Is the quantity-quality trade-off real? Quasi-experimental evidence from China

Yiyang Luo[*]

[*]PhD student, University of Essex, UK; Research fellow, University of Namur, Belgium

Abstract

China has implemented several family planning policies during the last four decades. For the purpose of understanding the implication and consequences of controlling population size, this paper investigates the relationship between fertility and children's educational outcome(quantity-quality(Q-Q) trade-off). It departs from previous literature in accounting for the non-linear distribution both within each birth parities and across birth parities. Specifically, the non-normal distribution within each birth parity is accounted for by applying Generalized Method of Moments(GMM). It respects the limited dependent nature of both outcome and endogenous variables. In contrast with conventional 2SLS estimates, GMM estimates give significantly negative effects of fertility on education outcome of children, which is consistent with the Becker-Lewis model that an increase in child quantity increases the shadow price of quality. Moreover, the Q-Q trade-off non-linearly decreases with family size and shows heterogeneous effects by birth order. The identification strategy exploits variation in family size that is induced by twin births and first child's gender. This paper is the first to apply first child's gender as an instrumental variable to explore the Q-Q trade-off in the context of China where son preferences are prevailing. This instrument offers the possibility to test the Q-Q trade-off at a lower fertility margin than the typical instruments based on twins and same-sex composition.

JEL Classification: C26, J13, J18, J24, O15, P20

Keywords: Fertility, education, quality-quantity trade-off, GMM, twinning, first child's gender

1 Introduction

Under the general assumption that parents invest in all their children equally, Becker and Nigel (1973) propose that an exogenous increase in the number of children increases the marginal cost of investment in their quality. Motivated by the theoretical model of the quantity-quality(Q-Q) trade-off (Becker and Nigel, 1973, 1976; Willis, 1973), the relationship between family size and child outcome has been one of the most enduring in economics. On the one hand, the declining quantity of children will free up resources for human capital investment; on the other hand, larger family size may benefit from the economies of scale and the social interaction of children, as Black et al. (2005) argued, children with fewer siblings may not be better off than if they are in a larger family.

If the Q-Q trade-off exists, it suggests that parents who have less children allocate more of their time and resources to each of the children assuming that parents invest in children evenly and holding income constant. Furthermore, the trade-off implies that those policies aiming at curbing population growth or subsidizing families with fewer children may contribute to better health, educational and behavioural outcomes of each child. In turn, it may reduce abortion rates, crime rates and poverty rates in the society (Schultz, 2007). For example, if couples perceive this Q-Q trade-off story and prefer better quality children, parents could be motivated to avoid unwanted births and reduce social problems. Therefore, a significant trade-off provides evidence for government to advance contraceptive knowledge and subsidize birth control.

However, population control policy is unsustainable for long term economic growth, especially for ageing societies such as some western countries and China. China implemented One-Child Policy(OCP) in 1978 to curb the rapid population growth, but it realized recently that the OCP gave substantial pressure for the only child to give elderly care for parents and for the society to bear the cost of ageing population structure. Thus the OCP was relaxed to some extent in 2014(selective two-child policy), and two-child policy has been implemented since early 2016. These shifts in family planning policy induces complicated demographic transition, thus it would be important to understand the relationship between fertility and human capital accumulation. If the Q-Q trade-off exists, policies aiming at maintaining sustainable economic growth should be equipped with policies that alleviate the adverse effect of fertility, such as more generous maternal leave, public education and health support for the family. It is also important to test the existence of a cutoff point for Q-Q trade-off. If non monotonicity and heterogeneity exist in quantity effects, then a trade-off may be found in a certain range of

family size and households with certain characteristics. Thus policies should be adjusted accordingly.

While the theoretical model implies a negative trade-off, empirical results of the Q-Q trade-off need to be treated critically. It is well established that there is no Q-Q trade-off in developed countries, but developing societies tend to have less conclusive findings. Due to the higher quality of public education system and better government support for childbearing and childcare, it is likely that the Q-Q trade-off is less evident in more developed economies. This study focuses on children born in 1970s with a majority of them living in rural China, they suffered from poor education system and were born before[1] the implementation of One Child Policy.

The compulsory schooling law in China was not implemented until 1986. It mandates six-year primary school and three-year junior high school. The primary school enrolment rate was quite low at the beginning of Chinese economic reform in 1978. Only 36.6 percent of the counties has widespread access to primary school education for schooling age children in 1985, and the enrolment rate of primary school and junior high school was 95.9 percent and 68.4 percent, respectively[2]. The implementation of compulsory schooling was not uniform across the country. It largely depended on local government budget which made the access to education more difficult in poor rural areas. In poor areas, even if public school was provided, it was not totally free, parents still had to pay the tuition and other fees[3]. Thus the school enrolment of children were subject to the choices of parents and local government budget and facilities prior to 1986. The sample under study in this paper is composed of children born in a period which is before the implementation of compulsory schooling law.

This paper presents new evidence on the child quantity(sibling size)-quality(schooling) trade-off in the context of China. The identification strategy exploits variation in family size that is induced by the gender of first birth and twin births. Typically previous literature tends to estimate the trade-off using linear probability approach and only considers non-linearity across birth parities[4]. This paper departs from previous literature in accounting for the non-linear distribution both

[1]The sample also includes children born in early 1980s, which may result in the estimates being contaminated by the OCP. However, robustness check based on an earlier period sample(1982 census) give similar result.

[2]source from the website of the National People's Congress of the People's Republic of China `http://www.npc.gov.cn/npc/xinwen/rdlt/fzjs/2009-02/13/content_1470214.htm`

[3]It is not until 2000 that the coverage of nine-year compulsory schooling reached 85% and the illiteracy rate reduced to less than 5%.

[4]Here birth parities means at least 1 birth sample(families with at least 1 birth), at least 2 births sample(families with at least 2 births) and at least 3 births sample(families with at least 3 births).

within each birth parities and across birth parities. Specifically, the non-normal distribution within each birth parity is accounted for by applying Generalized Method of Moments(GMM). It respects the binary nature of the outcome variables and the count nature of the endogenous variable, thus it gives a consistent and more efficient estimator of the Q-Q trade-off. The non-linear distribution of fertility effects across birth parities is accounted for by estimating Q-Q trade-off at each birth parity separately(Black et al., 2005; Angrist et al., 2010).

