college enrollment is lowered by the presence of a second-born brother, relative to a second-born sister.

1.2 Issue of Dependence of Family Size on Sibling Gender

In this subsection, we assume that child gender and family size are randomized assignment. The direct sibling gender effect on children's educational outcome is defined and measured by holding all predetermined covariates constant, and it is not mediated by other variable. However, family size itself is a potential outcome variable, not predetermined covariate, to (earlier-born) siblings' gender. It has been suggested that child gender, especially for earlier-born children, affect family size.[3] Therefore, with controlling for family size, indirect sibling gender effect is generated by the path

[2]We mimic the models in three previous studies as benchmark, the results are shown in Appendix 1 and Appendix tables

[3]The dependence of family size on child gender has been suggested by early work in Ben-Porath and Welch (1976) using 1970 U.S. census. Using more recent data, Dahl and Moretti (2008) also suggested that fertility choices and marital status can be driven by parental gender bias. Angrist and Evans (1998) have shown empirical evidence of parental preference for variety in offspring sex.

transmitted through family size to children's educational outcome. Figure **??** represents the relationships among educational outcome (Y), sibling gender ($B2$) and family size (N). We can see, in such case, it is difficult to isolate direct sibling gender effect from total estimated effect. Moreover, main family size may be also annoyed by indirect sibling gender effect.

To remove the problems caused by intermediate family size, a set of instrument variables(covariates) Z for family size is necessary (Pearl2005). Z must satisfy the following conditions: (1) Conditional on other covariates X, Z is independent of $B2$ and Y; (2) Z is an important explanatory variable for family size N; (3) N is determined by Z, $B2$ and X. Because that Z is independent of $B2$ and Y, the causal link form $B2$ to N is removed when using instrument Z for N. Figure **??** shows this implication. By doing this way, the indirect sibling gender is zero, and direct sibling gender effect and main family size effect can be separated. We should highlight that even family size is randomized, instrument variable Z is required to estimate the effects of direct sibling gender and family size on children's educational outcome. This paper uses the occurrence of twin birth at second birth as instrument for family size. The detailed is listed below.

1.3 Issue of Endogenous Family Size

The previous subsection has established that identification of the direct sibling gender effect and family size effect requires instrumental variable, given the condition where family size is exogenous. Nevertheless, recent literature on quality-quantity trade-offs has documented that family size is endogenous. Family size is the choice of parents, based on unobserved factors, such as genes or preference which may be correlated with children's educational outcome. The problem of endogenous family size leads the estimate of family size to noncausal interpretation. Conceptually, using a valid instrument variable to capture exogenous variation in family size, can remove this problem. Most of recent studies (Rosenzweig and Wolpin 1980; Caceres 2004; Black, Devereux and Salvanes 2005; Angrist, Lavy and Schlosser 2005; Conley and Glauber

2006; Qian 2008), using twins or sibling gender composition as instruments for family size, finds little evidence of quality-quantity trade-off. One exception is important work by Qian (2008). Using One-Child Policy in China as exogenous variation of family size, she finds that one additional child may increase school enrollment of first-born children.

The conditions for the valid instrument used to address endogenous family size includes condition (1) and (2) listed in previous subsection, and (3*) N is determined by $Z, B2, X$ and an unobserved error u, where u is independent of Y and $B2$. Following the strategies in Black et al. (2005) and Angrist et al. (2006), we exploit the occurrence of twins[4] at second birth as the instrument Z_i to measure the exogenous variation in family size[5], and propose Two-stage Least square Model. Ideally, substituting predicted family size by first-stage estimation for N in outcome regression (second-stage estimation) can identify effects of direct sibling gender and family size, as Figure **??** represents.

So far, in short, we use instrument variable method to solve 2 problems to identify direct sibling gender effect and causal family size effect separately, and investigate how the gender of subsequent sibling and the number of siblings affect first-born children's educational outcome.

1.4 Issues of Endogenous Child Gender

In the above subsections, we assume that the gender of second-born children is random. However, children's gender may have been determined endogenously, especially for economies with pro-male bias, while previous studies have had little concern about

[4] Rosenzweig and Wolpin (1980) first provided the idea of using twin birth as exogenous variation for family size. Because that their sample was very small, they used the ratio of the numbers of twin births to the number of total births as family size's instrument.

[5] The strategy of selecting subjects for method of twins IV follows Black et al. 2005. Angrist et al. 2006: the analysis sample is restricted to singleton first-born children who were born prior to twin at second birth. The reason of this strategy is that selection problems may arise, because that parents who choose to get additional child after the twin birth may have different concern from those who choose to get additional child after a consecutive singleton birth.

this issue. Endogenous child gender happens when parents with a strong son preference select their fetus' gender by resorting to sex-selective abortion under the legalized abortion law. In general, the more imbalanced the sex ratio of males to female is, it is more likely to exhibit the prevalent sex-selection practices.[6]

On one hand, if the gender of first-born children (subjects) is selected by parents, the sibling gender effect on children's educational outcome might be *understated*. That is, the "surviving" first-born girls are more likely to have parents who desired girls, and therefore, their subsequent brothers are less likely to hurt their education. On the other hand, if the gender of second-born children ($B2$) is selected, the sibling gender effect might be *overstated*. That is, boy-preferring parents, who give birth to a female first, may opt to have a son at the second birth and allocate more resources to the (second-born) son than their first-born daughter.

As an attempt to address the potential issues of sex selection, we restrict our data to the pre-1985 firstborn subpopulation. Prior to 1985, abortion was illegal in Taiwan, and access to technology of prenatal sex-selective abortion was limited.[7] Although not testable, the practices of prenatal sex selection in our pre-1985 data are not as prevalent as those after 1986, as suggested in Lin and Luoh (2008) and Lin, Liu and Qian (2008). Sex ratios at birth in Taiwan have ranged between 1.06 and 1.07 during the early 1970s, while technology of prenatal sex testing was not yet available. Although prenatal sex-testing methods (e.g. ultrasound) were initially introduced into Taiwan during the early 1980s, the facilities and technologies were not widespread till 1986. After 1986, sex imbalance became evident in Taiwan, and

[6] The abnormal masculinity of sex ratios in favor of males in Asia has been studied intensively in the past few decades. This skewed sex ratio reflects a high number of missing females in Asia (Sen 1990; Coale 1991)

[7] Abortion laws was first enacted in the end of 1984 and came into force in 1985. As Lin, Liu and Qian (2008) noted, although prenatal sex-testing methods, such as ultrasound, have been introduced into Taiwan during the early 1980s, the facilities and technologies started to be widespread only after 1986. As their Figure 1A showed, the sex ratios of firstborn population were very stable over the entire 1980s in Taiwan. Imbalanced sex ratios started to arise after 1986 only for the third-born population.

mostly arose at higher parity.

As Table 1 shows, the sex ratio at birth during 1978-84 were between 1.041 and 1.047, below the usual benchmark (1.05) of the "natural" sex ratio in Asia (e.g. James 1985; Johnasson and Nygren 1991; Coale and Bainster 1994; Das Gupta and Shuzhuo 1999; Hudson and den Boer 2005). It is worth noting that there is no evidence of female infanticide in Taiwan; mortality rates for infants and children are very low.[8] Consequently, we do not need to worry about the possibility of understating sibling gender effect induced by sex-selective abortions.

While the sex ratio at birth for the pre-1985 firstborn population is seemingly normal compared to the historical benchmark, their subsequent siblings, however, have some sign of sex imbalance. This is probably because some have been born after the legalization of abortion laws that came into force in 1985. Even so, Table 1 shows that the sex ratios of the second-born siblings during the pre-1985 period were between 1.066 and 1.067, still marginally within the range of the benchmark sex ratios during the early 1970s. To deal with gender endogeneity of second-born children with more caution, we employ two tests to check whether sex-selective abortions may haven been prevalent in case of second-born children. Contents and results of those tests are presented in the section 3.

2 Data and Descriptive Statistics

2.1 Administrative data and Sample Construction

For effective use of the proposed empirical methods in the previous sections, a credible and rigorous data set is a prerequisite. Our data set is constructed by linking the Birth Registry to the College Entrance Test (CET) records of the entire Taiwan. The Birth Registry covers all of the 7,053,190 births that took place between 1978 and

[8]Taiwan's infant mortality rate (death within one month) was about 0.004 in the early 1980's, and that also declined steadily over time, to under 0.0025, after 1990. In some years, mortality rate for females is even lower than that for males.

1999. It includes information on family backgrounds and individual demographics, such as birth date, birth place, birth order, birth weight, duration of pregnancy in weeks, and parental age and education. The College Entrance Test records include two sets of test scores: the SAT tests, conducted in February during the high school senior year, and the college enrollment tests, conducted in July after high school graduation. In this study we use the SAT tests, which are required for high school graduation. The CET data contain the SAT scores of all high school seniors during the period of 1996 to 2003.

