The Quantity-Quality Tradeoff of Children in a Developing Country: Identification Using Chinese Twins*

Hongbin Li
Department of Economics
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong

Junsen Zhang†
Department of Economics
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong

Yi Zhu
Department of Economics
Michigan State University
East Lansing, MI 48823

December 8, 2005

*We would like to thank Nancy Qian, Wing Suen, Li-an Zhou and participants in seminars and conferences for valuable comments. Any errors are the responsibility of the authors.

†Corresponding author. Tel.: 852-2609-8186; fax: 852-2603-5805; E-mail: jszhang@cuhk.edu.hk

The Quantity-Quality Tradeoff of Children in a Developing Country: Identification Using Chinese Twins

Abstract

Testing the tradeoff between child quantity and quality within a family is complicated by the endogeneity of family size. Using data from the Chinese Population Census, this paper examines the effect of family size on child educational attainment in China. We find a negative correlation between family size and child educational attainment, even after we control for the birth order effect. We then instrument family size by the exogenous variation that is induced by a twin birth, and find a causal link between family size and children's education. We also find that the effect of family size is more evident in rural China, where the public education system is poor. These findings suggest a quantity-quality tradeoff of children in developing countries.

JEL Classification: J13, J18, J24, O10

1 Introduction

The relationship between family size and child outcome has fascinated social scientists for decades, particularly since the emergence of the theory of the quantity-quality model that was developed by Gary Becker and his associates (Becker, 1960; Becker and Lewis, 1973; Becker and Tomes, 1976).[1] According to this model, an increasing marginal cost of quality (child outcome) with respect to quantity (number of children) leads to a tradeoff between quantity and quality. Numerous empirical studies have attempted to test the quantity-quality tradeoff, and have generally confirmed the prediction by observing a negative correlation between family size and child quality.[2] However, most studies simply treat family size as an exogenous variable, and thus cannot establish causality. Both child quantity and quality are endogenous variables, because childbearing and child outcome are jointly chosen by parents (Browning, 1992; Haveman and Wolfe, 1995), which means that they are both affected by unobservable parental preferences and household characteristics.

One important method for tackling endogeneity is to use the exogenous variations in family size that are caused by the natural occurrence of twins to isolate the causal effect of family size on child quality.[3] A pioneer study using twins as a means of identification is that of Rosenzweig and Wolpin (1980a), who find that family size (as induced by the birth of twins) has a negative effect on children's education attainment in a small sample (25 twins in approximately 1,600 children) from India. However, a recent study by Black et al. (2005) that also uses twins as the exogenous variation but with a large sample of the entire population of Norway, finds that the effect of family size is reduced to almost zero after

[1] Many aspects of household behavior have been considered to be associated with family size. For example, researchers have thoroughly documented evidence for the relationship between fertility and parental labor supply (Rosenzweig and Wolpin, 1980b; Angrist and Evans, 1998), maternal economic outcome (Bronars and Grogger, 1994), stability of marriage (Koo and Janowitz, 1983; Jacobsen et al., 2001), and children's attainments (King, 1987; Haveman and Wolfe, 1995).

[2] See King (1987) and Blake (1989) for a survey of early studies. Education and health are usually used as measures of child quality in the literature.

[3] In addition to twins, some researchers also employ the gender of the first child (Lee, 2004) or the gender composition of the first two children (Conley, 2004a; Angrist et al., 2005) as the instrument for family size. The former instrument is based on the prevailing preference for sons that is observed in Asian countries, and the idea behind the latter instrument is that parents of same-gender siblings are more likely to go on to have an additional child (Angrist and Evans, 1998).

controlling for birth order, and that there is a monotonic decline of educational attainment by birth order.[4] These new findings suggest that the omission of the birth order effect may lead to biased estimates of the effect of family size on child quality. Another recent study by Angrist et al. (2005) that uses both twin births and gender composition as the instrumental variables finds no evidence for a quantity-quality tradeoff of children in Israel.

The studies of Black et al. (2005) and Angrist et al. (2005) raise a provocative question: Is there a quantity-quality tradeoff as formulated by Becker? These new studies have made many improvements on the earlier study of Rosenzweig and Wolpin (1980a), in particular in terms of data quality and empirical specifications, and thus their evidence should be more robust. However, in addition to a larger sample size and improved model specification, another important difference between the new studies and that of Rosenzweig and Wolpin is that the latter draw on data from a developing country, whereas the former use data from developed countries. In a rich country with a comprehensive welfare system such as Norway, where there is both a good public education system (even college is free) and generous government support for childbearing and childcare, the cost of children, and in particular the educational expenditure, accounts for just a small proportion of the budget of parents. Thus, the quantity-quality tradeoff may not be obvious in these countries. In contrast, in a developing country such as India, where there is neither a well-functioning public education system nor generous support for childbearing and childcare, the cost of child quality is mostly borne by the parents. Thus, the quantity-quality tradeoff is more likely to occur in a developing country.[5] Therefore, it is important to use good data from developing countries to verify whether the findings of Black et al. (2005) can be replicated.

In this paper, we test the quantity-quality tradeoff by using the 1% sample of the 1990

[4]Sociologists and psychologists have documented the effect of birth order on child outcomes. See, for example, the summary of the findings by King (1987) and Conley (2004b). Several earlier empirical studies were conducted by economists. For example, Hauser and Sewell (1985) find no significant effect of birth order, Behrman and Taubman (1986) show that children born later tend to have a disadvantage in education, and Hanushek (1992) reports a U-shaped pattern of education by birth order for large families.

[5]There is also some evidence from developing countries in studies of epidemiology and public health, although the methods of these studies are usually different from those of economists. See, for example, the survey by Karmaus and Botezan (2002).

Chinese Population Census. China has a poorly functioning education system, especially in rural areas, where poverty is the main reason that forces children to drop out of primary and high school (Brown and Park, 2002). Using education level and school enrollment as measures for child quality, we find a negative correlation between family size and child quality. The negative correlation is robust to various specifications, including those that control for the birth order effect. The estimates suggest that having one more child in a family reduces a child's probability of being enrolled in school by around four percentage points. The negative effect of family size on child educational attainment is supported by two-stage least squares (2SLS) estimations using twin births as the instrument variable (IV) for family size, which suggests that family size has a causal effect on child quality. Our findings strongly support the prediction of Becker and his associates on the quantity-quality tradeoff of children, but differ from those of Black et al. (2005). To further test whether the quantity-quality tradeoff is less pronounced in more developed regions, we break the sample down into rural and urban sub-samples, and find that the quantity-quality tradeoff is far more pronounced in the less developed rural areas. Finally, we find that the effect of family size varies with gender and mother's education.

We are among the first to draw on twins data from a developing country to test the theory of the quantity-quality tradeoff of children. Given that the quantity-quality tradeoff is expected to be more pronounced in developing countries, it is surprising that few previous studies have drawn on twins data from developing countries, although this is probably due to difficulty in obtaining data. We are also among the first to explicitly examine the tradeoff in the context of China. Most of the previous related studies explore the determinants of Chinese children's educational attainment, and emphasize the rural-urban gap (Knight and Li, 1993, 1996; Hannum, 1999; Connelly and Zheng, 2003), gender inequality (Broaded and Liu, 1996; Hannum, 2002, 2003; Tsui and Rich, 2002), or poverty and credit constraints (Brown and Park, 2002). However, these studies either ignore the effect of family size or merely treat it as an exogenous control variable. To the best of our knowledge, the only exception is a working paper by Qian (2005), who attempts to use China's birth control policy as an identification to test the quantity-quality tradeoff.[6]

Knowing the true effect of family size on child quality has important policy implications for developing countries, and in particular for China. China started a unique birth control policy, the one-child policy, in 1979. Under this policy, each household is allowed only one child, although the policy has since been relaxed to allow two children if the first child is a girl in rural areas. Our findings suggest that the birth control policy has a potential positive effect in increasing the quality of children. If, as we have found, a smaller family size is generally associated with a better average education outcome for children, then the one-child (or two-child) policy has improved family welfare by reducing the number of children in a household. In particular, we find that the tradeoff between quantity and quality is more pronounced in rural areas, where the least well-off people live. This implies that the birth control policy, if it is as effective as expected by policy-makers, is fruitful in enhancing the welfare of the poor.

The rest of the paper is structured as follows. Section 2 specifies our empirical strategy. Section 3 describes the data. Section 4 presents our estimates of the effect of family size on children's educational outcomes in China, and Section 5 concludes.