This paper is the first to apply the variation in fertility generated by first child's gender into testing the Q-Q trade-off in the context of China where son preferences are prevailing. This is largely influenced by the Confucianism that it stresses the importance of patrilineality within Chinese society (Das Gupta et al., 2003). Patrilineality enforces sons to continue the family lineage and inheritance. It is also widely believed that girls marry "out" of their natal families and parents cannot rely on girls for old-age support. Therefore, families with son preferences will generally have larger family size if the first birth is a girl than a boy. This instrument uncovers a channel to test the Q-Q trade-off of a marginal second birth on the first born, while typically twins and parental preference for a mixed gender composition estimate the effect of a marginal third or higher birth order child on older siblings[5]. Moreover, using it as an instrumental variable provides a sensitivity test on twinning as an instrument.

I find significantly adverse effects of one additional birth on the education outcome of the firstborn child for families with at least one or two births, regardless of the instruments. The Q-Q trade-off is estimated separately for firstborn and secondborn children of families with three or more children, as it is reasonable to believe that children of different birth order would present different response to an additional birth. Indeed, Q-Q trade-off is bigger for the second-born children than for the first-born children. Generally the Q-Q trade-off is more evident in a relatively smaller family. It maybe due to the fact that conditional on a bigger family size, the resource is spread more thinly and thus the marginal decrease in the welfare of older sibling declines with child quantity. It provides some evidence on the non-linearity of trade-off as family size goes up.

Although exploring potential mechanisms through which family size affects children's outcome is beyond the scope of this paper, I explored mother's labour supply adjustment and economies of scale channels. Basically, mother's labour supply stayed stable when family size increased from 2 to 3 but dropped signif-

[5]In this paper marginal child refers to the child that is born due to the instrument, for example, a marginal third birth refers to the additional child induced by twins at the second birth.

icantly when fertility raised from 3 to 4. It has been documented that maternal time is very important in the early stage of the children's development. If reduction in labour supply indicates an increase in maternal supervision time with their children, then this labour supply adjustment could possibly explain the finding that 2+ families showed a significant Q-Q trade-off while 3+ families displayed no Q-Q trade-off. The evidence on economies of scale in terms of sharing schooling costs was mixed, but in general it favoured the channel that children from a bigger family enjoyed economies of scale[6].

Several tests are performed to test the validity of instrumental variables since twin births and girl at first birth may affect the schooling outcome independently from family size. Twinning may affect the outcome of lower parity children through zero spacing and lower birthweight. I test the existence of these mechanisms, and the results suggest that estimators of Q-Q effects are probably positively biased. Thus the true adverse effects of family size on the schooling performance of the children are bigger.

The validity of firstborn gender implies that it should be random so that it is not correlated to any parental preference. Thus it requires assumptions to be made on parental investment in education, i.e. parents have no gender-biased investment preference and child's ability is independent of gender. This paper tests this hypothesis in the following ways.

First, if the sex ratio at the first birth is within the natural range[7], then it can at least rule out pre-natal sex preferences. One would worry about the technology advancement in invalidating the gender-related instruments. Ultrasound diagnosis of fetal gender became available since early 1980s, but sex-selective abortion techniques became widely available only after 1986/7 (Lin et al., 2008). My sample period covers children born between 1974 and 1984, among which the younger cohorts could be affected by theses technologies but only to a weak extent. A simple way to check sex-selective abortion on the firstborn is to compare mother's age at first birth with boy versus girl. The idea is that if abortion exists, it should delay son's birth. However, I find no significant difference between mother's age at firstborn boy and at firstborn girl. Additionally, the infant mortality rate was 70 deaths per 1000 births, among which boy death was higher than girl death. Furthermore, to rule out the contamination of technology effect, results are also provided using 1982 census which is consists of children born before the availabil-

[6]Results on these mechanisms are not attached in this version of paper, but they can be provided upon request.

[7]The World Health Organization(WHO) defines natural sex ratio at birth to be 105 males for every 100 females. More information at `http://www.searo.who.int/entity/health_situation_trends/data/chi/sex-ratio/en/`

ity of these technologies. Second, possibly the validity can be tested by comparing the educational outcomes of children for different gender holding family size constant. If families with more boys have better outcomes than those with more girls, conditional on family size, then it indicates a post-natal sex-selective investment, assuming child ability is independent of gender. This paper find no evidence of post-natal gender biased investment behaviour from this perspective. Third, if gender at first birth does not affect child outcome independently from family size, then one would expect no reduced-form effects in a sample where the first child gender should matter the least. Unfortunately, this validity test fails to rule out the direct impact of first child's gender, but the direction of bias indicates that the true Q-Q tradeoff is smaller.

Previous twins-based studies mainly examine the effects of family size on lower birth order children, the effects on outcome of twins themselves are often overlooked. Rosenzweig and Zhang (2009) suggest that the effect of twinning on twins and non-twins provides upper and lower bound estimation of Q-Q trade-off, under the assumption that parents allocate resources towards non-twins due to the birthweight deficit. Therefore, this study also employs twin at first birth as instrument to bound the traditional twin estimators. Applying their estimation strategy, I find support of parental reinforcement behaviour, implying estimators of lower bound are positively biased. Thus, the true Q-Q trade-off based on twin instrument is bigger than what is showed in this study.

There is evidence of heterogeneity in supporting that the trade-off is mainly driven by children of rural households and of less educated mothers. Possible explanation could be that children from relatively disadvantaged background experienced less developed education system and government support, and were more likely to be financially constrained due to the less developed capital market.

The paper unfolds as follows: section 2 presents the institutional background and literature review on the analysis of quantity-quality trade-off. Section 3 provides the estimation strategy. Section 4 describes the data source and descriptive statistics. Section 5 provides empirical results and robustness checks. Finally, section 6 draws conclusion.

2 Institutional Background and Literature Review

2.1 Policy background

After two decades of rapid population growth, the Chinese central government enacted a series of birth planning campaign to curb population growth in 1971. At first, the main objective was to encourage people to get married later, to expand birth spacing between births and to have fewer children. Later, a stricter policy – the One Child Policy – was announced in 1978 and tightly enforced across the country in 1980. Each household was allowed to have only one child, a second child was permitted only under extreme circumstances. However, it only applied to individuals of Han ethnicity, ethnic minorities were exempt from this policy. It generates additional variation in family size across ethnicities.

Considering the difficulties in enforcing the policy and the high labour demand in rural areas, this restriction was relaxed in the manner that rural couples were allowed to have a second birth after a specified interval if the first child was a girl in 1984. However, in some rural areas, all couples were allowed to have two children. Violators of the family planning policy will mainly be punished in the form of monetary penalties. The amount of fines are generally big and varies with regions. Under certain circumstance, people can even lose their jobs or permanently lose the chance to get promoted, for instance if one of the couple works in the government or state-owned enterprise.