We link all birth records by mothers' unique identification numbers in order to create a mother-based birth registry, which provides an accurate measure for each child's birth order. As mentioned above, We restrict the data coverage to mothers who had their first child prior to 1985, between 1978 and 1984, and track if they gave birth to more children subsequently over a two decade period, from 1978 to 1999. By following this way, we are able to obtain accurate data of sex-composition of children and completed family size.

Our data shows that mothers who gave birth to their first child between 1978 and 1984 had no additional children after 1997, implying that our sex-composition and family size are very precise. Besides, we exclude a small number of mothers who born their first child before being 15 year-old, or after the age of 50. We also exclude children who are twins at first birth, and those who are from one-child families. Our resultant data set contains 2,393,874 children from 893,156 families with at least 2 children.

We work with first-born singleton births and separate them into two cohorts; those born between 1978 and 1979, and those born between 1980 and 1984. We put the dividing line at 1980 because ultrasound technology was available to parents after 1980.[9] Table 1 shows that our data set includes 265,284 and 627,872 subjects in the

[9] Ultrasound was introduced to Taiwan during the early 1980's, a time when abortion was still illegal. Ultrasound was used to monitor a mother's and her fetus' health. Eugenics Protection Law was relaxed on Jan. 1, 1985, after which abortion became legal, though only in instances where the fetus had a genetic disease.

earlier and later cohorts respectively. These large sample sizes allow us to get more precise estimates.

Children's educational outcome in this paper is measured by college attendance at age 18. We combine records of the first-born children in the mother-based birth registry data set with College Entrance Examination records, using children's IDs. Unlike Black et al. (2005) and Angrist et al. (2006), who used years of schooling to capture children's educational outcomes, we use a dummy for college entrance as children's long-term educational outcomes. Long-term influences of family background on children's educational outcomes can also reflect constraints of family resources, rather than short-term effects (e.g., Camerom and Heckman 1998; Heckman and Lonchner 1998; Cameron and Taber 2004).

Our matched data set is able to overcome potential problems emanating from using conventional cross-sectional data and panel surveys. Children's outcomes and characteristics cannot be fully observed in census data, particularly after they leave home. Thus, using census data to investigate the effects of sibling gender and family size on educational outcome could be problematic, due not only to biased sex-composition but also instance of incomplete family size.[10] On the other hand, panel survey requires a very long time to collect data on completed family size and children's long-term educational outcomes; it usually has small sample sizes, which may be vulnerable to larger measurement errors.

2.2 Summary Statistics

Table 1 presents other summary statistics of our subjects. We can observe that the birth year of second-born children is 1980 in the earlier cohort, and 1984 in the later cohort, both with small standard deviations. For sex-composition of the first two

[10]When using census data, the choice age of subjects is difficult. Choosing older children has the risk of omitting children who had left home to work or to get married; choosing younger children could leave out those who were not born yet and limit the findings to only short-term outcome. No matter how the sample's age is selected, it cannot capture the completed family size.

children, the possibility of having two boys is 12 percent, higher than of having two girls. On average, families in earlier cohorts have 2.814 children, larger than those in later cohorts, which have 2.649 children. The rate of twinning at second birth, which is used to instrument family size, is 0.007 for both cohorts. To further ensure our twinning instrument's validity to measure variation of family size, we go through birth weight controls. The average birth weight of first-born children is slightly lower than second-born children. Note that if second birth is twins, we measure the mean of their birth weights.

Interestingly, girls outperform boys in the college enrollment rate in our data. According to Table 1, we can see that about 11.9 to 16 percent of boys passed the Joint College Entrance Examination and enrolled in college at age 18, whereas a higher percentage (by around 8 percent to 14 percent) of girls go to college. Figure 1A also shows the same pattern. It shows that families in the earlier cohorts prefer larger families than those in the later cohorts. Although we can see in Table 1 and Figure 1B that a child's probability of entering a college increases as family size decreases, it could simply mean that cohort effects, such as family size, have shrunken over time, as the overall educational attainment has increased.

Parental ages are similar in both cohorts; mothers in the latter cohort have their first birth at an age of 23.6, slightly older than the 23.1 years of those in the earlier cohort. Parental education is a very important characteristic that affects children's outcomes. Our data indicates that the percentage of educated mothers, who have an academic high school degree or above is less than that of educated fathers. For example, the number of college-educated fathers is double that of college-educated mothers. Generally speaking, parents in latter cohort have higher education levels than those in the earlier cohort.

Statistics of first-born girls, by sibling gender and cohorts, are listed in Appendix Table 1. Columns (1)(2)(4)(5) show the means of variables for first-born girls, with second-born brother and sister. Column (3) and (5) show the differences in variables between second-born brother and second-born sister. As one can see, first-born girls with a younger sister are more likely to have been born in larger families and have lower probability of entering college, than those with a younger brother. It is worth noting that, in the latter cohort, it seems that the time lag between first and second birth was more in case of second-born boys, than in case of second girls (in row 4), which may be suggestive of possible sex-selection for second birth. Other variables are generally balanced between these two groups, for each cohort.

3 Estimations

3.1 Testing the randomization of Second-Born child's gender

Our empirical analysis begins with two empirical tests to examine the randomization of second-born child's gender. We find no evidence that the sex composition of the first two births can be a result of sex-selective abortion for the older cohort. However, we cannot reject that sex-selective abortion exists in younger cohort, which leads the sibling gender effect for younger cohort to being overestimated.

We first test if the probability of having a son at the second birth can be related to any of the observed family characteristics in our data, such as parental age, education, or residential areas. We estimate a linear probability model, where the indicator for having a son at the second birth is the dependent variable, using the full set of covariates used in outcome regression. F-tests indicate that the hypotheses of jointly zero coefficients on mother's education, father's education and birth counties, cannot be rejected at even 10 percent significance level for both cohorts. This may imply that the gender of second-born children can be seen as random, conditional on observable covariates. We do not present these results in the paper.

The construction of the second test is motivated by "sex selection hypothesis" in Ebenstein (2007). The main point of this hypothesis is that if male birth is more likely to be selected by boy-preferring parents, the birth interval should be distorted by the time it takes to abort a female fetus. So we would expect that, on average, male births will be preceded by a longer interval than female birth, if the first birth is female.[11] Based upon this point, the regression we estimate is:

$$BI_i = X_i'\phi + \pi_1 G1_i + \pi_2 G1B2_i + \nu_i, \tag{1}$$

where BI_i is the interval between first and second births measured in years for a family i; $G1$ is the dummy variable for first child's gender, equal to 1 if the first-born was girl; $G1B2$ refers to interaction term between the indicator of first-born being a girl and the indicator of the second-born being a boy. The observable family covariates (the vector of X_i) include parental education indicators (one for each education category),[12] parents' age cohorts indicators (one for each year of birth), indicators of year in which the child is eligible for college, and birth county indicators. The parameter π_1 captures the effect of having first-born girl(with a younger sister) on the birth interval, while the estimate of interaction term, π_2, measures the second boy/second girl ($G2/B2$) difference in birth spacing. If π_2 appears to be significantly positive, the sex selection hypothesis may not be rejected.

Results of testing equation (1) are reported in Table 2. Estimates of π_1 and $\pi_1 + \pi_2$ are negative, which implies that first-born girls are more likely to have a younger sibling within a shorter interval than first-born boys. This is evidence of son preference that shows that parents having daughters are keen to have another children till having a son.

The key estimate to determine whether the sex selection hypothesis holds is the coefficient of interaction term, π_2. As Table 2 shows, estimates of π_2 are significantly positive only for younger cohort. It suggests that estimates of π_2 are not manifested in older cohort, implying that the sex selection problem did not exist in older cohort. In contrast, in younger cohort, the time lag between the first-born girls and their second-

[11] Ebenstein (2007) used data of several Asian countries to show statistical evidence for this hypothesis, instead of regression analysis. He found that the distortion interval happened between second-born and third-born children in Taiwan, but it did not happen between first-born and second-born children

[12] We use 5 categories to capture parental education level: completed college degree or above; completed professional training college; completed high school degree; completed vocational high school degree; completed junior high school degree. (The excluded category is primary or below)

born boys was longer by 0.017 years, compared to second-born girls, indicating that gender of second-born children in younger cohort is more likely to be selected by parents.[13] As a result, estimates of sibling gender effect is upper-bound (overstated) for younger cohort.

3.2 OLS Estimates of Effects of Sibling gender and Family Size

3.2.1 Without controlling for family size

We first estimate a short regression of equation (2) without controlling for family size:

$$Y_i = X_i'\alpha + \beta B2_i + \omega_i, \tag{2}$$

where the outcome variable (Y_i) is binary, which takes the value of 1 if the subject enters a college, 0 otherwise. The vectors of covarites are the same as in regression (1). We divide our subjects into girls and boys and run the regressions separately.