2 Empirical Method

We follow the recent empirical literature and specify our general estimation as follows,

$$EDU = \beta_0 + \beta_1 SIZE + X\beta_2 + Z\beta_3 + BO\beta_4 + \epsilon, \tag{1}$$

[6]Qian (2005) uses the triple interaction of the sex of the first child, the birth cohort of the child and the local policy relaxation (sex*cohort*policy relaxation) as an IV, and finds that there is no quantity-quality tradeoff. However, her study is subject to discussion for the following two reasons. First, this triple interaction is likely to be endogenous, because it reflects the sex preference of parents. Families with sex preference are more likely to select the sex of their firstborn after the one-child policy was implemented (before the one-child policy, they can simply have more children in order to have a boy), i.e., the sex*cohort interaction within a locality picks up the sex preference. Moreover, localities with strong sex preference are more likely to relax the one-child policy, i.e., policy relaxation is a result of sex*cohort. Second, she uses a sample of children aged 9-28. Many of the older children of the households may have left (not tracked in the census), and thus the sex of the first child is not really observed for many households. Thus, the "first child" is in fact the oldest staying child, which cannot be random in China. She also uses twins as an IV in some specifications, and finds a quantity-quality tradeoff, which is similar to our finding.

where EDU is the educational attainment of the child as measured by the two educational outcome variables of education level and school enrollment. The variable $SIZE$ is the number of children in the family, and the coefficient β_1 is essentially what interests us. X is a vector of the variables that measure child characteristics, including age, gender, ethnic group, and place of residence, and Z stands for a set of parental attributes, including age and education level. We also add a vector of birth order indicators (BO) to examine whether the effect of family size is partially driven by birth order.[7] To allow the effect of family size to differ between rural and urban regions, we run regressions for the rural and urban samples separately.

The coefficient β_1 as estimated by the ordinary least squares (OLS) method may merely suggest a correlation, rather than a causal effect, because family size is likely to be endogenous. Following Rosenzweig and Wolpin (1980a) and Black et al. (2005), we use the birth of twins as an identifying instrument for family size. The first stage of the two-stage least squares (2SLS) estimation is given by

$$SIZE = \alpha_0 + \alpha_1 TWIN + X\alpha_2 + Z\alpha_3 + \upsilon, \tag{2}$$

and equation (1) becomes the second stage (without birth order included on the right-hand side to avoid a confusing interpretation where twins are involved). In equation (2), $TWIN$ is a dummy variable that equals 1 if the nth birth is a multiple birth and 0 otherwise, and all of the other variables are the same as specified in equation (1).

As noted by Rosenzweig and Wolpin (2000), the presence of any twin birth in a family makes for an inappropriate instrument, because its probability increases with the number of births. To avoid this problem in estimating the 2SLS models, we restrict the sample to households with at least n births so that we can be fairly confident that the families with twins at the nth birth have the same preference for number of children as those with singleton births. If the occurrence of multiple births is randomly assigned by nature, then twin births

[7] Following Black et al. (2005) in avoiding the ambiguous definition of birth order where twins are involved, we exclude twins from the sample that is used to estimate equation (1). Similarly, our results are not sensitive to the exclusion of twins.

should have little or no effect on children's education except through family size. Thus, the 2SLS estimate of β_1 would identify the causal effect of family size on child quality.

3 Data

In this paper we use the 1% sample of the 1990 Chinese Population Census that was collected by the Chinese National Bureau of Statistics (formerly the State Statistic Bureau). It is the fourth of its kind, following the three censuses that were conducted in 1953, 1964, and 1982.[8] The 1% sample covers 11,475,104 individuals from 2,832,103 households. The dataset contains a record for each household, and includes variables that describe the location, type, and composition of the households. Each household record is followed by a record for each individual residing in the household. The variables that relate to individuals include demographic characteristics, occupation, industry, education level, ethnicity, marital status, and fertility.

We use the relation identifier to match children to their parents within the households. Specifically, we identify individuals who are labeled "child" as the primary observation, and obtain the family size by counting the number of children in the household. We then attach the data of the parents, that is, those who are labeled "household head" or "spouse," to all of the children in the household. For each mother, we also have data on the total number of children born and the number of children still alive, which helps us to identify whether the family size is complete.

To facilitate our analysis, we use a sub-sample of the census data. First, we only use children of the household head, because we can only match the parental information and count the number of children of a couple for such children. Second, we drop households with no children or with a family size that exceeds the total number of children born, the latter of which is likely to be the result of data error.[9] Third, we restrict the sample to children

[8]The two earlier censuses are not available to researchers. The 1982 census is not very useful for our purposes due to a lack of school enrollment information and rural identifiers. The latest census, which was conducted in 2000, will be available soon.

[9]This discrepancy may arise as a result of adopted children, but there is no information in the data to

who were at least 6 years old and whose mothers were aged no more than 35 in the census year. We use 6 as the lower bound for the age of children because it is the minimum age of school enrollment in China, and no child younger than 6 in the sample was observed to have been enrolled. Restricting the mother's age to less than or equal to 35 makes it fairly certain that no adult children have moved out of a household. We impose such a restriction because we are unable to track children who had already left the household by the time of the survey.[10] Finally, we exclude some households with missing information on fathers,[11] and a small number of families with a birth that occurred before the mother was 16.

With these restrictions, we are left with a sample of 675,492 children from 447,159 households. As the census does not include an explicit twins identifier, we define children who were reported to be born in the same year and month to the same woman as twins. One percent of our sample comprises twin births. The first two columns of Table 1 report the summary statistics for the whole sample and the sample excluding twins. No significant differences can be observed between the two columns, as the statistics remain almost the same for each variable.

It is worthwhile to first outline the institutional background of non-tertiary education in China before we offer the definitions of the education variables. In 1986, the Law of Compulsory Education officially declared the implementation of nine compulsory years of schooling (six years of primary school and three years of junior high school) throughout China. However, the policy of compulsory education was not implemented uniformly across the country. The "Resolution on Educational System Reform," which was initiated in 1985, devolved the total responsibility of implementing compulsory education to local governments, and thus the provision of basic education depends on the local budget or level of economic development (He, 1996). As a result, access to education in rural areas is much worse than

distinguish between adoption and birth.

[10]With this restriction, only about one percent of the households have children that live outside of the household. We also conduct regressions using a sample that excludes these families and obtain the same results.

[11]Data on fathers were missing for 7 percent of all cases. In addition to dropping these observations, we also perform the estimations by creating categories of missing father variables, and the results are the same.

in urban areas, because rural citizens and governments are much poorer. In the poor rural areas, public schools are not widely available, and even in regions where the schooling system is publicly provided to all children, it is not totally free and parents still need to pay tuition and fees. Such a financial burden is one of the main reasons why poor families, who are often unable to borrow funds to finance their children's education, pull their children out of school (Brown and Park, 2002).

In this paper, we employ two education variables that are reported in the census: education level and school enrollment. Education level is defined as an ordered discrete variable that indicates three education levels: illiterate, primary school, and junior high school and above.[12] School enrollment is defined as a binary indicator that equals one if a child was enrolled in school or had graduated, and zero if a child had dropped out of school or never enrolled. Previous research has shown that school enrollment is a good indicator of educational attainment in developing countries (Glewwe and Jacoby, 1995; Alderman, et al., 2001; Glewwe et al., 2001). Table 1 shows that the average enrollment rate is 70 percent for the full sample, and that children at the three education levels account for 28, 69, and 3 percent, respectively. For children who are at least 8 years old, the enrolment rate rises to 91 percent and the education level also improves.

An important aspect of the data is that there is a large rural-urban difference in both education and fertility. In columns 3 and 4 of Table 1, we report the attributes of the rural and urban sub-samples, respectively. Of all of the children, 88 percent were from rural areas. Note that although there is no rural-urban difference for the education variables for the whole sample of children, there is a large difference among children of 8 years and over. The reason for the lack of difference in the whole sample is that rural children went to school earlier. In urban areas, the enrollment age was normally seven or eight for the generation

[12]The census codes the education level into seven categories: Illiterate, Primary School, Junior High School, Senior High School, Technical School, Junior College, and University. As the proportion of respondents with an education level of senior high school or above is rather small in the sample (less than 0.01 percent), we classify all of these observations into the third level of junior high school and above. Having more categories for education levels does not change our results significantly.

of children in our sample, and that age requirement has been strictly enforced. However, children in rural areas were able to go to school as early as 6. Note also that the fertility of rural families is much higher than that of urban families, with the rural-urban gap in the number of children being as large as 1.08. Over four fifths of the rural households had more than one child in comparison to only one fifth of the urban households.[13] These rural-urban differences make it important to analyze the rural and urban sub-samples separately.