There has been a gradual relaxation of China's family planning policies in recent years. The selective Two-Child policy was implemented in 2014. Couples in many parts of the country have been allowed to have two children if one parent was an only child. However, eligible couples did not illustrate strong wishes of giving additional birth. Thus this policy was far from reaching the government goal. In early 2016, Two-Child policy has been enacted to allow for maximum 2 children in all families.

2.2 Literature Review

A main implication of the quantity-quality trade-off model (Becker and Nigel, 1973, 1976) is that the marginal cost of quality increases with quantity, holding the income constant. However, empirical research, to some extent, diverges from the theory implication.

Rosenzweig and Wolpin (1980) are the first to confirm the hypothesis of Becker and Nigel (1973, 1976) by using twins as an exogenous increase in fertility in India. However, their estimation is imprecise due to the fact that the sample size is small(25 twins pairs), and the selection bias problem resulting from using outcomes of all children. Moreover, the internal validity of twinning as an instrumental variable relies on a large sample data which helps to mop up the infrequency of multiple births.

More recent studies based on developed societies tends to find negligible effects of family size on educational attainment of children. Using a data set on the entire population of Norway, Black et al. (2005) argue that previous work on the impacts of family size[8] would be biased if not controlling for birth order effects. In fact, they find negligible family size impacts on children's education when controlling for birth order. Instead of affecting the quality of each child, family size may only affect the marginal child through the impact of birth order. In particular, higher birth order has significant negative effects on child's education, adult earnings, employment and teenage childbearing. Theoretically, there are some models predicting birth order effects, say the optimal stopping model(continue to have children until they have a poor quality child). However, later research fails to find this pattern of birth order effects. One candidate explanation is that the Q-Q trade-off is offset by the good education system and generous welfare in well-off countries.

As pointed out by Imbens and Angrist (1994), the estimations generated by any particular IV approach capture only the effects on individuals affected by that instrument, which should be interpreted as the local average treatment effect (LATE). Angrist et al. (2010) combine evidence from multiple sources of variation in family size by separately and jointly using the incidence of twin births and same-sex sibling pairs. This method captures the effect of fertility across different treatment groups and therefore increases the range of variation. Their results show no evidence of Q-Q trade-off in Israel. However, their study fails to capture the effects on the marginal child, and the estimated effects on college attendance show large standard error. Additionally, a limitation of using twins at second/third birth and preference for a mixed gender composition of children is that they give only the marginal effects of a third/fourth child, while the marginal effects of a first or second child is technically inaccessible.

Developing countries differ from developed countries in many aspects, such as economy status, public education system, government support for childbearing

[8]This paper uses fertility effect, family size effect, sibling size effect and quantity effects interchangeably. They all refers to the effect of number of children.

and childcare, and social norms, which may render the trade-off more evident in the developing countries.

Lee (2008) studies the trade-off by exploiting variation in family size under son preferences in South Korea. He proposes an new instrumental variable – the first child's gender and directly tests the effect by using measurements of parental monetary investment in children's education, as he argues that the economics theory of Q-Q trade-off is about parental choices. He finds that per-child investment is reduced by 25.4% for children with one sibling, and by 42.4% for children with two siblings compare to the only child. The trade-off goes up as the family size is bigger. However, using parental monetary investment solely is not enough to capture the whole picture of Q-Q trade-off. For example, parental supervision time has been shown in economic and sociology literature that plays an important role for children's human capital accumulation, especially at the early stage.

Following the method of Black et al. (2005), Li et al. (2008) estimate the trade-off in China using 1990 Chinese Population Census, they find no effect of family size on first-born's education by exploiting the exogenous variation in family size induced by twins at second birth, and a significant trade-off on average educational outcome of first two births by using twins at third birth. They show that controlling for birth order do not alter the Q-Q trade-off effects, which contradicts the finding of Black et al. (2005). However, they treat an ordered discrete dependent variable using ordinary least square(OLS) estimation, which renders their results less convincing. Additionally, they do not estimate fertility effect for the first and second born children separately, in the at least three births family.

Rosenzweig and Zhang (2009) use a new data set, the Chinese Child Twins Survey(CCTS), to quantify the Q-Q trade-off of children. They also find a significantly negative relationship between family size and child quality. They exploit the effects by using the incidence of twin births that for the first time taking into account the effects of birth-weight deficit and close spacing of twins. Their study firstly proposes that the impacts of twinning on older non-twin births and on twin births themselves provide the lower and upper bounds of true impacts of the family size on average child quality. Moreover, the lower bound is positively biased if parents exhibit reinforcement behaviour. Thus true Q-Q trade-off should be bigger than what have been found in previous literature, if not accounting for variations in endowment. However, their strategy ignores the fact that birthweight itself is endogenous because it is affected by twinning. In addition, their approach is more likely to be a reduced form estimation by using twins, instead of using twins as an instrument to solve the endogeneity problem.

Based on different datasets in China, Li et al. (2008) find no evidence of Q-Q trade-off, but Rosenzweig and Zhang (2009) show negative effects of fertility rates on education outcomes of children by exploiting variation in family size using twins. Instead of the twins IV, China's family planning policy provides a good opportunity for researchers to investigate the quantity-quality trade-off. Using one child policy as an instrument provides a different LATE, Qian (2009) finds positive family size effects and Liu (2014) shows no significant effects on education outcome. The heterogeneity in the Q-Q trade-off suggests that it is sensitive to the sample period and composition.

Qian (2009) makes a breakthrough by using the relaxation of China's one child policy in rural China to instrument fertility. The sample size is substantially boosted by matching the 1989 CHNS data with the 1990 1% sample census at the county level. In particular, she exploits the exogenous variation by a triple interaction of the child's gender, year of birth and region of birth. Results show that an extra child increases school enrolment of first-born child by 16% for one-child family. She proposes several mechanisms for explaining the positive correlation and finds support in economies of scale in schooling cost and increasing mother labour supply for rising demand on cash for children's future education cost. However, the variation in the implementation of the One Child Policy is likely to correlate with the demand of children, thus the fertility effects derived from the policy are likely to be negatively biased.

In order to partly address the potential correlation problem, Liu (2014) uses two community level variables to control for the potential correlation between parental preference and local family planning polices. His study employs three instruments for fertility: fines for unsanctioned births, the eligibility for having two children, and their interactions. Results show weak fertility effects on children's educational attainment. He suggests that government interventions in education partly offset the effects of parental investments in children's schooling, resulting in a weak Q-Q trade-off.

Prior empirical studies that explore exogenous variations in family size due to twin births and sex composition mostly impose a constant fertility effects on child outcome, Mogstad and Wiswall (2016) re-examine this relationship by allowing an unrestricted functional form. They propose a non-parametric estimation that allows identification of marginal fertility effects, in comparison to the total fertility effects studied previously (Black et al., 2005). They find that previous conclusion of no Q-Q trade-off is an artifact of the linear specification in fertility, which masks significant marginal fertility effects. They conclude a Q-Q tradeoff in large families and Q-Q complementary in small families.