Estimates of sibling gender and standard errors are presented in columns (1) and (5) of Table 3. We can observe that, for first-born girls, the coefficient of sibling gender is 0.0027 (s.e.=0.0018) in earlier cohort, and 0.0015 (s.e=0.0013) in latter cohort. They are both insignificant but positive. It seems that having a second-born brother does not reduce first-born girls' possibility of entering a college. For first-born

[13]The best way to deal with endogeneity in latter cohorts is to find possible ways to capture exogenous variations in the gender of second-born children. The point of time when legalization of abortion took place is also a plausible way, though, it might just reflect cohorts effect. Apart from the time variation, we need to find other variation in demand for boys related to parental heterogeneity. We try to correlate the change of abortion law and the heterogeneity of parental characteristics, but we can not find any variation in demand for boys at second birth that is correlated with parental characteristics. Actually, in our data, the variation in demand for boys, associated with parental heterogeneity, takes place only at higher parity births. Figure 2 shows an obvious difference in sex ratio of third-born children between highly educated mothers and less educated mothers, after 1985/86, but this pattern is not observed for second births.

boys also, the effect of sibling gender is very small, almost zero. These results holds even if we exclude the controls of parental education.

3.2.2 Controlling for Family Size

Next we estimate the regression with control of family size. In OLS estimation, family size is treated exogenously. The estimating regression is:

$$Y_i = X_i'\alpha + \beta B2_i + \gamma N_i + \epsilon_i \tag{3}$$

Columns (2) and (6) of Table 3 report the results of coefficients of sibling gender and family size in regression (3).

For first-born girls in earlier cohorts, one can see that after adding the family size control, the estimates of sibling gender effect change sign from positive to negative, i.e. from 0.0027 (s.e.=0.0018) to -0.0017 (s.e.= 0.0019), but are still not manifested. The change between without and with controlling for family size, in the latter cohort, is apparent: from zero to significantly negative (-0.0036 (s.e.=0.0014)). Adding family size control seems to change the behavior of sibling gender. However, it actually reflects what we mentioned above: a confounding control problem of family size may bring in a bias term in sibling gender effect on educational outcomes. Hence it is difficult to interpret that having a younger brother reduces first-born girls' education. For first-born boys, having a second-born brother does not have any impact on their possibility of entering a college, in both regressions, with and without controlling for family size.

The estimate of family size here is small but significantly negative. For first-born girls, approximately speaking, having one more sibling will decrease a first-born girl's possibility of entering a college by around 1 percentage point. We find similar results for first-born boys, where having one more sibling decreases firs-born boys's possibility of enrolling in a college by around 1.2 to 1.6 percent. In summary, with controlling for family size, without addressing its potential endogeneity, the trade-off between children's education and number of siblings exists, as stated in most of the earlier findings in literature of Q-Q.

3.3 2SLS Estimates of Effects of Sibling Gender and Family Size

3.3.1 Family Size and Twin Births (First-Stage)

To address the potential endogeneity of family size, we propose to use the occurrence of twins at second birth to measure exogenous variations in family size. Twin births can increase the number of children beyond the parents' desired size.[14] The first-stage regression is:

$$N_i = X_i'\delta + \rho B2_i + \theta Z_i + \eta_i \tag{4}$$

The estimates and standard errors in first-stage regression (4) are reported in columns (1) and (3) of Table 4. The occurrence of twins at second birth leads families to have 0.62 more children in case of families with first-born girls, in older cohort; this estimate rises to 0.66 in younger cohort. It demonstrates that parents' preference for a smaller family size in younger cohort is higher than in older cohort. Therefore, twins have a larger effect on family size in younger cohort because they are more likely to increase the number of children beyond what parent's desire. This phenomenon is more apparent for families with first-born boys. We can see that parents having first-born sons are less likely to opt for further fertility, unless they have an unexpected twin birth. Hence, the twin effect on family size is larger, by around 0.71 to 0.73, for families with first-born boys than those with first-born girls.

Additionally, we can observe that having a boy at second birth will make families with first-born girls cut down the number of children by around 0.42 to 0.46. It implies that parents with two daughters are likely to have 0.42 to 0.46 more children than those with a first-born daughter and a second-born son. This is clear evidence of the fact that Taiwanese parents have a remarkably high degree of son preference. Parents are more likely to stop having an extra babies if they have already had a son.

[14] Some previous research of QQ also suggested that sibling sex composition at the first two births may also cause another possible exogenous variation in family size (Lee 2003, Angrist et al. 2005, Conley and Glauber 2006).

Families with first-born sons also exhibit a similar phenomenon. Furthermore, the first-stage also suggests that gender composition of the first-two children does have a strong impact on parents' fertility decision, such that the family size is an outcome variable of sibling gender composition.

The twinning instrument here satisfies the condition that twinning has a strong impact at the first-stage. Additionally, to be a valid instrument for family size, twins at second birth has to be uncorrelated with the error term in equation (3) . If the existence of twin births are not random, or related to family background characteristics, it may lead to inconsistent estimates. Although it is not possible to test the relationship between twins birth and unobservable family backgrounds, we examine whether the probability of twins corresponds to observable parental education. The null hypothesis is that the coefficients of mother's and father's education are jointly zero in regressing parental education levels, on the possibility of having twins at second birth. The F-tests suggest that we can not reject the null hypothesis at 5 percent significance level. This result explains that the possibility of having twins at second birth is not correlated with parental education.

Furthermore, the other condition of a valid instrument is that twinning cannot affect children's educational outcome, other than through affecting family size. Some research has documented that children with LBW (Low Birth weight) have lower health and human capital outcome (Behrman and Rosenzweig (2004), Almond, Chay and Lee 2005 and Lin, Liu and Chou 2007); twins are more likely to have low birth weights. This indicates that twinning can not only affect children's outcome through family size, but also through (low) birth weight. Rosenzweig and Zhang (2006) suggested that while using twinning as an instrument, one has to further add controls of birth weights. To improve the validity of the twinning instrument, we further control for the subject's and the second-born birth's weight at the first-stage (equation (4)).

Results of the first-stage with birth weight controls are shown in columns (2) and (4) of Table 4. Note that if second birth is twins, we use their mean birth weight. Generally speaking, after controlling subject's and second-birth's birth weights, the twinning effect on family size declines by a small percentage, with coefficients of around 0.59 to 0.64 for families with first-born girls, and around 0.7 for those with first-born boys.

Sibling gender effect on family size has no obvious changes. In addition, children's birth weight is negatively associated with family size. It may imply that parents who care more about their children's health (indicated by birth weight) have fewer children.

3.3.2 Results of Second-Stage

The general second-stage regression for subject i is the same as outcome regression (3). The second-stage estimates of effects of sibling gender and family size on children's possibility of entering a college are presented in columns (3), (4) and (7), (8) of Table 3, which are corresponding to first stage in Table 4. The difference between (3)(7) and (4)(8) is controls of subject's and sibling's birth weights.

Estimates of sibling gender effect remain close to zero in older cohorts. It is worth noting that after addressing the potential endogeneity problem of family size, the sibling gender effect for younger cohort is changing from significantly negative OLS estimate (-0.0036) to zero, comparing columns (6) and (7). Given our enormous sample size, these estimates can reasonably be assumed to be quite precise. This supports the discussion in the section on framework, that addressing the endogeneity of family size also eliminate the indirect effect of sibling gender intermediated via family size. In addition, adding controls of birth weights does not alter the results, while subjects' birth weight is found to be positively correlated to their likelihood of entering a college, consistent with the literature.

As a result, we find that the effect of second-born sibling's gender is zero. Even this estimate may be upper-bound for younger cohort, the zero estimate implying that the causal effect is zero. Alternately, this result is also supported by the same estimated result, using twin births in Chen, Chen and Liu 2008, which exploited the first-born twins sample, who exhibit random nature sex-ratio, to investigate the analogous twin-sibling gender effect. In short, surprisingly, given the overwhelming son preference, we can not find evidence to show that having a brother will crowd out a first-born sister's possibility of entering a college.

For family size effect at first-born girls, the 2SLS estimates of the increase in family size, induced by twins at second birth, is -0.038 to -0.042 in older cohort, when comparing results with and without controlling children's birth weight. These estimates are statistically significant at 5-percent level. It shows that the first-born girls in older cohort received less education if they are from larger families. This phenomenon does not exist for first-born boys in older cohort. In younger cohort, family size effect reduces to zero after instrumenting for family size. The estimates are -0.013 (s.e.= 0.012) for first-born girls, and -0.006 (s.e.=0.011) for first-born boys. These numbers indicate that there is no significant effect of family size on children's educational attainment in younger cohort. In sum, taking it together with the very little family size effect estimated from the OLS method, our 2SLS results show little evidence of a trade-off relationship between number of children number and their educational outcomes, except for first-born girls in the older cohorts. These results are similar to recent literature on quality-quantity trade-off.