To gain a picture of how education may vary with family size, we present in Table 2 children's enrollment rate by family size for both the rural and urban sub-samples. To control for the age effect, we report the primary school enrollment rate for young children (aged 13 or below) and the junior high school enrollment rate for older children (aged over 13). Several aspects are worth noting. First of all, there is a clear pattern that a greater family size is associated with a lower average enrollment. Although only children aged below 13 appear to have a lower enrollment rate than children in two-child families, there is a monotonically decreasing trend for family sizes of two to six and above. Moreover, the advantage of two-child families over single-child families disappears for children who are older than 13. Second, on average, urban children seem to have a higher enrollment rate than rural children. Except for young children in the only-child group, urban children are more likely to be enrolled in school (or graduated), regardless of family size and gender. Finally, male children consistently have a higher enrollment rate than female children in the rural sample, but the gender-based difference is less explicit in the urban sample.

4 The Effect of Family Size on Children's Education

In this section, we systematically test whether family size has a negative effect on children's educational attainment in China. We first employ a probit model to estimate equation (1) controlling for different sets of variables. We also examine whether the effect of family size

[13] Although the one-child policy had been in force for ten years by the time of the census in 1990, there is empirical evidence that the policy was more effective in deterring second births in urban areas than in rural areas (Zhang and Spencer, 1992; Ahn, 1994)

is different in rural versus urban areas by performing regressions for the separate samples. We then run 2SLS estimations to isolate the causal effect of family size using the birth of twins as an IV. Finally, we check the heterogeneity of the effect of family size by performing IV estimations under various sample stratifications. For all of the estimations in this paper, we regress the two dependent variables as defined in Section 3 and basically control for a set of child attributes that comprises the cubic form of age, gender, indicator of being *Han* Chinese, and rural (if applicable) and provincial dummies.

4.1 Probit Estimations

We begin by examining the effect of family size on children's education for the whole sample. Table 3 reports the ordered probit estimates with education level as the dependent variable. The first column reports a simple regression with the number of children, the male, *Han*, and rural dummies, age, age squared, age cubed, and provincial dummies as independent variables. Note that family size is negatively correlated with children's education level, as suggested by the highly significant coefficient of the number of children in column 1.[14] When we replace the linear form by the indicator of family size (one-child category omitted) in column 2, we observe a similar negative relationship. Although a two-child family seems to have an advantage over a one-child family,[15] the effect is significantly negative for larger families, and increases in magnitude with family size, which is consistent with the pattern that is observed from the summary statistics in Table 2. Some control variables have the expected signs. In general, male or *Han* children have an educational advantage over female or minority children. Due to space limitations, we do not report the results for the other demographic variables in the table.

Interestingly, the estimated effects of family size are not very sensitive to the inclusion of the parental demographic attributes and birth order variables. In columns 3 and 4, we include the demographic attributes of parents, including age and age squared, and indicators

[14] The statistics that are reported here, as in all of the regressions in this paper, allow for the correlation of errors for any two children in the same household.

[15] Such an advantage is also found in Hauser and Kuo (1998) and Black et al. (2005).

of education level for each parent. As shown, controlling for these variables only slightly reduces the effect of family size. In columns 5 and 6, we also control for the birth order effect. Again, including birth order variables only changes the effect of family size marginally, which suggests that the negative relationship between family size and children's education may partially be driven by birth order, but not to a large extent. This finding is in stark contrast to that of Black et al. (2005), who find that the effect of family size becomes trivial once the birth order effect is controlled. We also find little evidence of a monotonic decline of child quality by birth order, as distinct from Black et al. (2005). Rather, the positive coefficients of higher birth orders indicate that children who are born later in large families are more likely to have an advantage over children who are born earlier.

Our findings strongly support the prediction of Becker (1960) and Becker and Lewis (1973) of the quantity-quality tradeoff of children, but differ from those of Black et al. (2005). The most important difference between our study and that of Black et al. (2005) is that they draw on data from Norway, which is a developed country, whereas we draw on data from China, which is a developing country. In a rich country with a comprehensive welfare system such as Norway, where there is both a good public education system and generous government support for childbearing and childcare, the quantity-quality tradeoff may not be obvious. However, in a developing country such as China, where there is neither a good public education system nor generous support for childbearing and childcare, the cost of child quality is mostly borne by the parents. Thus, the quantity-quality tradeoff is more likely to happen in the Chinese case.

We next turn to school enrollment, which is the other educational outcome, and present in Table 4 the probit estimation results with the same specifications as before. As school enrollment is a binary variable, we report dF/dx, or the marginal change in the probability of being enrolled when the independent variable increases. Similar to the estimates in Table 3, family size appears to have a significantly negative impact on the probability of school enrollment for children, and the effect decreases only marginally when we add the demographic attributes of parents and account for the birth order effect. The coefficients of family size suggest that, everything else being constant, having one more child in the household reduces a child's probability of enrollment by approximately four percentage points. Also consistent with the results in Table 3, we find a positive two-child family effect and a non-negative effect of high birth orders in large families.

4.2 Effects in Rural and Urban Areas

The foregoing analysis uses the whole sample, and does not take into account the interaction with urban-rural differences. As discussed in Section 2, there is a considerable rural-urban gap in access to and completion of schooling in China. This gap is the result of both supply- and demand-side factors. On the supply side, the average school quality is much better in urban China than in rural China. While urban public schools receive substantial subsidies from local governments, many rural schools are badly funded, and thus short of well-trained teachers. The lack of government funding compels many rural schools to become self-financed, which forces many rural children out of school because their parents cannot afford to pay the school fees (Brown and Park, 2002). On the demand side, rural parents may have lower educational aspirations for their children than urban parents. This is probably due to the lower return and higher opportunity cost of sending children to school for rural families, because rural children can contribute to the household income by carrying out farm and house work even at a very young age.[16]

Because of the rural-urban education gap, we expect the effect of family size on child quality to be different in rural and urban areas. Given that public education is more prevalent and children's education held to be more important in urban China, having an additional child in the family may result in a smaller adverse impact on the average child education compared to the effect in a rural family. In fact, the rural-urban difference within China resembles the difference between China and Norway. To allow for disparity in the effect of

[16]See Becker (1991), Johnson (1994), Dasgupta (1995), and Ray (1998) for arguments on the benefits of children in developing countries.

family size between rural and urban areas, we perform the same estimations for separate rural and urban sub-samples, and report the results in Tables 5 to 8.

The estimates indeed show that the effect of family size is smaller in urban areas than in rural areas. As can be seen from Tables 5 and 6, the estimates for the rural sample are very close to those for the whole sample, especially when we look at the coefficients of the number of children (columns 1, 3, and 5). In contrast, the estimated coefficients of family size for the urban sample, as reported in Tables 7 and 8, are smaller in magnitude, although they remain statistically significant. For example, having one more child in urban families lowers the probability of enrollment by less than two percentage points (Table 8), which is less than half of the estimated effect for rural families (Table 6).

It should be noted that the negative effect of family size becomes even larger when we include birth order variables in the regressions for the urban sample (Tables 7 and 8), although the statistics for many of the birth order variables are insignificant. It is also worth noting that ethnic- and gender-based differences are less explicit among urban children. Although there is a clear educational advantage for male or *Han* children in rural areas (Tables 5 and 6), the evidence from urban children (Tables 7 and 8) shows an insignificant ethnic effect and even a negative male effect.[17]

In summary, we find that family size is negatively correlated with children's educational attainment in China, both when we measure education by discrete levels and by the probability of being enrolled in school. The negative effect is not very sensitive to the inclusion of controls for parental characteristics and birth order variables. By dividing the sample into rural and urban sub-samples, we find that the adverse impact of family size is smaller in urban China. We also observe some evidence of a positive two-child family effect, but do not identify a significant negative effect of higher birth orders in large families.

[17]The absence of an educational advantage for boys in urban China has also been observed in recent literature (for example, Tsui and Rich, 2002; Connelly and Zheng, 2003).

4.3 OLS and 2SLS Estimations

The negative correlation between family size and education is only suggestive, because fertility is most likely to be endogenous. In this section, we use twins at the nth birth ($n = 1, 2, 3$) as an instrumental variable (IV) to identify the casual effect of family size on child quality. In particular, we employ 2SLS estimations that are specified by equations (1) and (2), and compare the results with those of the OLS estimates.