This paper departs from previous empirical work in accounting for the non-normal distribution of outcome and endogenous variables within each birth parity, rather than only across parities. In addition, it explores an additional dimension of marginal fertility effect.

3 Methodology

3.1 Instrumental Variables

Child quantity and quality are affected by parental preferences, household characteristics and child's innate ability. It is not possible to fully observe all the factors that affect child quality, hence child quality is endogenously determined by quantity. In order to identify a causal relationship, I use the gender of firstborn and twin births as instrumental variables to capture potentially exogenous variations in family size.

There are three main samples in my estimation, at least one birth, at least two births and at least three births sample. This analysis is limited to lower birth order (*N-1* births) child(ren) in at least *N* births sample as they are the only unconditional treatment units in the household. In specific, at least one birth sample(1+) consists of firstborns from families with one or more births, and the family size is instrumented by the gender of first birth. At least two births sample(2+) consists of firstborns from families with two or more births, and sibling size is instrumented by twins at second birth. At least three births sample(3+) consists of firstborns and secondborns from families with three or more births, fertility is instrumented by twinning at third birth. These three samples are further divided by outcome variables, as it will be explained in section 4, due to different exposed age groups.

This paper is the first one to apply the gender of first birth to test Q-Q trade-off in the context of China where traditionally son preferences is influenced by the Confuciusim. The argument for relevance of this instrument is that families with son preferences will generally have larger family size if the first birth is a girl than a boy. Moreover, given the biological time constraint for childbearing, families with a first born girl will shorten the birth spacing if they try to have a boy sooner. By the same logic, a girl at first birth is also a good predictor for a third birth. For example, if families have first two girl births and prefer to have a son, they will try to have an additional birth and should do so more quickly. In general, a girl at first birth is likely to contribute to a higher fertility rate and

shorter birth spacing. Lee (2008) proposes two relevant tests of first child's sex as an instrumental variable for fertility.

The first test estimates a parity progression model under the hypothesis that parents having daughters in lower birth orders should end up with larger family sizes if there were son preferences. This model requires that households' decisions about future childbearing are fully observed, so that it will be tested only among those who have completed fertility. The sample is confined to mothers aged above 40 as conventional demographic paper assumes that this is the age range that women are no longer fertile. I estimate a logit model where the main variable of interest is the sex composition of siblings and the dependent variable is an indicator for whether the family has more than N child/children birth. Mother's age at first birth, current age, birth year, education, ethnicity, geographic location of parents are controlled for. The first two columns of table 1 show that a girl at first birth increases the probability of having a second child by 5.5% and the likelihood to have a third child by 9.5%. The second column also shows that having same-sex children at first two births increases the probability to have a third child by 11.2% while the third column suggests that the increase in family size is mainly due to two girl births (9.4%). Having two sequential boy births does not affect the family size.

The other model for testing son preferences is a hazard model of fertility timing. The hypothesis in this model is that families should advance the timing of having a second child if the first birth is a girl than a boy given the time constraint of childbearing biologically. Similarly, the birth spacing between the second and third child should be shortened following two consecutive girl births. I estimate a Weibull hazard model where the key variables of interests are sex compositions of siblings and the outcome variable is the birth spacing between corresponding births. To avoid bias that caused by changes of time-variant variables during the interval, only time-invariant variables are included as controls. Right panel of table 1 reports relative hazard ratios. The fourth column shows that the hazard rate of having a second child increased 56% by a first girl birth. As before in the parity progression model, the positive effect of samesex sibling in inducing an additional birth is mainly resulted from two consecutive girl births(95%). Two consecutive son births decrease the hazard rate of having a third child by 26%. These results are consistent with the hypothesis that an initial girl birth(births) advances the timing of the second(third) child while boy births delay the additional childbearing.

The rational for twins instrument is that twin births at any parity is random and the subjects of interest are older siblings. The internal validity of twins instrument

relies on its randomness conditional on some observed biological factors. One concern would invalidate this instrument is the use of fertility drugs that boosts the chance of having multiple births. However, the introduction of this drug started in late 1980s, which is unlikely to affect my sample. In addition, several robustness checks are carried out in this study with regard to the closer spacing, lower birthweight, economies of scale issues of using twinning as an instrument.

Angrist et al. (2010) point out that combining evidence from multiple sources of variation is important since different instruments are potentially subject to different omitted variables biases. For example, the occurrence of twins varies with maternal age, and twin births may affect outcome directly through birth spacing. Although first child's sex instrument is not suffering from those biases, it may directly affect outcome through gender specific characteristics and postnatal investment preferences of parents. Therefore, estimations based on different instruments provide specification check on themselves.

Table 2 shows strong first stage effects. A girl at first birth increases fertility by about 28% to 30% for 1+ households, and twin births increase the family size by about 37% to 51% for 1+ sample, by 56% for 2+ sample, and by 42% to 55% for 3+ sample, depending on the exposed subsamples and birth orders. Figure 1 plots first stage estimates of the effect of twins and first child's sex on fertility. Twinning captures shifts in family size over a narrow range that is close to the parity of twins birth, while the first child's sex instrument captures a relatively wider range of fertility variation. For example, the effect of twins at second birth increases the fertility from 2 to 3 with no further effects on higher birth parities. While those families who have son preferences will tend to have children until obtaining a son, and significance of the effects fades out when family size go beyond 5.

3.2 Rationale for Using Non-linear Model

Previous work mainly uses conventional 2SLS to estimate the fertility effects, while typically the normal distribution assumptions for outcome and endogenous variables are violated. The conventional 2SLS approach can be characterized as follows (Black et al., 2005; Angrist et al., 2010):

$$N_i = Z_i\alpha + X_i\beta + e_i \tag{1}$$

$$Y_i = \gamma N_i + X_i\theta + v_i \tag{2}$$

where N_i is the number of children in the family(quantity), Z_i stands for instru-

ments, it's a dummy for the gender of the firstborn at 1+ sample, or twin births at parity N for at least N births sample, X_i is a set of control variables, Y_i is a measure of child's educational outcome(quality).

3.2.1 Graphic Evidence

Fertility, an endogenous variable in Q-Q trade-off literature, has traditionally been estimated using Ordinary Least Squares(OLS) with normal error terms. Since fertility is a discrete count variable, the normal distribution assumption is inappropriate. More importantly, the assumption of constant variance is invalid. Figure 2 explores the form of mean-variance association for households with at least one child. To make the evidence graphically representative, the moments are computed conditional on the education level of mother and mother's age at giving first birth. Clearly, the variance is not constant. Although the variance is not precisely equal to the mean, the plot suggests a proportional relationship between them. Thus it can do more justice to the data by using Poisson estimations than by relying on OLS (Rodriguez, 2007).