3.4 Other 2SLS Results

3.4.1 By Every Single Year

In above studies, we found that there is no gender effect, and only adverse family size effects in earlier cohorts. In this subsection, we examine the effects of sibling gender and family size for each year of birth, in order to see whether these two effects vary by cohorts of each year. The estimating regression here has controls of subjects' and second-born children's birth weights in first- and second -stages. The coefficients and standard errors are reported in columns (1) to (5) of Table 5.

Interestingly, we find that having a second-born brother and coming from a larger family matter for the college-entering possibility of first-born girls; however, this applies to only those born in 1978. For first-born girls born in 1978, the coefficient of sibling gender is -0.0245 (s.e.=0.0122), which is statistically significant at 5-percent level. Recalling that estimates of sibling gender effect in earlier cohorts are unbiased, this result shows that first-born girls, who were born in 1978, had a 2.4 percent lower possibility of entering college if they had a second-born brother, rather than a second-born sister. These girls also suffered from their larger family size. Having an additional younger brother decreases their possibility of entering college by 6.3 percent. Except for first-born girls born in 1978, in other cohorts, there are no effects of sibling gender and family size, regardless of birth year or gender of the first-born children. We omit the results of 1982 and 1983 in the table, because they are quite similar to results of 1984.

Robust Check – Columns (6) to (9) of Table 5 report the results for first-born children whose subsequent siblings were born before 1985, for robust checking. This check is to see whether subjects with subsequent sibling born in the pre-legalized-abortion period have the same estimation results as those presented earlier. Under this new sample construction, the sex ratio of second-born children is similar to our original construction, but the test of "sex selection hypothesis" is rejected, implying that sex selection issue is less likely to have existed. Owing to the time gap between first- and second-born children getting smaller, we investigate the results only for first-born children born between 1978 and 1981. Comparing the results presented here (columns (6) to (9)) to earlier(columns (1)-(4)), one can see that they are roughly similar. Effects of sibling rivalry and family size are reflected only in case of first-born girls born in 1978, while the estimate of sibling rivalry effect is significant at 10-percent level but not at 5-percent level.

3.4.2 By Parental Education

In addition, effects of sibling gender and family size may vary with parents' economic status. For example, parents with higher income may allocate their resources to children more evenly, since they have fewer budget constraints. Owing to the absence of parental income variables in our data, we use parental education as the proxy for family income. We define higher educated parents as those who had high school degree or above, including vocational high school. The coefficients and standard errors are listed in Table 6. The table shows that, in general, although the coefficients of sibling gender effect and family size are stronger for those who had relatively more educated parents, estimates of effects of these two factors are not significant on children's education.

4 Interpretation

We find the significant effects of sibling rivalry (adverse sibling gender) and family size on children's education, but only for first-born girls born in the earliest year, 1978, in our data. For the recent cohorts, we can not find the evidence on negative effects of having a brother (relative to a sister) and having larger family size on women's education. This is particularly surprising given the long-lasting and extraordinarily strong preference for sons in Taiwan.

There are two possible explanations for the absence of sibling gender effect on girls' probability of entering a college in recent years. First, the decreasing gender gap in "return to education" may have led parents to allocate family resources more equally between boys and girls. In fact, Taiwanese women have been two to three percent more likely to enroll in colleges than men, on average, in the last one decade or so. The increase in women's education level corresponds with the increase in women's wages. Using Taiwan Labor Force sample surveys across 1990's to 2000's, we find that the gender gap in wage is generally shrinking over time. The ratio of women's average wage to men's increased from 0.65 in 1980 to 0.85 in 2000.[15]

The second reason for the zero gender effect is that having a brother may benefit girls externally. For example, parents may devote more time to nurture all their children if they have a son. Using Taiwan 1990 and 2000 census , we select parents

[15]Taiwan Labor Force annual survey has around 5,000 to 6,000 samples, aged between 15 and 65. It provides monthly wage information. To estimate the gender gap in wage, we select a sample of those aged between 25-34, to investigate average wage of men and women.

who had only one child of less than two years age in the census day,[16] and look at how the child gender affects mothers' and fathers' labor supply. We find that mothers having an infant son are 0.5 to 0.6 percent more likely to exit the labor market and stay at home to nurse their child, than those having a daughter. Girls may get more care from mothers if the subsequently born sibling is male, not a female. This suggests that the negative impact of having a brother on daughters' education may be offset by the benefit from the increase in parenting time of the mother, due to the concern about the son. Pabilonia and Ward-Batts (2007) also reported similar findings on the gender effect of children on parental labor supply . Additionally, some researchers also discovered that the existence of a male child will increase parent's marriage stability or father's income (Lundberg and Rose 2003; Lundberg 2005; Dahl and Moretti 2008). Having a young male sibling may generate spill-over benefits for females.

Our study find little evidence on family size effect in recent years, which is consistent with Angrist et al. (2006) and Black et al. (2005), who also used an enormously large sample and looked at children's long-term outcomes. One of the possible explanations for this finding is that allocation of parental resources to each child may not necessarily get diluted with the increase in number of children. For example, fathers may work longer to earn more money, and spend less on themselves in response to the increasing family size. However, this view is difficult to be proved because of the lack of detailed consumption expenditure data.

Alternatively, there is another possible way to relax the parental budget constraint in Taiwan. Most grand parents support their grandchildren, in terms of nursing support or financial support. In particular, with the rapid economic growth in 1970s and 1980s, grand parents may have become richer than before. Support from grand parents is a powerful resource that relaxes parental budget constraints in raising

[16]Our birth registry records do not contain parents' status in terms of labor supply. Hence, we use census data to examine child gender effect on parents' labor supply. The reason that we choose parents with young only-child is to avoid the family size effect on parent's labor supply.

children. However, it is difficult to show the supporting evidence of this view as well, since we do not have suitable data covering transfers and assistance from grandparents to their grand children.

5 Conclusions

This paper constructs a unique data base by combining 1978-1999 Taiwan Birth Registry records and 1996-2003 College Entrance Examination data to study the causal effect of sibling gender and number, simultaneously, on children's educational attainments, by overcoming the major difficulties previous researches faced. Our empirical methodology removes the problem caused by confounding control of family size on sibling gender effect, and address sources of endogeneity of family size and child gender.

We restrict our analysis to first-born children, and look at the influence of second-born sibling's gender and family size, on subjects' possibility of entering a college, using the occurrence of twins' birth as an exogenous variation instrument for family size. Having a subsequent brother and a larger number of siblings have significant effects on children's education only for girls born in 1978. Surprisingly, given the strong son preference, we cannot find the significant evidence on effects of sibling gender and family size on children's possibility of entering colleges for post-1978 cohorts.

A number of plausible reasons are presented to explain the persistence of the absence of effects of sibling gender and family size. Due to the growing return to education of women, parents do not necessarily exercise gender discrimination in allocation of family resources, especially in the last two decades. Moreover, having brothers may provide external benefit to girls , which tends to counter the negative impact of brothers on their educational outcomes. In addition, the reason of zero family size effect for first-born children may be that family resources do not get diluted with the increases of number of children, because of change in consumption pattern, or of supports from grand parents.

References

[1]

[2] Ananat, Elizabeth O. and Michaels, Guy. 2004"The Effect of Marital Breakup on the income and Poverty of Children?" MIT Working paper

[3] Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), "Identication of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91:444-472.

[4] Angrist, Joshua D. and W. Evans. 1998."Children and Their Parents' Labor Supply: Evidence from Exogenous Variations in Family Size." *American Economic Review* , 88(3): 450-477.

[5] Angrist, Joshua D., Victor Lavy and Analia Schlosser. 2006. "New Evidence on the Causal Link between the Quantity and Quality of Children." MIT Department of Economics Working Paper

[6] Angrist Joshua D., and Jorn-Steffen Pischke. 2008. "Mostly Harmless Econometrics: An Empiricist's Companion." Princeton University Press.

[7] Ansley J. Coale. 1991."Excess Female Mortality and the Balance of Sexes in the Population: An Estimate of the Number of 'Missing Female'. " *Population and Development Review*, 17(3): 517-523.

[8] Backer, Gary S., 1960. "An Economic Analysis of Fertility." *Demographica and Economic Change in Developed Countries*, Princeton, NJ: Princeton Unversity Press

[9] Becker, Gary and H. G. Lewis, 1973. "On the interaction between the Quantity and Quality of Children". *Journal of Political Economy*, 81(2), part (2): s279-288.

[10] Behrman Jere R., Robert Pollak and Paul Taubman. 1982. "Parental Preferences and Provision for Progeny" *Journal of Political Economy* ,90(1):52-73.

[11] Ben-Porath , Yoram and Finis Welch.1976."Do Sex Preferences Really Matter?" *The Quarterly Journal of Economics*, 90(2):285-307

[12] Behrman Jere R., and Paul Taubman. 1986. "Birth Order, Schooling, and Earnings". *Journal of Labor Economics* 4:S121-S145.

[13] 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): 669-700.

[14] Blake, Judith. 1981 "Family Size and the Quality of children". *Demography* , 18:421-42.