Before reporting the estimation results, we first discuss the validity of using twin births as our IV. A good IV should be highly correlated with the number of children in a household, but should not affect the child outcome except through family size. That is to say, a valid IV should not be correlated with unobserved parental and household characteristics that are captured by the error term in equation (1). The birth of twins is an important source of exogenous variation in fertility that has been used in previous research (Rosenzweig and Wolpin, 2000), and is believed to be unlikely to depend on family background. Although the correlation between twin births and unobserved household attributes is untestable by design, we follow Black et al. (2005) and examine whether the occurrence of twins is associated with certain observed characteristics, such as the education level of parents. Similar to the findings of Black et al., the F-tests based on linear probability models suggest that the probability of having a twin birth is uncorrelated with the education level of either mothers or fathers in our sample.

Table 9 presents the OLS and 2SLS estimates for the whole sample, along with the coefficients of the instrumental variable (twins at the n-th birth) in the first-stage.[18] The results with education level as the dependent variable are reported in the upper panel, and the results with school enrollment as the dependent variable are reported in the lower panel. From left to right, we list in columns the estimates for families with at least n births (using twins at the n-th birth as the corresponding IV for the 2SLS estimation) in increasing order

[18]The birth order variables are not included, because the order in the presence of multiple births is confusing. Controlling for birth order has very little effect on our estimates, as shown in Section 4.4.

of n from 1 to 3.

The 2SLS estimates in Table 9 suggest a significantly negative effect of family size on children's education in most cases. The 2SLS coefficients of the effect of an increase in family size that is induced by the occurrence of twins are negative in all three cases (columns 2, 4, and 6), and are significant at the five-percent level for columns 2 and 6. The results are similar for both dependent variables. These results suggest that the causal link between family size and child education exists for families with at least one child, and for a smaller sample of families with at least three births. Finally, note that the first-stage estimation is very significant for all of the specifications, with t-ratios of well above 40, which suggests that having a twin birth indeed increases the family size. Moreover, consistent with the previous literature, the effect of a twin birth on family size increases with a higher parity, which ranges from 0.6 to 0.8 in our sample.

In Tables 10 and 11, we report the results of the same regressions using the rural and urban sub-samples, respectively. Interestingly, the quantity tradeoff only exists for rural families. For the rural sub-sample (Table 10), we find the tradeoff relationship between family size and children's education to be similar to that of the whole sample. However, for the urban sample (Table 11), none of the 2SLS estimates is statistically different from zero at the five-percent level, which indicates that there is no causal effect of family size on urban children's education. Also similar to the probit estimates, the ethnic- and gender-based gaps in rural children's education no longer hold in urban areas.

4.4 The Heterogeneous Effects of Family Size

In this subsection, we test the sensitivity of our estimates to different model specifications and sub-samples. We first estimate models in which the birth order variables are controlled, and then estimate the effect of family size by gender or by mother's education. As the effect has been shown to differ between rural and urban areas, we skip the estimations for the full sample and perform the sensitivity test only for the rural and urban sub-samples.

We report both the OLS and 2SLS estimates in Tables 12 and 13 for the rural and urban samples, respectively. For all of the specifications and sub-samples, we control for the full set of variables that is used in Tables 10 and 11.

Our first sensitivity analysis shows that the previous results are robust when we control for the birth order effect. Column 1 of Tables 12 and 13 presents the results from the estimations that control for the birth order variables. Compared to Table 10, the addition of the birth order variables has very little effect on the 2SLS estimates of family size for the rural sample, as both the coefficients and t-statistics are very close to the previous levels. This is also the case with the estimates for the urban sample, which remain largely insignificant. This finding suggests that the effect of family size is not driven by the birth order effect.

In columns 2 and 3 we break the samples down by gender to see whether the effect of family size differs between boys and girls. Although the OLS estimates show that the effect of family size is more negative for girls than for boys, the picture from the 2SLS estimates is not as clear. The effect appears to be more pronounced for rural girls when we use the IV of twins at the first birth, but becomes larger for rural boys in families with at least three births. Despite the mixed results for the 2SLS estimations for the rural sample, we continue to identify a rural-urban gap that is independent of gender, namely, a smaller effect of family size in urban areas.

In the last three columns we stratify our sample by mother's education level. As household income is not observed in our sample, we use mother's education as a control for financial constraints. If better educated mothers are less financially constrained, then we should see a smaller effect of family size on the educational outcomes of their children. We break mother's education in the rural sample into the three groups of illiterate, primary school, and all the other levels above primary school. As urban women are generally better educated than rural women, to avoid a group with too few observations we break the urban sample down into the categories of below junior high school (illiterate and primary school), junior high school, and above junior high school.

To some extent, the results by educational group are consistent with our expectations. With the OLS estimates, the effect of family size decreases in magnitude with the level of the mother's education for both rural and urban children, although a few OLS coefficients for the urban sample (Table 13) are not statistically significant. However, the evidence is less explicit when we look at the 2SLS estimates. Again, for the rural sample the variation in effects across educational groups depends on the IV (and the sample) that we use. For the urban sample, we do not detect a tangible effect of family size for any sub-group, as almost all of the 2SLS estimates are insignificant.

5 Conclusions

In this paper, we test the quantity-quality model using a representative census dataset from China. We find evidence that family size is negatively correlated with children's education. The negative effect of family size is robust to various specifications, including those that control for parental characteristics and birth order effect. In terms of school enrollment, having one more child in a family is associated with an approximately four percentage point decline in the average probability of a child being enrolled. We also instrument family size with twin births to explore the causal link between family size and child education, and find supportive evidence. We further find that the effect of quantity on quality is not uniform across regions. More precisely, the tradeoff relationship is more evident in rural China, but the effect diminishes or even vanishes for urban China. We also find that the effect differs according to the gender of the child and the mother's education level.

Although this study has its limitations, it is among the first to explicitly measure the effect of family size on child outcome in China. Previous empirical tests are often limited by a small sample size or by the fact that they do not take into account the endogeneity of family size, both of which are tackled in this paper. Given that public education is insufficiently funded in many areas of China, our findings suggest a plausible determinant of children's education in China that has not been well explored in the literature. Nonetheless, due to the data limitations, we are unable to examine more aspects of child quality (such as health and labor market outcome), and are thus ill inclined to generalize our results to a broader extent. Future work may rely on more comprehensive and traceable household data that give researchers information on the completed education of children even if they have left the family.

This paper may shed some light on other issues in China, such as the one-child policy. Since its inception in the late 1970s, China's one-child policy has been controversial, and has drawn attention from politicians, the mass media, and academics alike. Although there is still no consensus on many of the positive or negative aspects of this forced birth control policy, a recent study by Li and Zhang (2005) does show that the population reduction as a result of the dramatic population control policy has indeed helped the growth of the Chinese economy since the late 1970s. This study indicates that a possible effect may be that children are of better quality under the policy, because the size of their family would have been larger had the policy not existed. However, to better understand the long-term effect on child outcome in adulthood, more work is badly needed in this area.

References

Ahn, Namkee, “Effects of the one-child policy on second and third births in Hebei, Shaanxi and Shanghai,” *Journal of Population Economics*, 7, 1994, 63-78.

Alderman, Harold, Jere Behrman, Victor Lavy, and Rekha Menon, "Child health and school enrollment: a longitudinal analysis," *Journal of Human Resources*, 36(1), 2001, 185-205.

Angrist, Joshua and William Evans, “Children and their parent’s labor supply: evidence from exogenous variation in family size,” *American Economic Review*, 88(3), 1998, 450-477.

Angrist, Joshua, Victor Lavy, and Analia Schlosser, “New evidence on the causal link between the quantity and quality of children,” MIT working paper, 2005.

Becker, Gary S., “An economic analysis of fertility,” *Demographic and Economic Change in Developed Countries*, Gary S. Becker, ed., Princeton: Princeton University Press, 1960.

--, *A Treatise on the Family*, Cambridge: Harvard University Press, 1991.

Becker, Gary S. and H. Gregg Lewis, “On the interaction between the quantity and quality of children,” *Journal of Political Economy*, 81(2), 1973, S279-S288.

Becker, Gary S. and Nigel Tomes, “Child endowments and the quantity and quality of children,” *Journal of Political Economy*, 84(4), 1976, S143-S162.

Behrman, Jere R. and Paul Taubman, “Birth order, schooling, and earnings,” *Journal of Labor Economics*, 4, 1986, 121–145.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes, “The more the merrier? The effect of family size and birth order on children’s education,” *Quarterly Journal of Economics*, 120(2), 2005, 669-700.