Recent literature has realized the non-linear fertility effects and anaylized the Q-Q trade-off for different family size, separately. However, they only consider the non-linearity across parities, but the non-normal distribution within parity is overlooked. Specifically, this paper argues for a GMM approach to account for the count nature of the endogeneous variable and binary nature of the outcome variables. This analysis shows that the use of this approach makes a significant difference in estimating the fertility effects. Following the visual assessment strategy proposed by Silva et al. (2014), figure 3 provides visual comparisons of the traditional approach and the proposed method on model fitness[9]. For ease of demonstration, equation 1 and 2 can be rewritten as the following,

$$N_i = V_i\sigma + e_i \tag{3}$$

$$Y_i = W_i\phi + v_i \tag{4}$$

where V_i includes all the covariates in the first stage of equation 1 and W_i includes all the covariates in the second stage of 2.

Figure 3 focus on families with at least one birth, examining the effects of an additional birth on the probability of completing primary school for the first-born children. The instrument used is the gender of first birth. Panel (a) and (b)

[9]Bear in mind that the true estimation process is not constructed in two stages, but showing the estimation step by step helps to provide clear rationale for using non-linear approach.

of figure 3 estimate the first stage and compare the fitness of OLS and Poisson regression. Plots show the value of N_i and parametric fits of $E[N_i|V_i]$ versus the estimated single index $V_i\hat{\sigma}$. In addition, nonparametric fits are displayed to provide a robustness check of the parametric fit, in the sense that parametric one proposes a model to fit the data while nonparametric one let the data commands the regression. Here the nonparametric fit uses a kernel regression of N_i on the fitted value obtained above. Panel (c) and (d) of figure 3 estimate the second stage and compare the fitness of OLS and Logit model. They show plots of Y_i and of parametric fits of $E[Y_i|W_i]$ versus the estimated single index $W_i\hat{\phi}$. Similarly nonparametric fits are added to plots.

All of the OLS estimators have the expected sign but the magnitude of marginal effects are clearly erroneous. For example, figure 3(c) shows that the fitted value of $E[Y_i|W_i]$ goes beyond zero and way below the upper bound of probability one. As a result, the parametric and non-parametric fits largely diverge from each other, which indicates a mismeasurement of partial effects for most of the observations. The departure of OLS fit from kernel fit may explain the failure of proving a Q-Q trade-off previously. In contrast, figure 3(d) displays an overlapping, to a large extent, of Logit and kernel regression, which shows a large improvement on the model fitting. The slight diverge of two regression fits at the upper tail is likely due to the fact that fewer children are enrolled at junior high school than those who are not.

Similar analysis is applied in figure 4 where the outcome variable is the probability of enrolling in Junior high school with twins at the second birth as an instrument variable[10]. Non-linear fitting shows better performance at the distribution boundaries. While the performance of non-linear fitting is not necessarily better than OLS within the boundaries, the non-linear estimation at least provides another bound of the Q-Q trade-off.

To sum up, the merit of using non-linear specification are manifold. First, it allows for non-normal distribution of the error term. Second, it relaxes the constant variance assumption imposed on Ordinary Least Square estimation. Third, it captures the discrete nature of the data, and allows inference to be analysed on the probability of event occurrence. Although coefficients can be estimated consistently by OLS, the inference based on the estimated covariance matrix is not valid.

Figure 3 and 4 have shown that non-linear fitting approximates data better than

[10]Here only two of the many regression are showed graphically, more plots are provided on request.

ordinary least squares at the distribution boundaries. This is of particular importance if the true value of outcome variable have higher probability locating closer to 0 or 1. The econometric estimation will be presented as follows.

3.2.2 Econometrics Reasoning

Equation 5 uses a Poisson regression to estimate first stage, but directly plugging in the first-stage fitted value to a Logit estimation—equation 6 would fall into the forbidden regression problem (Angrist and Pischke, 2008).

$$E(N_i|X_i, Z_i) = exp(Z_i\alpha + X_i\beta) \tag{5}$$

$$Y_i = F(\gamma N_i + X_i\theta) + \varepsilon_i \tag{6}$$

Estimation in stages, by replacing N with its estimated conditional mean, does not give consistent estimates of the parameters (Windmeijer and Santos Silva, 1997; Wooldridge, 2010).

$$Y_i = F(\gamma exp(Z_i\alpha + X_i\beta) + X_i\theta + \gamma\eta_i) + \varepsilon_i \tag{7}$$

where

$$\eta_i = N_i - exp(Z_i\alpha + X_i\beta) \tag{8}$$

Equation 9 shows that the moments of η_i depend on parameters and regressors.

$$E(\eta_i^2|X_i, Z_i) = exp(Z_i\alpha + X_i\beta) \tag{9}$$

Angrist and Pischke (2008) propose a simple alternative to the forbidden second step by using the nonlinear fitted values as instruments, instead of directly plugging in nonlinear fitted values. It has been shown graphically that nonlinear estimators give better approximations to the first stage conditional expectation function than the linear model, thus an additional advantage of non-linear modelling is the improvement in efficiency.

Furthermore, as suggested by Windmeijer and Santos Silva (1997) and Wooldridge (2010), a consistent estimator would be the Generalized Method of Moments(GMM) estimator, and the nonlinear fitted values are used as instruments. Following Windmeijer and Santos Silva (1997), a GMM approach with $exp(Z_i\hat{\alpha} + X_i\hat{\beta})$ as an instrument has been applied to this analysis. Using the fitted values, obtained from the Poisson regression, as instruments provide a best approximation to the

optimal instrument set. Figure 5 and 6 show the comparison of 2SLS and GMM estimator, clearly the advantage of GMM estimation exhibits at the boundaries. Thus the GMM estimation is appropriate for this study in the sense that outcome variables are more located at the upper tail of the distribution.

Additionally, number of children in a household are likely to display under-dispersion property, an ideal model to control for dispersion (Winkelmann and Zimmermann, 1994; Wang and Famoye, 1997) would be Generalized Poisson regression. However, this attempt to improve the instrument will have only a marginal impact on the final estimator, but will result in more of computational complication due to the additional parameters to be estimated. Thus, this paper uses Poisson regression to obtain the instrument.