[15] Butcher, F. Kristin and Anne Case. 1994."The Effect of Sibling Sex Composition on Women's Education and Earnings." *Quarterly Journal of Economics*, 109(3):531-563.

[16] Cameron, Stephen and James J. Heckman. 1998. "Life-Cycle Schooling and Educational Selectivity: Models and Choices." *Journal of Political Economy*, 106(2):206-333.

[17] Cameron, Stephen and Christopher Taber. 2004. "Estimation of Educational Borrowing Constraint Using Returns to Schooling." *Journal of Political Economy*, 112(1): 132-183.

[18] Chen, H. Stacey, Yenchien Chen, and Jin-tan Liu. 2008. "The Impact of Sibling Sex Composition on Women's educational Achievements: A Unique Natural Experiment by Twins Gender Shocks," mimeo, SUNY-Albany, Department of Economics.

[19] Caceres, Julio. 2004. "Impact of Family Size on Investment in Child Quality: Multiple Births as Natural Experiment." mimeo, Department of Economics, University of Maryland.

[20] Case, Anne and Angus Deaton. 2002. "Consumption, Health, Gender and Poverty". World Bank. Research Program in Development Studies.

[21] Conley, Dalton. 2004. " What is the 'True' Effect of Sibship Size and Birth Order on Education? Instrumental Variable Estimates from Exogenous Variation in Fertility." New York University, Working Paper.

[22] 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*,41(4): 722-37

[23] Dahl, Gordon and Enrico Moretti. 2008 Forthcoming. “The Demand for Sons: Evidence from Divorce, Fertility and Shotgun Marriage.” *Review of Economics and Statistics*.

[24] David, Daniel J. 1977. “Birth Order and Intellectual Development: The Confluence Model in the Light of Cross-cultural Evidence.” *Science*, 196: 1470-1472.

[25] Ebenstein, Avraham Y. 2007. “Fertility Choices and Sex Selection in Asia: Analysis and Policy.” Mimeo, Harvard University.

[26] Featherman, David L. and Robert M. Hauster .1978. “Opportunity and Change.” New York: Academic Press James

[27] Garg, Ashish and Jonathan Morduch. 1998. “Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana.” Journal of Population Economics. 11: 471-493.

[28] Hanushek, Eric.1992. “The Trade-off between Child Quantity and Quality.” *The Journal of Political Economy*, 100(1):87-117

[29] Heckman, James J., Lance Lochner and Christopher Taber. 1998. “ General-Equilibrium Treatment Effects: A Study of Tuition Policy. ” *The American Economic Review*, 88(2): 381-386.

[30] Hudson, Valerie M. and Andrea M. den Boer. 2005. “Bare Branches”.MIT Press.

[31] Imbens, G. W., and Angrist, J. D. (1994), “Identication and Estimation of Local Average Treatment Effects,” *Econometrica*, 62:467-476.

[32] James, William H. 1985.“The Sex Ratio of Oriental Births.” *Annals of Human Biology*, 12(5):485-487.

[33] Johansson, S and O. Nygren. 1991. “The Missing Girls of China– A New Demographic Account .” *Population and Development Review*, 17:35-51.

[34] Kaestner, Robert. 1997. “Are Brothers Really Better? Sibling Sex Composition and Educational Achievement Revisited.” *Journal of Human Resources*, 32(2):250-284.

[35] Koolwal, Gayatri B. 2007. “Son Preference and Child Labor in Nepal: The Household Impact of Sending Girls to Work.” *World Development*, 35(5): 881-903

[36] Lee,Jungmin, 2003. “Children and Household Education Decisions: An Asia Instrument.” Working paper

[37] Lundberg, S. J. and E. Rose. 2002. “The Effect of Sons and Daughters on Men’s Labor Supply and Wages.” *Review of Economics and Statistics*, 84: 251-268.

[38] Lundberg, S. J. and E. Rose. 2003, “Child Gender and the Transition to Marriage.” *Demography* .40(2): 333-349.

[39] Lundberg, S. J. 2005. “Sons, daughters, and parental behavior.” *Oxford Review of Economic Policy* 21(3): 340-356.

[40] Lundberg, S.J., S. McLanahan, and E. Rose. 2007. “Child Gender and Father Involvement in Fragile Families.” *Demography* , 44(1): 79-92.

[41] Morduch, Jonathan. 2000. “Sibling Rivalry in Africa.” *The American Economic Review*, 90(2):405-409.

[42] Qian, Nancy. 2008. “Quantity-Quality: The Positive Effect of Family Size on School Enrollment in China”. Department of Economics, Brown University.

[43] Lin, Ming-Jen, Jin-Tan Liu and Nancy Qian. 2007. “Missing Women More and More: The Effect of Access to Abortion on Sex Ratios at Birth and Excess Female Mortality.” Mimeo. National Taiwan University and Brown University.

[44] Pabilonia, S.W. and J. Ward-Batts. 2007. “The Effect of Child Gender on Parents’ Labor Supply: An Examination of Natives, Immigrants, and their Children.” *American Economic Association Papers and Proceedings*, 97(2), 402-406.

[45] Parish, William and Robert Willis. 1993. “Daughters, Education and Family Budgets: Taiwan Experiences.” *Journal of Human Resources*,28(4): .863-898.

[46] Powell, Brian and Steelman, Lara C. 1989. “The Liability of Brothers: Paying for College and Sex Composition of the Family.” *Sociology of Education*, 62: 134-47.

[47] —. 1993. “The Educational Benefit of being Spaced out: Sibship Density and educational Progress.” *American Sociological Review*, 58:188-192.

[48] Rose, Elaina. 2000. “Gender Bias, Credit Constraints and Time Allocation in Rural India.” *Economic Journal* 110(465): 738-58.

[49] Rosenzweig, Mark R. and Kenneth I. Wolpin. 1980 "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica*, 48(1): 227-240.

[50] Rosenzweig Mark R. and Junsen Zhang. 2006. "Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's 'One-Child' Policy." Yale Working paper.

[51] Sen, Amartya. 1992. "Missing Women: Social Inequality Outweighs Women's Survival Advantage in Asia and North Africa." *British Medical Journal*, 304(6827): 577-579.

[52] Thomas, Duncan. 1990. "Intra-household Resource Allocation: an Inferential Approach", *Journal of Human Resources*, 25(4): 635-664

Appendix 1: Previous Approaches

We present estimates in Appendix Table 2 using models proposed in previous studies, Parish and Willis (1993), Butcher and Case (1994) and Grag and Morduch (1998), for the purpose of comparison. In these models, we use all female children born between 1978 and 1984 in our dataset, irrespective of birth order. According to Appendix Table 3, we cannot find a consistent conclusion among these models. Model I shows that an increase in number of siblings, regardless of age or gender, decreases girls' possibility of entering a college. Model II shows that having a brother has no effect on girls' education, while the existence of a sister helps girls' education. In contrast with model II, model III shows that having a brother increases girls' education. The coefficients of family size are significantly negative in model II and model III.

One of the causes leading to the confounding estimates is "birth order". Because children with different birth orders have different endowments, pooling children with different birth orders together into indicators of sibling sex composition may make the results hard to interpret. For instance, some of existing psychology literature has suggested that children with lower birth order (earlier-born) can learn more from teaching and caring their younger siblings, and, therefore, can perform well in education (Zajonc 1976, Behrman and Taubman 1986). Hence, the effect of younger siblings on children's education may be different from the effect of older siblings. If we generate a variable of sibling sex composition as "the indicator of any brother" which clubs older and younger brother together, the estimate of this variable may get confounded by birth order effect.

To overcome the problem from pooling children with different birth orders, we modify the above models by further separating all female samples into 3 categories, by birth order, and looking at the effects of older and younger siblings separately. The results are listed in Appendix Table 3. When we analyze first-born girls, we look at the effect of presence of a younger brother or a younger sister; while analyzing second-born or after-third-born children, we go through the indicators of an older brother (sister) and younger brother (sister). By doing this, we can find a common conclusion, among second-born and later-born girls: children with higher birth orders receive less education if they have an older brother. For first-born girls, it shows that having at least one younger brother increases their possibility of entering a college by 0.3 percent, relative to having only younger sisters, while having sisters does not affect their education attainment.

However, the modified estimates still have a disadvantage in the "rough" indicators of having an older or younger brother (sister). Consider a case where a girl has three younger brothers. The older of the brothers may be competing more for parental resources, than the youngest brother. On the other hand, the girl may have learning benefit from teaching the youngest brother. Hence, if the variable of having any younger brother sums up all younger brothers together, the estimate would be confounding. Besides, sex selective issues may make the estimates of sibling sex composition more complex, if this variable includes children with different birth orders. For example, parents may be more likely to choose the gender of a later-born child because age and family resources become constraints.