Blake, Judith, *Family Size and Achievement*, Berkeley and Los Angeles, CA: University of California Press, 1989.

Broaded, Montgomery C. and Chongshun Liu, “Family background, gender and educational attainment in urban China,” *The China Quarterly*, 145, 1996, 53–86.

Bronars, Stephen G. and Jeff Grogger, “The economic consequences of unwed motherhood: using twin births as a natural experiment,” *American Economic Review*, 84(5), 1994, 1141–1156.

Brown, Philip H. and Albert Park, “Education and poverty in rural China,” *Economics of Education Review*, 21(6), 2002, 523–541.

Browning, Martin, "Children and household economic behavior," *Journal of Economic Literature*, 30(3), 1992, 1434-1475.

Conley, Dalton, "What is the 'true' effect of sibship size and birth brder on Education? instrumental variable estimates from exogenous variation in fertility," 2004a, mimeo.

--, *The Pecking Order: Which Siblings Succeed and Why,* New York: Pantheon Books, 2004b.

Connelly, Rachel and Zhenzhen Zheng, "Determinants of school enrollment and completion of 10 to 18 year olds in China," *Economics of Education Review*, 22, 2003, 379-388.

Dasgupta, Partha, "The population problem: theory and evidence," *Journal of Economic Literature*, 33, 1995, 1879-1902.

Glewwe, Paul W. and Hanan Jacoby, "An economic analysis of delayed primary school enrolment in a low income country: the role of early childhood nutrition," *Review of Economics and Statistics*, 77(1), 1995, 156-169.

Glewwe, Paul W., Hanan Jacoby, and Elizabeth King, "Early childhood nutrition and academic achievement: a longitudinal analysis," *Journal of Public Economics*, 81(3), 2001, 345-368.

Hannum, Emily, "Political change and the urban-rural gap in basic education in China," *Comparative Education Review*, 43(2), 1999, 93-211.

--, "Ethnic differences in basic education in reform-era rural China," *Demography*, 39(1), 2002, 95–117.

--, "Poverty and basic education in rural China: villages, households, and girls' and boys' enrollment," *Comparative Education Review*, 47(2), 2003, 141-159.

Hanushek, Eric A., "The trade-off between child quantity and quality," *Journal of Political Economy*, 100(1), 1992, 84–117.

Hauser, Robert M. and Hsiang-Hui Daphne Kuo, "Does the gender composition of sibships affect women's educational attainment?" *Journal of Human Resources*, 33, 1998, 644-657.

Hauser, Robert M. and William H. Sewell, "Birth order and educational attainment in full sibships," *American Educational Research Journal*, 22, 1985, 1–23.

Haveman, Robert and Barbara Wolfe, "The determinants of children's attainments: a review of methods and findings," *Journal of Economic Literature*, 33 (4), 1995, 1829-

1878.

He, Dongchang ed, *Education in Contemporary China* (in Chinese), Beijing: Contemporary China Press, 1996.

Jacobsen, Joyce P., James W. PearceIII and Joshua L. Rosenbloom, "The effects of child-bearing on women's marital status: using twin births as a natural experiment." *Economics Letters*, 70, 2001, 133-138.

Johnson, D. Gale, "Effects of institutions and policies on rural population growth with application to China," *Population and Development Review*, 20(3), 1994, 503-531.

Karmaus, W. and C. Botezan, "Does a higher number of siblings protect against the development of allergy and asthma? A review," *Journal of Epidemiology and Community Health*, 56, 2002, 209-217.

King, Elizabeth M., "The effect of family size on family welfare: what do we know?" in D.Gale Johnson and Ronald D. Lee eds., *Population Growth and Economic Development: Issues and Evidence,* Madison, WI: University of Wisconsin Press, 1987.

Knight, John and Shi Li, "The determinants of educational attainment in China," in Keith Griffin and Renwei Zhao eds., *The Distribution of Income in China*, 285-330, London: Macmillan Press, 1993.

--, "Educational attainment and rural-urban divide in China," *Oxford Bulletin of Economics and Statistics*, 58(1), 1996, 83–117.

Koo, Helen P. and Barbara K. Janowitz, "Interrelationships between fertility and marital dissolution: results of a simultaneous Logit model." *Demography*, 20, 1983, 129-145.

Lee, Jungmin, "Sibling size and investment in children's education: an Asian instrument," *IZA Discussing Paper*, 2004.

Li, Hongbin and Junsen Zhang, "Do high birth rates hamper economic growth," *Review of Economics and Statistics*, forthcoming, 2005.

Qian, Nancy, "Quantity-quality: the positive effect of family size on school enrolment in China" Working Paper, Brown University, 2005.

Ray, Debraj, *Development Economics*, Princeton: Princeton University Press, 1998.

Rosenzweig, Mark R. and Kenneth I. Wolpin, "Testing the quantity-quality fertility model: the use of twins as a natural experiment," *Econometrica*, 48(1), 1980a, 227-240.

--, "Life-cycle labor supply and fertility: causal inferences from household models," *Journal of Political Economics*, 88(2), 1980b, 328-348.

l
--, “Natural ‘natural experiments’ in economics,” *Journal of Economic Literature*, 38(4), 2000, 827-874.

Tsui, Ming and Lynne Rich, “The only child and educational opportunity for girls in urban China,” *Gender and Society*, 16(1), 2002, 74-92.

Zhang, Junsen and Byron Spencer, “Who signs China’s one-child certificate, and why?” *Journal of Population Economics* 5, 203-215, 1992.

Table 1: Descriptive Statistics of the 1% Sample of the 1990 Population Census

Variables	Full sample		Rural vs. Urban	
	All families (1)	Families without twins (2)	Rural (3)	Urban (4)
Children (Observations)	675,492	665,738	595,729	79,763
Age	8.71 (2.39)	8.72 (2.39)	8.78 (2.42)	8.27 (2.08)
Male	0.52 (0.50)	0.52 (0.50)	0.52 (0.50)	0.52 (0.50)
Han	0.91 (0.28)	0.91 (0.28)	0.91 (0.29)	0.93 (0.26)
Rural	0.88 (0.32)	0.88 (0.32)	-	-
(All ages)				
Enrolled	0.70 (0.46)	0.71 (0.46)	0.71 (0.46)	0.70 (0.46)
Illiterate	0.28 (0.45)	0.28 (0.45)	0.28 (0.45)	0.30 (0.46)
Primary school	0.69 (0.46)	0.69 (0.46)	0.69 (0.46)	0.67 (0.47)
Junior high school and above	0.02 (0.15)	0.02 (0.15)	0.02 (0.15)	0.03 (0.17)
(Age>7)				
Enrolled	0.91 (0.28)	0.91 (0.28)	0.91 (0.29)	0.97 (0.16)
Illiterate	0.07 (0.43)	0.07 (0.26)	0.08 (0.27)	0.02 (0.16)
Primary school	0.89 (0.31)	0.89 (0.31)	0.89 (0.31)	0.92 (0.27)
Junior high school and above	0.04 (0.19)	0.04 (0.19)	0.03 (0.18)	0.05 (0.22)
Family (Observations)	447,159	442,423	376,680	70,479
Number of children	2.10 (0.90)	2.09 (0.89)	2.26 (0.87)	1.27 (0.57)
Having more than one child	0.75 (0.43)	0.75 (0.43)	0.85 (0.36)	0.23 (0.42)
Having a multiple birth	0.01 (0.10)	-	0.01 (0.10)	0.01 (0.09)
Mother’s age	31.61 (2.80)	31.61 (2.80)	31.45 (2.87)	32.47 (2.18)
Father’s age	34.23 (3.77)	34.23 (3.77)	34.13 (3.86)	34.78 (3.23)

Note: Standard deviations are shown in parentheses. All sampled children were aged at least 6 in 1990, with non-missing information on both mothers and fathers. Mother’s age is restricted to be no more than 35 at the census year.