4 Data

The dataset is the 1% sample of 1990 Chinese Population Census, it covers 11,835,947 individuals from 3,260,243 households. The 1990 census questionnaire contains detailed demographic information, including the following: name, relationship, sex, age, nationality, registration status, education, industry, occupation, unemployment status, marital status, residence in 1985, number of children ever born, number of children surviving, birth order in 1989 etc. The census provides no direct information on parental-children linkage, the relationship is thus inferred using the census question on what is each household member's relationship to the head and using the IPUMS matching rule[11].

For the purpose of this analysis, information on both parents and children are needed. However, the drawback of using census data is that it has no record on individual who left the family. Thereby this analysis requires households to have parents and children co-residence in the household[12], which drops 33% of the households. This reduction in sample size is largely due to elderly/very young families with no children presenting or families with no parents presenting. This analysis also requires observations of all alive children in the family since the test of Q-Q trade-off is built on resource competition or sharing. Thus it excludes families with reported number of children alive unequal to the constructed family size. There are 29% (690,924) of the households with reported number of children alive bigger than constructed family size. It suggests that those families having

[11]Information on the IPUMS algorithm is available at http://www.ipums.umn.edu

[12]Divorce was considered both immoral and bourgeois during 1950s to 1970s, and it required approval from the state, which made it basically impossible. In 1980s, a new marriage law made divorce relatively easier but the divorce rate remained very low, at less than 0.1%.

non-coresident children. Additionally 1% of the households have reported number of children smaller than the constructed family size, which could be attributed to measurement error or misreporting.

This study further restricts the sample by excluding mothers giving birth before age 16(0.2% of the sample); excluding families that migrated in the last 5 years[13]; excluding families with multiple birth (except twins birth) since they have closer spacing and endowment deficit compared to the singleton child(125 households are dropped). Finally, the sample size is 1,656,679 households.

This analysis focuses on three domains of outcome[14]: the probability of primary school enrolment, primary school graduation and junior high school enrolment. School enrolment is a binary variable that equals to one if a child was enrolled in school(attending) or dropped out or graduated, and zero if a child had never enrolled school. Graduation is defined as one if a child had graduated from school, zero if still attending or dropped out, conditional on enrolled in primary school. I use relatively early educational outcomes as measurements of child quality because the economic theory of the Q-Q trade-off analyses parental choices. It is more convincing to use outcomes that take place closer in time to parental choice-making on quantity and quality to test the theory (Lee, 2008). Therefore, college enrolment and labour market outcomes are not my outcomes of interest; they are more likely to be affected by various factors that are not correlated with parental choices.

The sample is restricted to children aged between 6 and 16 to ensure that all children have satisfied the minimum entry age for primary school and that there is no adult children living in the house since they are not traceable in the survey. The sample is further restricted to children age smaller than or equal to 13[15] for estimation of the effect on enrolment of primary school, and restricted to children age older than 13 for estimation of completing primary school and junior high school enrolment. Overall, this paper focuses on nine subsamples, which are categorized by family size and outcome variable.

[13]This is the only information on migration history that the 1990 census provided. China had very rigid hukou registration system that made migration very difficult. Before 1990s, the household mobility in China was almost zero. Indeed, 98% of the households reported to live in the same county and 99% reported to live in the same province during the last 5 years.

[14]This study focuses on public school which was the only type of primary and secondary school that existed in China during the sample period. The number of private school expanded in 2003 due to a change in the education law that admitted the legitimacy of non-government education.

[15]The years of primary school education varied between 5 and 6 years. Most of the children were expected to complete primary school by the age of 13, conditional on enrolled in primary school and no drop-out.

A set of family demographic characteristics are controlled for. In the 1+ sample, the age of the firstborn, mother's age at first birth, education and ethnic group of parents, an rural indicator and provincial indicators are included as controls. It is important to control for maternal age at first birth due to its correlation with twin births and investment behaviour. I also control for cohort effects by including the birth year of the firstborn child and father. It will help address the problem that younger children are more likely to be in school and have few siblings (Qian, 2009). Parental cohort effects are also included in the sense that parents of first-born children are likely to be different from parents of other parity born children. In addition, it controls for the year specific effects. For example, one would expect certain differences between parent who experienced the cultural revolution[16] and those who did not. In the 2+ sample, the gender of firstborn child and birth spacing between first two births are additionally controlled for. In the 3+ sample, I also control for the sex composition of the first and second borns, birth spacing between the second and third born children.

As shown in table 3, the primary school enrolment rate is 86% for firstborn in 1+ sample, 86% for firstborn in 2+ sample, 84% for firstborn and 84% for secondborn in 3+ sample. Primary graduation rate is relatively lower, 75% for 1+ sample, 74% for 2+ sample, 74% for firstborns and 68% for secondborns in 3+ sample. Following the similar pattern, the rate of junior school enrolment drops quite a lot in all subsamples. While older siblings have similar enrolment rate of primary school, the rate of primary school graduation and secondary school enrolment decreases with household size. It indicates a negative correlation between family size and child schooling outcome, but not necessarily a causal one. It may simply due to the fact that children of smaller families are from relatively younger cohorts and benefit from better education system.

The gender of the first birth is biologically normal(0.51) in the 1+ sample, and it decreases in 2+ sample and declines further in the 3+ sample. In addition, sex ratio increases with birth order[17]. The sex ratio at different sample and birth order provides evidence for son preferences. The twinnng rate at each birth is within the natural range[18].

[16]The Cultural Revolution lasted from 1966 to 1976, and it caused a large scale eduction system disruption.

[17]Yi et al. (1993) show that the sex ratio of the first child is biologically balanced, but it goes up at higher birth orders especially for households with no previous boy birth. Almond et al. (2013) confirm that the sex ratio of the first child is at natural rate 1.05, but the sex ratio of the second child quite differs, with first girl families have persistently higher rate than families with a first boy.

[18]Huang et al. (2016) show that twins rate is around 0.4% for households with children born between 1974-1980, and slightly higher 0.45% in the period 1980-1984.

Table 4 shows that children are relatively older and mothers are younger at first child birth in households with higher fertility, while mother's and father's age at census are quite similar across subsamples. Parents have higher education level are associated with fewer children. There are smaller proportion of parents of Han ethnicity and from urban areas in higher fertility families.

5 Empirical Results

This section tests the Q-Q trade-off for the firstborn child in the 1+ sample, 2+ sample and 3+ sample, respectively; for the secondborn child in the 3+ sample. Four specifications are used for the test in each sample, the author will start the estimation with ordinary least square (OLS) estimator, then move to two stage least square (2SLS) estimator. Since 2SLS estimator deals with the endogenous issue using two stages of linear estimation, this paper considers two additional non-linear estimators–nonlinear first-stage(NLIV) and generalized method of moments(GMM). Due to the forbidden regression problem mentioned earlier, NLIV uses Poisson regression to estimate the first stage effects on fertility and then uses it as an instrument for fertility in a linear regression of outcome variable. GMM estimator employs non-linear estimations at both firth and second stage, i.e. obtaining Poisson estimator for fertility and then use it as instrument for fertility in a binary regression for the schooling outcome variable.