The other important reason for confounding estimates of gender sibling effect in previous studies is endogenous family size. Bad-control family size contaminates the sibling gender effect on children's education, which is because the gender composition of children remarkably influences parents' desired family size, and the endogeneity of family size make the situation even worse. Therefore, if this disturbance of family size is not addressed, the effect of sibling gender on children's education will be confounded with the effect of sibling gender on family size. Although some of the previous studies sensed this problem, they were unable to tackle it, owing to the absence of valid instruments in their data. As a result, the complexity of interactions among birth orders, sibling gender composition and family size, confound previous estimates of effects of sibling gender composition, making them hard to interpret.

Appendix 2: The Proof

Prior empirical literature regarding the effect of sibling sex composition on child outcomes[17] is built upon 3 assumptions: (i) sibling sex composition is randomly assigned, (ii) family size is exogenous, and (iii) family size is predetermined, not affected by changes in sibling sex composition. We first show that violation of assumption (iii) can lead to biased results, even if family size and sibling sex composition are both randomized.

We illustrate our empirical strategies using the notation of counterfactuals.[18] Define Y_{1i} as the i-th firstborn's *potential outcome* if the subsequent sibling is male, and Y_{0i} if female. The observed educational outcome can be written as $Y_i = B2_iY_{1i} + (1 - B2_i)Y_{0i}$. Let vector X_i denote i's family backgrounds and demographics (or briefly "covariates"). The first parameter of interest is the *overall effect* of having a subsequent brother, relative to a sister, on the firstborn's outcome, conditional on covariates

$$\Delta(x) \equiv E[Y_{1i} - Y_{0i}|X_i = x].$$

The following identification condition has been adopted by the pervious studies on sibling rivalry and spillover (see footnote 17), and it will be used in our initial analysis:

Assumption 1 Random Treatment: *Conditional on exogenous covariates X_i, potential outcomes (Y_{1i}, Y_{0i}) are jointly independent of $B2_i$.*

Under this condition, $\Delta(x)$ can be identified by a simple statistics; that is, the conditional mean difference in the observed outcomes by the subsequent sibling's sex,

$$E[Y_i|B2_i = 1, X_i = x] - E[Y_i|B2_i = 0, X_i = x] = \Delta(x),$$

whereby Assumption 1 the equality holds.

[17]See, e.g., Parish and Willis (1993), Butcher and Case (1994), and Garg and Morduch (1998).

[18]See, e.g., Rubin (1974, 1977), Pearl (2005), Peterson, Sinisi and van der Lann (2006).

Alternatively, as in Butcher and Case (1994) and Garg and Morduch (1998), we are also interested in the *controlled direct effect* of having a brother, as opposed to a sister, with family size fixed, $N_i = n$. It can be written as

$$r(x, n) \equiv E[Y_{1i} - Y_{0i} | X_i = x; N_i = n].$$

Estimation and interpretation of $r(x, n)$, however, present conceptual and practical difficulties, because family size can change with sibling sex composition; it is impossible to hold family size constant in such a way that the effect of a change in sibling sex composition could be isolated. We next discuss this issue, and we show that Instrumental-Variables (IV) methods provide one way to address it.

Although often taken as an exogenous and predetermined covariate in the previous studies on sibling rivalry, family size (N_i) is another outcome variable, directly affected by treatment status $B2_i$. In this subsection, we maintain the assumption that family size is exogenous, but we relax the condition that family size is predetermined. In this subsection, we first consider cases where family size is determined by a function of exogenous treatment and covariates, $B2_i$ and X_i.

We show below that, even if both child gender and family size are exogenous, the indirect sibling gender effect via changing family size cannot be separated from the controlled rivalry effect, without additional assumptions. We further propose a new way of defining the controlled direct effect of sibling gender shock, using notations of counterfactuals. We will show that as long as exogenous family size depends on exogenous sibling sex, identification of the direct and indirect effects (and the main effect of family size) requires the existence of an instrument for family size.

Let N_{1i} (or N_{0i}) denote the i-th firstborn's *potential family size,* if i's subsequent sibling is male (or female). We note that N_{1i} and N_{0i} cannot be both observed for the same child i. If parents favor boys over girls, other things being equal, we may have $N_{1i} \leq N_{0i}$ because parents are more likely to have another baby after having a girl. In cases where family size can vary with treatment status, the controlled direct effect $r(x, n)$ is not well-defined (so it cannot be identified by any statistics).

Alternatively, we consider another measure for the controlled net rivalry effect, conditional on *potential family size*. Precisely defined,

$$E[Y_{1i} - Y_{0i}|X_i = x, N_{1i} = n] \equiv r_1(x, n) \tag{5}$$

$$E[Y_{1i} - Y_{0i}|X_i = x, N_{0i} = n] \equiv r_0(x, n). \tag{6}$$

Parameter $r_1(x, n)$ (or $r_0(x, n)$) measures the direct sibling gender effect on firstborn's outcome, under the scenario where family size would be n if the subsequent sibling is male (or female). To motivate a consistent estimator for the net rivalry effects, we consider a simple regression model, similar in spirit to Butcher and Case (1994) and Garg and Morduch (1998):

$$\begin{aligned} & E[Y_i|B2_i = 1, X_i = x, N_i = n] - E[Y_i|B2_i = 0, X_i = x, N_i = n] && (7) \\ = \; & E[Y_{1i}|B2_i = 1, X_i = x, N_{1i} = n] - E[Y_{0i}|B2_i = 0, X_i = x, N_{0i} = n] \\ = \; & E[Y_{1i}|X_i = x, N_{1i} = n] - E[Y_{0i}|X_i = x, N_{0i} = n], && (8) \end{aligned}$$

where the validity of the second equality requires an independence condition stronger than Assumption (1); that is,

Assumption 2 STRONG RANDOM TREATMENT: *Conditional on exogenous covariates X_i, potential outcomes and potential family size ($Y_{1i}, Y_{0i}, N_{1i}, N_{0i}$) are jointly independent of $B2_i$.*

Even with the condition of Strong Random Treatment, the result of simple regressions in equation (7) does not necessarily provide a correct measure for the direct sibling gender effects, $r_0(x, n)$ and $r_1(x, n)$. To show this, we rearrange equation (8) as follows:

$$\begin{aligned} & E[Y_{1i} - Y_{0i}|N_{1i} = n, X_i = x] + \{E[Y_{0i}|N_{1i} = n, X_i = x] - E[Y_{0i}|N_{0i} = n, X_i = x]\} \\ & = r_1(x, n) + bias_0(x, n) = r_0(x, n) + bias_1(x, n), && (9) \end{aligned}$$

where the nuisance terms are:

$$bias_{B2}(x, n) \equiv E[Y_{B2,i}|X_i = x, N_{1i} = n] - E[Y_{B2,i}|X_i = x, N_{0i} = n], \tag{10}$$

for $B2 \in \{0, 1\}$. Where $r_{B2}(x, n)$ indicates direct sibling gender effect, while the nuisance term measures the indirect sibling gender effect intermediated through family size. In extreme cases where family size is independent of sibling gender conditional on covariates, the nuisance terms equal zero, and simple regression results identify both of the overall and direct sibling gender effects, $r_1(x, n) = r_0(x, n) = \Delta(x)$. In most cases, however, sibling gender affects potential family size. When covariates are fixed, the invariance of potential family size across sibling gender, $N_{1i} = n = N_{0i}$, suggests the existence of another factor Z_i that affects potential family size.

In order to formulate this, let Ψ be the support of Z_i. For each $Z_i \in \Psi$ and each $B2 \in \{0, 1\}$, potential family size induced by X_i and Z_i can be written as an additive form: $N_{1i} = \widehat{N}_1(X_i, Z_i) + u_i$ and $N_{0i} = \widehat{N}_0(X_i, Z_i) + u_i$, where u_i is an unobserved error term. For family size is assumed to be exogenous in this subsection, potential outcomes are independent of u_i and Z_i. Since Z_i affects potential family size but is independent of potential outcomes, Z_i is acting as an instrumental variable for family size, even though treatment status and family size are exogenous. Formally,**

Assumption 3 Existence of an Instrument: *Conditional on exogenous covariates X_i, the following 2 conditions hold for all $z \in \Psi$:*

(i) $(Y_{1i}, Y_{0i}, \hat{N}_{1i}(X_i, z_i), \hat{N}_{0i}(X_i, z_i))$ are jointly independent of Z_i,

(ii) the conditional distribution of potential family size is not a trivial function of z.

By Dawid's (1979) Lemma 4.2, the condition of Strong Random Treatment implies that potential outcomes (Y_{1i}, Y_{0i}) are also independent of $\hat{N}_{B2,i}$ for each $B2 \in \{0, 1\}$ conditional on X_i. Replacing $N_{B2,i}$ with $\hat{N}_{B2,i}$ in equation (10), we have $bias_{B2}(X_i) = 0$, because $\hat{N}_{B2,i}$ can be ignorable. Therefore, after instrumenting for family size, the direct sibling gender effect which is equalling to overall effect is able to obtain. We should note that, even family size is exogenous, the use of valid instrument for family size is necessary when estimating sibling gender effect with controlling family size.