Table 2: Descriptive Statistics of School Enrollment Rate by Family Size

	Family size					
	1-child	2-child	3-child	4-child	5-child	6+child
Full Sample	116,766	296,082	183,606	59,846	15,046	4,146
	Enrolled in or graduated from primary school (age<=13)					
All	0.68	0.73	0.68	0.64	0.61	0.57
Male	0.69	0.73	0.69	0.66	0.63	0.59
Female	0.64	0.72	0.66	0.63	0.60	0.56
	Enrolled in or graduated from junior high school (age>13)					
All	0.52	0.42	0.33	0.24	0.17	0.17
Male	0.51	0.43	0.36	0.28	0.22	0.21
Female	0.54	0.40	0.29	0.21	0.14	0.15
Rural Sample	61,784	277,474	179,236	58,579	14,639	4,017
	Enrolled in or graduated from primary school (age<=13)					
All	0.70	0.73	0.67	0.64	0.61	0.57
Male	0.72	0.73	0.69	0.66	0.63	0.59
Female	0.64	0.72	0.66	0.62	0.60	0.56
	Enrolled in or graduated from junior high school (age>13)					
All	0.42	0.37	0.31	0.22	0.16	0.15
Male	0.43	0.39	0.35	0.27	0.21	0.19
Female	0.40	0.34	0.27	0.19	0.13	0.13
Urban Sample	54,982	18,608	4,370	1,267	407	129
	Enrolled in or graduated from primary school (age<=13)					
All	0.65	0.78	0.77	0.72	0.67	0.59
Male	0.65	0.78	0.76	0.72	0.66	0.65
Female	0.65	0.77	0.77	0.72	0.67	0.55
	Enrolled in or graduated from junior high school (age>13)					
All	0.78	0.78	0.71	0.73	0.59	0.56
Male	0.75	0.76	0.68	0.73	0.54	0.56
Female	0.83	0.80	0.73	0.73	0.63	0.56

Note: Standard deviations are shown in parentheses. All sampled children were aged at least 6 in 1990, with non-missing information on both mothers and fathers. Mother's age is restricted to be no more than 35 at the census year.

Table 3: Ordered Probit Estimations of the Effect of Family Size on Children's Education Level: Full Sample

Dependent variable: Education level
(1=illiterate, 2=primary school, 3=junior high school or above)

	Children's controls only		Parents' demographic controls		Parents' and birth order controls	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of children	-0.150***		-0.136***		-0.126***	
	(49.41)		(44.15)		(36.74)	
2-child family		0.105***		0.132***		0.147***
		(17.55)		(21.29)		(23.03)
3-child family		-0.067***		-0.020***		0.008
		(9.31)		(2.70)		(0.99)
4-child family		-0.311***		-0.257***		-0.229***
		(28.69)		(23.43)		(19.58)
5-child family		-0.568***		-0.510***		-0.488***
		(26.45)		(24.25)		(22.13)
6+child family		-0.861***		-0.822***		-0.821***
		(18.26)		(17.90)		(17.16)
Second child					-0.009*	-0.049***
					(1.88)	(9.57)
Third child					-0.087***	-0.082***
					(9.13)	(8.58)
Fourth child					-0.089***	0.029
					(4.34)	(1.39)
Fifth child					-0.050	0.191***
					(0.87)	(3.19)
Sixth or later child					0.120	0.422**
					(0.74)	(2.51)
Male Indicator	0.127***	0.130***	0.131***	0.135***	0.135***	0.138***
	(32.72)	(33.38)	(33.34)	(34.16)	(33.97)	(34.69)
Han Indicator	0.293***	0.300***	0.228***	0.235***	0.228***	0.235***
	(29.04)	(29.85)	(23.08)	(23.86)	(23.13)	(23.90)
Rural Indicator	-0.023***	-0.147***	0.125***	0.012	0.124***	0.014*
	(3.54)	(21.19)	(16.52)	(1.48)	(16.30)	(1.77)
Observations	665,738	665,738	665,738	665,738	665,738	665,738

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. Households with twins are excluded. All regressions include age, age squared, and age cubed, and provincial dummies. Parents' controls include parents' age and age squared, and indicators for parents' education levels.

Table 4: Probit Estimations of the Effect of Family Size on Children's School Enrollment: Full Sample

Dependent variable: whether or not enrolled in school (1=yes, 0=no)

	Children's controls only		Parents' demographic controls		Parents' and birth order controls	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of children	-0.046***		-0.042***		-0.038***	
	(52.65)		(48.28)		(39.70)	
2-child family		0.022***		0.028***		0.032***
		(11.08)		(13.89)		(15.71)
3-child family		-0.039***		-0.026***		-0.017***
		(16.45)		(10.99)		(6.57)
4-child family		-0.116***		-0.098***		-0.085***
		(32.77)		(28.17)		(23.04)
5-child family		-0.187***		-0.166***		-0.149***
		(29.06)		(26.83)		(23.37)
6+child family		-0.255***		-0.238***		-0.220***
		(19.90)		(19.52)		(17.76)
Second child					-0.005***	-0.015***
					(3.38)	(9.23)
Third child					-0.034***	-0.031***
					(11.43)	(10.41)
Fourth child					-0.037***	-0.015**
					(5.78)	(2.40)
Fifth child					-0.027	0.001
					(1.48)	(0.07)
Sixth or later child					0.009	0.026
					(0.19)	(0.56)
Male Indicator	0.044***	0.044***	0.044***	0.044***	0.045***	0.046***
	(35.78)	(35.81)	(36.00)	(36.15)	(36.93)	(37.00)
Han Indicator	0.081***	0.084***	0.058***	0.060***	0.058***	0.060***
	(27.23)	(27.98)	(20.54)	(21.28)	(20.57)	(21.29)
Rural Indicator	0.017***	-0.015***	0.069***	0.036***	0.069***	0.037***
	(8.49)	(7.34)	(28.56)	(14.67)	(28.39)	(15.09)
Observations	665,738	665,738	665,738	665,738	665,738	665,738

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. Households with twins are excluded. All regressions include age, age squared, and age cubed, and provincial dummies. Parents' controls include parents' age and age squared, and indicators for parents' education levels.

Table 5: Ordered Probit Estimations of the Effect of Family Size on Children's Education Level: Rural Sample

Dependent variable: Education level (1=illiterate, 2=primary school, 3=junior high school or above)

	Children's controls only		Parents' demographic controls		Parents' and birth order controls	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of children	-0.152***		-0.143***		-0.126***	
	(49.02)		(45.36)		(36.21)	
2-child family		0.040***		0.059***		0.081***
		(6.09)		(8.65)		(11.57)
3-child family		-0.109***		-0.074***		-0.032***
		(14.24)		(9.36)		(3.88)
4-child family		-0.337***		-0.299***		-0.251***
		(30.51)		(26.69)		(21.05)
5-child family		-0.576***		-0.537***		-0.486***
		(27.07)		(25.66)		(22.23)
6+child family		-0.838***		-0.819***		-0.782***
		(18.01)		(18.16)		(16.65)
Second child					-0.050***	-0.074***
					(9.81)	(14.41)
Third child					-0.130***	-0.130***
					(13.65)	(13.64)
Fourth child					-0.129***	-0.035*
					(6.26)	(1.67)
Fifth child					-0.093	0.097
					(1.61)	(1.64)
Sixth or later child					0.134	0.365**
					(0.82)	(2.17)
Male Indicator	0.144***	0.147***	0.148***	0.152***	0.154***	0.157***
	(34.88)	(35.62)	(35.23)	(36.11)	(36.43)	(37.23)
Han Indicator	0.312***	0.314***	0.247***	0.248***	0.247***	0.248***
	(29.51)	(29.75)	(23.79)	(23.99)	(23.81)	(24.02)
Observations	587,058	587,058	587,058	587,058	587,058	587,058

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. Households with twins are excluded. All regressions include age, age squared, and age cubed, and provincial dummies. Parents' controls include parents' age and age squared, and indicators for parents' education levels.

Table 6: Probit Estimations of the Effect of Family Size on Children's School Enrollment: Rural Sample

Dependent variable: whether or not enrolled in school (1=yes, 0=no)

	Children's controls only		Parents' demographic controls		Parents' and birth order controls	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of children	-0.046***		-0.043***		-0.038***	
	(50.72)		(47.87)		(38.26)	
2-child family		0.009***		0.013***		0.019***
		(3.85)		(5.49)		(7.99)
3-child family		-0.046***		-0.037***		-0.024***
		(17.75)		(14.00)		(8.74)
4-child family		-0.119***		-0.107***		-0.088***
		(32.22)		(29.06)		(22.96)
5-child family		-0.186***		-0.172***		-0.147***
		(28.41)		(27.05)		(22.61)
6+child family		-0.247***		-0.237***		-0.210***
		(19.04)		(19.16)		(16.72)
Second child					-0.015***	-0.021***
					(9.26)	(12.80)
Third child					-0.044***	-0.042***
					(14.53)	(13.88)
Fourth child					-0.047***	-0.030***
					(7.20)	(4.57)
Fifth child					-0.038**	-0.019
					(2.10)	(1.05)
Sixth or later child					0.011	0.017
					(0.23)	(0.35)
Male Indicator	0.049***	0.050***	0.049***	0.050***	0.051***	0.052***
	(37.75)	(37.90)	(37.76)	(38.00)	(39.10)	(39.27)
Han Indicator	0.087***	0.088***	0.063***	0.064***	0.063***	0.064***
	(27.56)	(27.81)	(21.11)	(21.35)	(21.09)	(21.32)
Observations	587,058	587,058	587,057	587,057	587,057	587,057

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. Households with twins are excluded. All regressions include age, age squared, and age cubed, and provincial dummies. Parents' controls include parents' age and age squared, and indicators for parents' education levels.