5.1 At Least One Birth Sample

Panel a of table 5 shows the effect of a marginal child at second birth on the education outcomes of the firstborn. All the specifications provide evidence on significant Q-Q trade-off, irrelevant of the outcome variables. OLS estimation suffers from endogeneity problem that parents with different preference for family size may also have different preference for quality of children, and unobserved child's ability may affect parents' investment decision and child's performance. The exogenous variation of sibling size is induced by gender of the first child in the 1+ sample.

In all the cases, 2SLS estimator gives larger Q-Q trade-off than OLS, which implies that the OLS estimator is downward biased. NLIV estimator gives even larger effects than the 2SLS estimator, this is mainly due to the difference in the distribution assumption imposed on fertility. 2SLS assumes that fertility is normal distributed, and NLIV uses Poisson distribution to model fertility. Comparing

to the NLIV estimator, GMM estimator gives a slightly greater effect of family size on the probability of enrolling in primary school, and a smaller effect on the probability of graduating from primary school and enrolling in junior school. For example, GMM estimates shows that an additional birth induced by a firstborn girl decreases the probability of primary school enrolment by 19 percent [19], which is 1% higher than the NLIV estimator [20]. The different magnitude of Q-Q trade-off between NLIV and GMM is resulting from the non-linear distribution assumption imposed on the outcome variable. For example, high density of children's outcome are located at the upper tail of the distribution of the nonlinear model, thus it gives better approximation to the data generating process than the linear probability model.

Note that the magnitude of Q-Q trade-off is considerably higher in the case of junior school enrolment. OLS estimator states that the probability of enrolling in junior school for the firstborn is reduced by 11.1%, and 2SLS suggests a reduction in the probability by half. NLIV gives slightly higher estimate of 61.5%, while GMM estimator shows a 54.3% reduction in the likelihood of enrolling in the junior school if there is a marginal second birth.

By exploring the exogenous variation in family size that is induced by first child's gender, this paper is able to fill the gap of examining the effect of family size increasing from 1 to 2 on firstborn. It provides the technical foundation for family planning policies, particularly it may imply that China's One Child Policy is actually beneficial for child's human capital by restricting the family size, thus reducing resource competition. Furthermore, if policy maker aims at encouraging child birth, then corresponding policy (for example, education and health) reform is needed to cooperating with the significant Q-Q trade-off.

5.2 At Least Two Births Sample

Panel b of table 5 gives the estimation results in the households with at least 2 births. All the OLS estimations show significantly negative effects of family size on the outcome of firstborn children, which are likely to be spurious inverse relationships. An additional child at second birth decrease the probability of enrolment by 3%, the probability of primary graduation by 7% and the likelihood of enrolling in junior school by 12.5%. However, OLS estimation neglects unobserved

[19] Note that column (4) of panel a gives estimates of -0.162, which corresponds to a 19%(0.162/0.857) reduction in the probability of primary school enrolment for the firstborn in 1+ sample.

[20] Column (3) of panel a gives NLIV estimates of -0.15, which corresponds to a 18% reduction.

endowment and parental heterogeneity in preferences for quality of children and quantity of children.

By using twins at second birth as an instrumental variable, the conventional 2SLS estimations show smaller effects in magnitude but no evidence of significant Q-Q trade-off, which is in line with previous literature that using the same methodology. The twins instrument generates an average treatment effect on the non-treated, thus the estimator can be interpreted as the impact of an additional child at second birth on the educational attainment of firstborn child in two singleton births family(non-twins family). Note that 2+ sample shows a different direction of OLS bias than the 1+ sample. Comparing to 2SLS, OLS underestimates the true trade-off on the first born in at least 1 birth family while overestimates the effects in 2+ sample. A marginal child at second birth insignificantly increases the probability of primary school enrolment by 0.6%, the probability of primary graduation by 6.1% and of junior enrolment by 4.4%, for the firstborn.

Move to the NLIV estimates, the Q-Q trade-off turns significant, the effects are -2%, -20% and -35%, respectively. GMM estimation supports a significant Q-Q trade-off by accounting for the non-normal distribution of both outcome and endogenous variables. The schooling outcome of firstborn child is reduced by 8.3%, 12.2% and 15.5%, respectively for the probability of primary enrolment, primary graduation and junior enrolment. This finding contrasts the majority of previous literature which finds no Q-Q trade-off, but somehow in line with the results of Rosenzweig and Zhang (2009). They use a logit model and show that twinning reduces the likelihood of college attendance by 27%.

5.3 At Least Three Births Sample

Estimation results from the at least three births sample give a different picture than the at least two births sample. Panel c of table 5 shows effects of a marginal fourth child on firstborn and secondborn children, separately. As it in the case for 2+ sample, OLS estimators show significant negative effects of an extra child on the outcome of elderly siblings, with the effect is slightly greater in magnitude on the secondborn than the firstborn.

This paper examines the effects on the firstborn and secondborn separately, because children of different birth orders experience different degrees of sibling interaction and family resource allocation. Generally the 2SLS estimates show no significant trade-off, with one exception that the probability of enrolling in junior school is significantly reduced by 22.4 percent for the secondborn child. NLIV

estimate shows greater effects in magnitude than 2SLS. It gives insignificant Q-Q trade-off on primary school enrolment, and significant trade-off on the other two measurements of schooling outcome. GMM estimator shows that the only significant adverse effect on the firstborn children is the 3.3 percent reduction on the likelihood of primary school graduation, while there is no significant trade-off on enrolment of primary school and junior high school. The secondborn sibling suffers more of an additional fourth child than the firstborn sibling. An additional child decreases the probability of primary school enrolment by 5.5%, primary school graduation by 8.4%, and junior high school enrolment by 10.1% for the secondborn children. This is consistent with Black et al. (2005) in that they find a monotonic decline in educational attainment as the birth order goes up.

It is worth noting that in the summary statistics, firstborn children in 3+ sample have similar rate of primary school enrolment with 1+ and 2+ sample, but lower rate of primary school graduation and junior high school enrolment. It may imply that 3+ families are more financially constrained if resource competition is more intense in families with more children. Families make different decisions on whether to send kids to junior high school or not, it could be the case that more affluent families have higher propensity sending kids to junior high school. Therefore the trade-off is less obvious for those families who are less likely to financially affected by the additional fourth child.