Table 1: Descriptive statistics of firstborn singletons, with at least one sibling

	Older cohort Born during 1978-1979		**Younger cohort** Born during 1980-1984	
	Mean	Deviation	Mean	Deviation
Sex-ratio (boys/girls) of firstborns	1.047	(0.500)	1.041	(0.500)
Sex-ratio of the second-born siblings	1.067	(0.500)	1.066	(0.500)
Sex-ratio of the second-born siblings before 1985				
Birth years of the second-born siblings	1980	(1.902)	1984	(2.177)
Sex composition of the first and second births				
Two boys	0.262	(0.440)	0.262	(0.440)
Two girls	0.235	(0.424)	0.237	(0.425)
Family size	2.814	(0.865)	2.649	(0.778)
Twins at 2nd birth	0.007	(0.077)	0.007	(0.083)
Subject's birth weight (kg)	3.224	(0.452)	3.205	(0.442)
Second-born sibling's birth weight (kg)*	3.288	(0.466)	3.272	(0.465)
Outcome: college attainment				
Boys	0.119	(0.324)	0.160	(0.367)
Girls	0.139	(0.346)	0.184	(0.387)
Demographics				
Cohort information				
Mother's year of birth	1955	(3.278)	1958	(3.554)
Father's year of birth	1951	(4.748)	1955	(4.465)
Mother's age at first birth	23.147	(3.250)	23.575	(3.365)
Father's highest degree completed				
College degree or above	0.060	(0.159)	0.064	(0.169)
Professional training degree	0.065	(0.188)	0.076	(0.201)
High school (HS) degree	0.086	(0.221)	0.094	(0.246)
Vocational HS degree	0.151	(0.361)	0.187	(0.401)
Junior HS degree	0.172	(0.398)	0.256	(0.452)
Mother's highest degree completed				
College degree or above	0.026	(0.237)	0.030	(0.244)
Professional training degree	0.037	(0.246)	0.042	(0.265)
HS degree	0.052	(0.280)	0.065	(0.291)
Vocational HS degree	0.154	(0.358)	0.201	(0.390)
Junior HS degree	0.198	(0.378)	0.285	(0.436)
Sample size	265,284		627,872	

Source: Birth Registry records from 1978 to 1984, linked with the 1996-2003 College Entrance Tests records. Only singletons from families with at least 2 children are included. Standard deviations in (.) * If the second birth is twins, this statistics refers to the mean of twins' birth weight.

Table 2: Regression results of the sex of the firstborn on birth spacing from the second birth

Dependent variable: birth interval between 1st and 2nd births, in years

	Older cohort		**Younger cohort**	
	(1)	(2)	(3)	(4)
G1=1 if 1st birth was girl	-0.0478*	-0.0464*	-0.0548*	-0.0531*
	(0.0086)	(0.0085)	(0.0060)	(0.0059)
G1*B2	0.0147	0.0128	0.0175*	0.0163*
	(0.0098)	(0.0097)	(0.0068)	(0.0067)
Parental education controls	No	Yes	No	Yes
R-squared	0.0489	0.0676	0.0498	0.0678
Sample size	265,284	265,284	627,872	627,872

Note: Control variables include the firstborn's birth place and age on the college entrance test day; the full set of dummies for parental age and education levels; the mother's age at first birth; and eligible college years. The coefficients of all parental educations are significantly negative at 5% level.

Table 3: OLS estimates of the effects of sibling gender and family size on firstborns' college attendance, by gender and by cohort

Dependent variable: indicator for college attendance

	Older cohort				Younger cohort			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) Firstborn girls								
B2=1 if 2nd birth was boy	0.0027	-0.0017	-0.0151	-0.0164	0.0015	-0.0036*	-0.0036	-0.0034
	(0.0018)	(0.0019)	(0.0085)	(0.0091)	(0.0013)	(0.0014)	(0.0051)	(0.0053)
Family size		-0.0095*	-0.0386*	-0.0415*		-0.0122	-0.0123	-0.0102
		(0.0009)	(0.0181)	(0.0197)		(0.0008)	(0.0118)	(0.0125)
Subject's birth weight				0.0171				0.0206
				(0.0022)				(0.0016)
Sibling's birth weight				-0.0003				0.0029
				(0.0024)				(0.0016)
Instrumenting for family size	No	No	Yes	Yes	No	No	Yes	Yes
Parental education controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	129,601	129,601	129,601	129,601	307,729	307,729	307,729	307,729
B) Firstborn boys								
B2=1 if 2nd birth was boy	-0.0011	-0.0027	-0.0039	-0.0037	0.0007	0.0006	0.0002	0.0002
	(0.0017)	(0.0017)	(0.0026)	(0.0027)	(0.0012)	(0.0012)	(0.0015)	(0.0016)
Family size		-0.0122*	-0.0213	-0.0201		-0.0157*	-0.0059	-0.0034
		(0.0009)	(0.0156)	(0.0164)		(0.0008)	(0.0106)	(0.0111)
Subject's birth weight				0.0133				0.0205
				(0.0019)				(0.0014)
Sibling's birth weight				0.0006				0.0011
				(0.0020)				(0.0015)
Instrumenting for family size	No	No	Yes	Yes	No	No	Yes	Yes
Parental education controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	135,683	135,683	135,683	135,683	320,143	320,143	320,143	320,143

Note: All regressions are based on the matched firstborn sample, controlling for mother's age, father's age, mother's age at first birth, children's birth place and cohort effects. Parental education includes a full set of dummies for categorical education levels separately for father and mother. The coefficients of parental education are all insignificant at the 5 percent level.

Table 4: First-stage estimates, using twins at the 2nd birth as IV for family size
Dependent variable: family size

	Earlier cohort		Later cohort	
	(1)	(2)	(3)	(4)
A) Firstborn girls				
Twins at 2nd birth	0.6174*	0.5881*	0.6612*	0.637*
	(0.0262)	(0.0272)	(0.0126)	(0.0129)
B2=1 if 2nd birth was boy	-0.4619*	-0.4571*	-0.4166*	-0.4169*
	(0.0046)	(0.0047)	(0.0027)	(0.0028)
Subject's birth weight		-0.0164*		-0.0175*
		(0.0057)		(0.0035)
Sibling's birth weight		-0.0421*		-0.0319*
		(0.0056)		(0.0033)
Parental education controls	Yes	Yes	Yes	Yes
R-squared	0.2014	0.2018	0.1873	0.1891
Sample size	129,601	129,601	307,729	307,729
B) Firstborn boys				
Twins at 2nd birth	0.7097*	0.6953*	0.7251*	0.7049*
	(0.0224)	(0.0229)	(0.0109)	(0.0112)
B2=1 if 2nd birth was boy	-0.1282*	-0.1271*	-0.0873*	-0.0866*
	(0.0039)	(0.0040)	(0.0023)	(0.0024)
Subject's birth weight		-0.0263*		-0.0291*
		(0.0049)		(0.0029)
Sibling's birth weight		-0.0249*		-0.0286*
		(0.0049)		(0.0029)
Parental education controls	Yes	Yes	Yes	Yes
R-squared	0.1367	0.1367	0.1197	0.1236
Sample size	135,683	135,683	320,143	320,143

Note: All regressions are based on the firstborn sample, controlling for mother's age, father's age, mother's age at first birth, eligible college years, and birth place. Parental education includes a full set of dummies for categorical education levels separately for father and mother. The coefficients on interactions between sibling gender and parental education are all insignificant at the 5 percent level.

Table 5. Effects of sibling gender and family size on firstborn children's education outcome, by gender and by single birth year

Dependent variable: indicator of entering a college

	No restriction for birth year of second-born siblings					Second-born siblings born before 1984			
	1978	1979	1980	1981	1984	1978	1979	1980	1981
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A) Firstborn girls									
B2=1 if 2nd birth was boy	-0.0245*	-0.0083	-0.0007	-0.0048	-0.0015	-0.0251*	-0.0101	-0.0052	-0.0101
	(0.0122)	(0.0133)	(0.0131)	(0.0124)	(0.0113)	(0.0131)	(0.0145)	(0.0153)	(0.0160)
Family size	-0.0631*	-0.0231	0.0058	-0.0167	-0.0064	-0.0586*	-0.0251	-0.0058	-0.0270
	(0.0256)	(0.0292)	(0.0296)	(0.0283)	(0.0284)	(0.0270)	(0.0307)	(0.0329)	(0.0335)
Sample size	62,836	66,765	62,461	64,201	58,134	60,640	62,967	56,126	51,902
Average college enrollment	***								
B) Firstborn boys									
B2=1 if 2nd birth was boy	0.0011	-0.0071	0.0009	0.0019	0.0051	0.0017	-0.0069	0.0001	-0.0001
	(0.0039)	(0.0037)	(0.0037)	(0.0035)	(0.0036)	(0.0041)	(0.0039)	(0.0042)	(0.0042)
Family size	-0.0120	-0.0267	0.0002	0.0189	0.0162	-0.0090	-0.0312	-0.0001	0.0018
	(0.0231)	(0.0231)	(0.0253)	(0.0253)	(0.0239)	(0.0242)	(0.0244)	(0.0286)	(0.0307)
Sample size	65,679	70,004	64,708	66,835	60,766	63,343	66,070	57,978	54,061
Average college enrollment	***								

Note: Same as the previous table. In all regressions, we use the occurrence of twinning at the second birth to instrument for family size. We include all covariates (such as parental education and birth weights) as in the previous table. The coefficients on interactions between sibling gender and parental education are all insignificant at the 5 percent level.