Table 7: Ordered Probit Estimations of the Effect of Family Size on Children's Education Level: Urban Sample

	Dependent variable: Education level (1=illiterate, 2=primary school, 3=junior high school or above)					
	Children's controls only		Parents' demographic controls		Parents' and birth order controls	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of children	-0.091***		-0.030*		-0.060***	
	(5.82)		(1.95)		(2.99)	
2-child family		0.036*		0.099***		0.063***
		(1.89)		(4.98)		(2.85)
3-child family		-0.172***		-0.038		-0.108**
		(4.41)		(0.98)		(2.41)
4-child family		-0.209***		-0.030		-0.114
		(2.70)		(0.39)		(1.32)
5-child family		-0.814***		-0.591***		-0.708***
		(3.57)		(2.83)		(3.14)
6+child family		-1.598***		-1.261***		-1.428***
		(3.53)		(2.81)		(3.03)
Second child					0.160***	0.111***
					(5.25)	(3.75)
Third child					0.072	0.135
					(0.90)	(1.63)
Fourth child					-0.098	0.110
					(0.66)	(0.69)
Fifth child					0.209	0.907**
					(0.53)	(2.13)
Sixth or later child					-0.877***	0.096
					(3.60)	(0.20)
Male Indicator	-0.038***	-0.036***	-0.034***	-0.032**	-0.037***	-0.035***
	(2.97)	(2.80)	(2.67)	(2.47)	(2.89)	(2.71)
Han Indicator	-0.034	-0.034	-0.032	-0.032	-0.037	-0.037
	(1.03)	(1.04)	(0.98)	(0.98)	(1.15)	(1.13)
Observations	78,680	78,680	78,680	78,680	78,680	78,680

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. Households with twins are excluded. All regressions include age, age squared, and age cubed, and provincial dummies. Parents' controls include parents' age and age squared, and indicators for parents' education levels.

Table 8: Probit Estimations of the Effect of Family Size on Children's School Enrollment: Urban Sample

	Dependent variable: whether or not enrolled in school (1=yes, 0=no)					
	Children's controls only		Parents' demographic controls		Parents' and birth order controls	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of children	-0.019***		-0.007*		-0.018***	
	(4.88)		(1.69)		(3.68)	
2-child family		0.016***		0.027***		0.014**
		(3.10)		(5.13)		(2.30)
3-child family		-0.039***		-0.008		-0.037***
		(3.24)		(0.73)		(2.72)
4-child family		-0.074***		-0.029		-0.069***
		(3.04)		(1.27)		(2.60)
5-child family		-0.289***		-0.206***		-0.276***
		(5.35)		(4.44)		(5.39)
6+child family		-0.492***		-0.373***		-0.465***
		(5.67)		(4.71)		(5.57)
Second child					0.051***	0.038***
					(6.62)	(4.81)
Third child					0.027	0.045**
					(1.49)	(2.46)
Fourth child					-0.009	0.051
					(0.24)	(1.52)
Fifth child					0.065	0.132**
					(0.85)	(2.52)
Sixth or later child					-0.228***	0.014
					(3.07)	(0.21)
Male Indicator	-0.004	-0.004	-0.004	-0.003	-0.005	-0.004
	(1.30)	(1.09)	(1.13)	(0.92)	(1.43)	(1.25)
Han Indicator	-0.007	-0.006	-0.006	-0.005	-0.008	-0.007
	(0.82)	(0.72)	(0.75)	(0.64)	(0.96)	(0.84)
Observations	78,680	78,680	78,680	78,680	78,680	78,680

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. Households with twins are excluded. All regressions include age, age squared, and age cubed, and provincial dummies. Parents' controls include parents' age and age squared, and indicators for parents' education levels.

Table 9: OLS and 2SLS Estimations of the Effect of Family Size on Children's Education Outcomes: Full Sample

	All families (IV: twins at 1st birth)		Families with at least 2 births (IV: twins at 2nd birth)		Families with at least 3 births (IV: twins at 3rd birth)	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)
	Dependent variable: Education level (1=illiterate, 2=primary school, 3=junior high school or above)					
First Stage	-	0.556***	-	0.696***	-	0.824***
	-	(42.44)	-	(57.57)	-	(51.43)
Number of children	-0.028***	-0.040***	-0.038***	-0.011	-0.044***	-0.029**
	(42.42)	(2.84)	(47.88)	(1.08)	(29.07)	(2.11)
Male Indicator	0.027***	0.024***	0.032***	0.039***	0.048***	0.050***
	(31.09)	(5.96)	(33.00)	(14.59)	(31.48)	(18.98)
Han Indicator	0.049***	0.047***	0.055***	0.059***	0.067***	0.069***
	(22.63)	(12.06)	(23.73)	(21.17)	(19.98)	(19.28)
Rural Indicator	0.028***	0.038***	-0.041***	-0.051***	-0.085***	-0.086***
	(16.80)	(3.39)	(15.62)	(11.67)	(14.49)	(14.05)
	Dependent variable: whether or not enrolled in school (1=yes, 0=no)					
First Stage	-	0.556***	-	0.696***	-	0.824***
	-	(42.44)	-	(57.57)	-	(51.43)
Number of children	-0.027***	-0.030**	-0.036***	-0.009	-0.040***	-0.027**
	(43.27)	(2.20)	(47.11)	(0.91)	(27.86)	(2.02)
Male Indicator	0.029***	0.028***	0.034***	0.040***	0.051***	0.053***
	(34.47)	(7.13)	(36.63)	(15.66)	(34.73)	(20.60)
Han Indicator	0.045***	0.044***	0.050***	0.054***	0.060***	0.061***
	(21.52)	(11.83)	(22.30)	(19.99)	(18.43)	(17.81)
Rural Indicator	0.040***	0.042***	-0.011***	-0.020***	-0.038***	-0.040***
	(26.47)	(3.93)	(5.25)	(5.14)	(8.42)	(8.16)
Observation	675,492	675,492	555,059	555,059	256,909	256,909

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. All regressions include age, age squared, and age cubed, parent's age and age squared, indicators for parents' education level, and provincial dummies.

Table 10: OLS and 2SLS Estimations of the Effect of Family Size on Children's Education Outcomes: Rural Sample

	All families (IV: twins at 1st birth)		Families with at least 2 births (IV: twins at 2nd birth)		Families with at least 3 births (IV: twins at 3rd birth)	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)
	Dependent variable: Education level (1=illiterate, 2=primary school, 3=junior high school or above)					
First Stage	-	0.505***	-	0.688***	-	0.826***
	-	(33.62)	-	(54.89)	-	(50.64)
Number of children	-0.030***	-0.042**	-0.038***	-0.012	-0.044***	-0.028**
	(43.54)	(2.46)	(47.04)	(1.18)	(28.93)	(1.99)
Male Indicator	0.031***	0.027***	0.034***	0.040***	0.050***	0.052***
	(33.14)	(5.09)	(34.02)	(14.22)	(32.02)	(19.52)
Han Indicator	0.055***	0.053***	0.056***	0.060***	0.068***	0.069***
	(23.48)	(13.39)	(23.35)	(20.95)	(19.97)	(19.33)
	Dependent variable: whether or not enrolled in school (1=yes, 0=no)					
First Stage	-	0.505***	-	0.688***	-	0.826***
	-	(33.62)	-	(54.89)	-	(50.64)
Number of children	-0.029***	-0.032*	-0.036***	-0.011	-0.040***	-0.023*
	(43.11)	(1.88)	(46.30)	(1.03)	(27.52)	(1.68)
Male Indicator	0.032***	0.031***	0.035***	0.041***	0.052***	0.054***
	(36.26)	(6.00)	(37.02)	(14.99)	(34.91)	(20.91)
Han Indicator	0.050***	0.049***	0.051***	0.055***	0.061***	0.063***
	(22.11)	(12.88)	(21.95)	(19.76)	(18.44)	(17.91)
Observation	595,729	595,729	531,038	531,038	251,050	251,050

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. All regressions include age, age squared, and age cubed, parent's age and age squared, indicators for parents' education levels, and provincial dummies.

Table 11: OLS and 2SLS Estimations of the Effect of Family Size on Children's Education Outcomes: Urban Sample

	All families (IV: twins at 1st birth)		Families with at least 2 births (IV: twins at 2nd birth)		Families with at least 3 births (IV: twins at 3rd birth)	
	OLS (1)	2SLS (2)	OLS (3)	2SLS (4)	OLS (5)	2SLS (6)
	Dependent variable: Education level (1=illiterate, 2=primary school, 3=junior high school or above)					
First Stage	-	0.790***	-	0.846***	-	0.702***
	-	(38.15)	-	(21.01)	-	(8.13)
Number of children	-0.003	-0.024	-0.022***	-0.008	-0.024**	-0.045
	(1.46)	(1.13)	(4.95)	(0.26)	(2.34)	(0.47)
Male Indicator	-0.006***	-0.008***	-0.011**	-0.009*	-0.012	-0.014
	(2.64)	(2.66)	(2.46)	(1.78)	(1.18)	(1.01)
Han Indicator	-0.008	-0.019	-0.001	0.002	0.006	0.004
	(1.46)	(1.51)	(0.10)	(0.23)	(0.27)	(0.15)
	Dependent variable: whether or not enrolled in school (1=yes, 0=no)					
First Stage	-	0.790***	-	0.846***	-	0.702***
	-	(38.15)	-	(21.01)	-	(8.13)
Number of children	-0.002	-0.021	-0.017***	0.006	-0.025***	-0.146*
	(0.88)	(1.06)	(4.69)	(0.22)	(3.44)	(1.79)
Male Indicator	-0.002	-0.004	0.002	0.004	0.008	-0.005
	(0.94)	(1.36)	(0.55)	(0.94)	(0.98)	(0.41)
Han Indicator	-0.004	-0.015	0.004	0.009	-0.001	-0.014
	(0.87)	(1.26)	(0.65)	(1.08)	(0.08)	(0.66)
Observation	79,763	79,763	24,021	24,021	5,859	5,859

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. All regressions include age, age squared, and age cubed, parent's age and age squared, indicators for parents' education levels, and provincial dummies.

Table 12: OLS and 2SLS Estimations of the Effect of Family Size on Children's Educational Outcomes by Gender and Mother's Education: Rural Sample

		Birth order	Gender		Mother's education		
			Male	Female	Illiterate	Primary	>Primary
		(1)	(2)	(3)	(4)	(5)	(6)
		Dependent variable: Education level (1=illiterate, 2=primary school, 3=junior high school or above)					
Twins at 1st birth	OLS	-0.025***	-0.024***	-0.035***	-0.051***	-0.021***	-0.016***
		(33.81)	(26.27)	(37.77)	(37.94)	(22.91)	(11.85)
	2SLS	-0.042**	-0.022	-0.061***	0.023	-0.018	-0.140***
		(2.56)	(0.95)	(2.78)	(0.68)	(0.72)	(4.24)
Twins at 2nd birth	OLS	-0.033***	-0.033***	-0.041***	-0.057***	-0.026***	-0.024***
		(38.27)	(29.21)	(39.42)	(38.20)	(24.20)	(14.13)
	2SLS	-0.015	-0.025*	-0.001	-0.029	-0.009	-0.004
		(1.50)	(1.84)	(0.07)	(1.08)	(0.65)	(0.23)
Twins at 3rd birth	OLS	-0.040***	-0.038***	-0.046***	-0.061***	-0.029***	-0.024***
		(24.59)	(16.96)	(25.10)	(23.58)	(14.58)	(6.74)
	2SLS	-0.027**	-0.046**	-0.015	-0.061**	-0.011	-0.007
		(1.99)	(2.33)	(0.86)	(2.09)	(0.55)	(0.29)
		Dependent variable: whether or not enrolled in school (1=yes, 0=no)					
Twins at 1st birth	OLS	-0.024***	-0.022***	-0.034***	-0.049***	-0.019***	-0.015***
		(33.92)	(25.52)	(37.71)	(38.07)	(21.95)	(11.44)
	2SLS	-0.030*	-0.015	-0.048**	0.044	-0.013	-0.134***
		(1.89)	(0.67)	(2.23)	(1.32)	(0.53)	(4.08)
Twins at 2nd birth	OLS	-0.031***	-0.029***	-0.040***	-0.054***	-0.024***	-0.022***
		(38.03)	(27.79)	(39.47)	(38.14)	(22.96)	(13.87)
	2SLS	-0.013	-0.019	-0.004	-0.030	-0.003	-0.008
		(1.37)	(1.44)	(0.26)	(1.14)	(0.23)	(0.45)
Twins at 3rd birth	OLS	-0.036***	-0.032***	-0.044***	-0.057***	-0.025***	-0.022***
		(23.58)	(15.32)	(24.35)	(22.97)	(12.77)	(6.89)
	2SLS	-0.022*	-0.037**	-0.013	-0.058**	-0.008	-0.0001
		(1.68)	(1.97)	(0.76)	(1.97)	(0.43)	(0.01)
Observation		595,729	307,542	288,187	173,157	277,027	145,545

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. All regressions include age, age squared, and age cubed, parent's age and age squared, indicators for parents' education levels, and provincial dummies.

Table 13: OLS and 2SLS Estimations of the Effect of Family Size on Children's Educational Outcomes by Gender and Mother's Education: Urban Sample

		Birth order	Gender		Mother's education		
			Male	Female	<Junior	Junior high	>Junior
		(1)	(2)	(3)	(4)	(5)	(6)
		Dependent variable: Education level (1=illiterate, 2=primary school, 3=junior high school or above)					
Twins at 1st birth	OLS	-0.004	-0.0003	-0.006**	-0.018***	-0.001	0.013***
		(1.52)	(0.08)	(2.08)	(4.15)	(0.21)	(2.86)
	2SLS	-0.021	-0.018	-0.027	-0.078	-0.038	0.008
		(1.06)	(0.65)	(1.03)	(0.98)	(1.09)	(0.30)
Twins at 2nd birth	OLS	-0.019***	-0.019***	-0.026***	-0.026***	-0.018***	-0.008
		(3.54)	(2.87)	(4.63)	(3.89)	(2.68)	(0.73)
	2SLS	-0.009	0.022	-0.027	-0.040	0.035	0.020
		(0.29)	(0.45)	(0.71)	(0.78)	(0.59)	(0.39)
Twins at 3rd birth	OLS	-0.017	-0.023	-0.027**	-0.024*	-0.020	-0.021
		(1.60)	(1.58)	(2.29)	(1.73)	(1.31)	(0.74)
	2SLS	-0.045	-0.060	-0.027	-0.113	0.114	-0.162
		(0.49)	(0.51)	(0.21)	(0.65)	(1.08)	(0.82)
		Dependent variable: whether or not enrolled in school (1=yes, 0=no)					
Twins at 1st birth	OLS	-0.004*	0.001	-0.004	-0.015***	0.004	0.013***
		(1.72)	(0.42)	(1.44)	(4.35)	(1.19)	(3.18)
	2SLS	-0.018	-0.020	-0.022	-0.065	-0.028	0.001
		(0.97)	(0.76)	(0.87)	(0.84)	(0.82)	(0.02)
Twins at 2nd birth	OLS	-0.014***	-0.012**	-0.020***	-0.020***	-0.007	-0.008
		(3.56)	(2.51)	(4.24)	(3.85)	(1.46)	(0.87)
	2SLS	0.004	0.034	-0.014	0.034	0.045	-0.024
		(0.17)	(0.85)	(0.45)	(0.69)	(1.22)	(0.52)
Twins at 3rd birth	OLS	-0.021***	-0.022**	-0.029***	-0.026**	-0.014	-0.033*
		(2.69)	(2.13)	(2.96)	(2.44)	(1.32)	(1.90)
	2SLS	-0.140*	-0.153*	-0.134	-0.276*	0.003	-0.156
		(1.82)	(1.81)	(1.17)	(1.70)	(0.09)	(0.81)
Observation		79,763	41,595	38,168	14,009	33,004	32,750

Note: *, **, and *** represent significance levels of 10, 5, and 1 percent. Robust t-statistics, which allow for correlation of errors within family, are shown in parentheses. All regressions include age, age squared, and age cubed, parent's age and age squared, indicators for parents' education levels, and provincial dummies.