Table 6. Effects of sibling gender and family size on education by parental education, after instrumenting for family size

Dependent variable: indicator of entering a college

	Father		Mother	
	HS or above	Below HS	HS or above	Below HS
Variables	(1)	(2)	(3)	(4)
A) Firstborn girls				
B2=1 if 2nd birth was boy	-0.0070	-0.0047	-0.0098	-0.0030
	(0.0067)	(0.0060)	(0.0066)	(0.0060)
Family size	-0.0252	-0.0118	-0.0281	-0.0108
	(0.0173)	(0.0126)	(0.0178)	(0.0128)
Parental education controls	Yes	Yes	Yes	Yes
Sample size	175,925	261,405	137,286	300,044
B) Firstborn boys				
B2=1 if 2nd birth was boy	-0.0001	-0.0004	-0.0006	-0.0008
	(0.0022)	(0.0017)	(0.0025)	(0.0016)
Family size	-0.0231	0.0058	-0.0070	-0.0077
	(0.0152)	(0.0110)	(0.0173)	(0.0104)
Parental education controls	Yes	Yes	Yes	Yes
Sample size	272,336	272,336	143,092	312,734

Note: Same as the previous table.

Appendix Table 1: Descriptive Statistics for first-born girls with at least one sibling

Variables	Older cohort			Younger cohort		
	with a 2nd-born brother	with a 2nd-born sister	Diff=(2)-(1).	with a 2nd-born brother	with a 2nd-born sister	Diff=(5)-(4)
	(1)	(2)	(3)	(4)	(5)	(6)
Family size	2.734	3.199	0.465	2.571	2.989	0.418
	(0.805)	(0.981)	[0.000]	(0.723)	(0.888)	[0.000]
Percentage of entering a college	0.141	0.137	-0.004	0.185	0.182	-0.003
	(0.348)	(0.344)	[0.040]	(0.388)	(0.386)	[0.025]
Age of mother at birth	23.179	23.150	-0.030	23.612	23.567	-0.045
	(3.268)	(3.281)	[0.102]	(3.375)	(3.383)	[0.002]
Second-born sibling's birth year	1980	1980	0.000	1984	1984	-0.020
	(1.879)	(1.913)	[0.256]	(2.451)	(2.479)	[0.024]
Mothers' birth year	1955	1955	0.000	1958	1958	0.000
	(3.297)	(3.309)	[0.075]	(3.565)	(3.569)	[0.005]
Fathers' birth year	1952	1952	0.000	1955	1955	0.000
	(4.806)	(4.837)	[0.495]	(4.491)	(4.532)	[0.279]
Subject's birth weight (kg)	3.178	3.175	-0.003	3.159	3.156	-0.003
	(0.439)	(0.444)	[0.1763]	(0.431)	(0.429)	[0.412]
Second-born sibling's birth weight	3.346	3.239	-0.107	3.331	3.226	-0.106
(kg)	(0.474)	(0.452)	[0.000]	(0.452)	(0.469)	[0.000]
Mothers' highest grade completed						
College or above	0.027	0.026	-0.001	0.030	0.028	-0.002
	(0.161)	(0.159)	[0.547]	(0.171)	(0.166)	[0.009]
Professional training degree	0.036	0.037	0.000	0.043	0.042	-0.001
	(0.187)	(0.188)	[0.892]	(0.203)	(0.200)	[0.197]
High School	0.052	0.053	0.001	0.065	0.064	0.000
	(0.222)	(0.224)	[0.416]	(0.246)	(0.245)	[0.592]
Vocational HS	0.157	0.150	-0.007	0.201	0.201	0.000
	(0.364)	(0.357)	[0.805]	(0.401)	(0.400)	[0.759]
Junior HS	0.197	0.197	-0.001	0.284	0.285	0.002
	(0.398)	(0.397)	[0.257]	(0.451)	(0.452)	[0.276]
Fathers' highest grade completed						
College or above	0.060	0.060	0.000	0.064	0.063	-0.002
	(0.238)	(0.238)	[0.876]	0.245	0.242	[0.071]
Professional training	0.065	0.064	-0.001	0.076	0.075	-0.001
degree	(0.246)	(0.244)	[0.516]	0.266	0.264	[0.268]
High School	0.087	0.085	-0.001	0.094	0.093	-0.001
	(0.282)	(0.279)	[0.356]	0.292	0.290	[0.169]
Vocational HS	0.152	0.149	-0.003	0.187	0.186	0.000
	(0.359)	(0.356)	[0.109]	0.390	0.389	[0.824]
Junior HS	0.171	0.174	0.003	0.254	0.256	0.002
	(0.376)	(0.379)	[0.140]	0.435	0.437	[0.135]
Rural areas	0.660	0.663	0.044	0.654	0.656	0.002
	(0.474)	(0.473)	[0.289]	0.476	0.475	[0.251]
Sample size	67,366	62,235		158,888	148,841	

Note: Same as Table 1. Standard deviation in (.); and p-values in [.].

Appendix Table 2: Benchmark estimates-- sex-composition effects from reproducing models in the literature, female sample.

Dependent variable: Indicator of entering a college

	Model I	Model II		Model III
	(1)	(2)	(3)	(4)
Number of older brothers	-0.0462	-	-	-
	(0.0009)			
Number of older sisters	-0.0391	-	-	-
	(0.0007)			
Number of younger brothers	-0.0077	-	-	-
	(0.0006)			
Number of younger sisters	-0.0079	-	-	-
	(0.0005)			
Dummy for any brother	-	-0.0014	-	0.0055
		(0.0010)		(0.0014)
Dummy for any sister	-	-	0.0045	-
			(0.0010)	
Number of sisters	-	-	-	0.0042
				(0.0011)
Number of sisters squared	-	-	-	0.0002
				(0.0003)
Birth order	-	-	-	-0.0345
				(0.0006)
Family size		-0.0267	-0.0301	-0.0117
		(0.0017)	(0.0018)	(0.0009)
Family size squared	-	0.0015	0.0018	-
		(0.0002)	(0.0002)	
Sample size	829,621	829,621	829,621	829,621
R-squared	0.1094	0.1061	0.1061	0.1094

Note: Models I, II, and III reproduce the results of Parish and Willis (1993), Butcher and Case (1994) and Garg and Morduch (1998), respectively. Samples are female born between 1978 and 1984, pooling all categories of birth order. All regressions also control for the indicators of parental education level, parental age, birth place, and the year of being eligible for college entry.

Appendix Table 3: Sex-composition effects, categorizing the female sample by birth order

Dependent variable: Indicator for college attendance

	First-born girls		Second-born girls		Third and later-born girls			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dummy of any old brother	-	-	-0.0063	-0.0065	-0.0098	-0.0098	-	-
			(0.0013)	(0.0013)	(0.0023)	(0.0022)		
Dummy of any old sister	-	-	-	-	-	-	0.0046	0.0046
							(0.0025)	(0.0025)
Dummy of any young brother	0.0034	-	0.0027	-	0.0008	-	0.0030	-
	(0.0013)		(0.0015)		(0.0027)		(0.0038)	
Dummy of any young sister	-	0.0025	-	0.0011	-	-0.0002	0.0016	-0.0008
		(0.0013)		(0.0016)		(0.0028)	(0.0040)	(0.0028)
Birth order	-	-	-	-	-0.0130	-0.0134	-0.0142	-0.0161
					(0.0028)	(0.0028)	(0.0036)	(0.0027)
Family size	-0.0139	-0.0148	-0.0115	-0.0098	-0.0205	-0.0196	-0.0204	-0.0162
	(0.0024)	(0.0025)	(0.0030)	(0.0030)	(0.0054)	(0.0053)	(0.0074)	(0.0053)
Family size squared	0.0004	0.0005	0.0003	0.0001	0.0015	0.0015	0.0015	0.0013
	(0.0003)	(0.0003)	(0.0004)	(0.0004)	(0.0005)	(0.0005)	(0.0006)	(0.0005)
R-squared	0.1124	0.1124	0.1065	0.1065	0.0738	0.0738	0.0736	0.0736
Sample size	443,441	443,441	287,340	287,340	98,840	98,840	98,840	98,840

Note: The samples include females born between 1978 and 1984. All regressions control for indicators for parental education, parental age, and children's birth place and cohort effects.