Comparing the trade-off effects in 1+, 2+ and 3+ sample, I find that the trade-off is more evident in a relatively smaller family. It maybe due to the fact that conditional on a bigger family size, the resource is spread more thinly and thus the marginal decrease in the welfare of older sibling declines with child quantity. In other words, for those families who prefer to have a larger family size, an extra child affect the division of family resources less as family size goes up. The smaller magnitude of the estimators obtained for a larger family size provides some evidence on the non-linearity of the trade-off as quantity goes up. In addition, there is evidence of birth order effect. Generally, Q-Q trade-off is greater for the secondborn than for the firstborn in at least 3 births family. Note that the majority of the sample is consists of rural households, thus the estimation is more reflexive for less developed part of China.

5.4 Robustness Check

5.4.1 Instruments Validity

There are concerns about the validity of the instruments if the gender of firstborn and twins affect the schooling outcome independently from family size.

Twinning may affect the outcome of lower births children through zero spacing, lower birthweight and economies of scale. The spacing channel on the outcome of the firstbirth can be tested through 3+ sample without twins but with two tightly spaced following births (Black et al., 2005). Panel (a) of table 6 shows that spacing has nearly zero effect on the outcome of older sibling. If this can be extrapolated to the case of twins, then twins will not affect outcome of older sibling independently through family size.

Alternatively, the validity of twins instrument can be tested as proposed by Angrist et al. (2010), they estimate reduced forms in the no-first-stage sample. The idea is that a valid instrument should not affect outcome independently of family size, therefore by shutting down the channel of family size, twin births should have no effect on outcome. The no-first-stage sample consists of families who are likely to have a bigger family anyway, regardless of the twins birth. Indicators for preferring a larger quantity of children could be a tighter spacing of subsequent births and an younger age of mother at first child birth. Panel (b) of table 6 shows results based on no-first-stage samples, there is no evidence to reject the validity of twin births as instruments.

Rosenzweig and Zhang (2009) propose that birthweight deficit of twins may induce parents to allocate resources toward lower birth order non-twins children, if this reinforcement story is true, then the estimators obtained through twins at second and third births will be biased positively. My data provides no information on birthweight, although birthweight itself is endogenous. This paper tests the Q-Q trade-off following the bounding strategy proposed by Rosenzweig and Zhang(2009),

$$Y_{ij} = \alpha_1 T_j + \alpha_2 F_{ij} + \alpha_3 (T_j * F_{ij}) + X_i \theta + e_{ij} \tag{10}$$

where T_j is an indicator for a household j with twin at second(third) birth, F_{ij} is a dummy for singleton first-born(singleton first two births) and their interaction are included in the regression. The variable of interest is the interaction term, with a positive coefficient α_3 implying parental reinforcement behaviour. I observe reinforcement effects toward singleton birth child in the at least two births sample, while the effects are not evident in the at least three births sample (table 7).

Following the idea of Rosenzweig and Zhang (2009), the effects of twinning on non-twins older sibling provide the lower bound, and the effects of twinning on twins themselves provide the upper bound. As it indicated by table 7, the Q-Q trade-off obtained through twins at second birth gives the lower bound, it would be informative to also know the upper bound by using twins at first birth. Twins at first birth generates bigger effects of fertility on schooling outcome (table 8). These estimates are big in magnitude but comparable to the previous finding of Rosenzweig and Zhang (2009) where they find a 20 percent decrease in expected college enrolment and literature grade. Since twining reduces birthweight which is positively correlated with the reinforcement behaviour, the estimations without conditional on birthweight are negatively biased. There is caveat using twin at first birth as an instrumental variable. Twins themselves are not directly comparable to non-twins due to birthweight deficit, the reader should bear in mind that the true Q-Q trade-off effect is possibly overestimated in the case of using twins at first birth as instrumental variable.

Note that estimation based on twins at first birth is not directly comparable to twins at higher birth order and gender of the first birth for three reasons. First, twins at first birth identifies the effect of an additional birth on the average outcome of twins compare to the outcome of singleton birth, while twins at higher birth orders gives the effects of an additional birth on singleton births. Second, twins at first birth is more likely to be an exogenous shock in the sense that parents have no prior on their children, therefore they are less likely to have reinforcement or compensate behaviour. On the contrary, twins at higher birth order may induce parents who value the quality of children to allocate the resource away from twins, which partly offsets the true Q-Q trade-off. Third, different instruments generate different complier population and capture different local average treatment effects(LATE) (Angrist et al., 2010). Twins at first birth affect family size at the incidence of twinning, while girl at first birth affects fertility at a relatively bigger range for families have son preferences and the range will be wider if this preference is stronger. Overall, twins at first birth provide the upper bound of Q-Q trade-off as suggested by Rosenzweig and Zhang (2009), which is consistent with the finding that estimators generated by twins at first birth are bigger in magnitude compares to the case of twins at higher birth order and gender at first birth. As Angrist et al. (2010) point out, different instrument is suffering from different omitted variable bias, therefore comparing the estimation results from two different instrumental strategies would provide a specification check itself.

An additional issue of using twins as instrument is the possibility that twins may

affect outcome through economies of scale, depending on the tighter spacing and the gender composition of twins birth. If twins of the same gender benefit from sharing clothes and rooms, then one would expect smaller trade-off to be found in the same gender twins. It twins of the mixed gender benefit from the interaction of different-gender sibling, then a smaller trade-off would be found in the mixed gender twins. The best scenario for twins to be a valid instrument is to have estimation results irrelevant of gender composition of twins. I compare results based on two subsamples of 1+ sample. The same-sex subsample is consists of singleton boy(girl) at first birth and same-sex twins at first birth; the mixed-sex subsample is comprised of singleton at first birth and mixed-sex twins at first birth. If economies of scale do exist, then the formal subsample will be expected to exhibit significantly less Q-Q trade-off. Table 9 shows no evidence that twins affect outcome variable directly through economies of scale. There are no significant differences between two subsamples among all outcome variables. For example, the first two columns show the slope of regression line -0.194 and -0.210, clearly their confidence interval are largely overlapped. Therefore, it seems that the internal validity of twins instrument is not threatened in this context.

There is also concern about using the gender of firstborn as instrument for fertility, it may affect outcomes through parental preference in terms of resource allocation. A direct test is to compare the average outcomes of children by number of boys in the family and family size (Lee, 2008). Panel a of table 10 shows that households with more boys have higher probability enrolling in primary school conditional on family size, however, the probability of completing primary school and enrolling in secondary school decreases with more boys in the household. This can be regarded as evidence that there is no systemic gender bias in parental investment. Note that this comparison is intuitive, while it does not account for any endogeneity. If parents prefer to invest more in boy rather than girl, then the trade-off is likely to be positively biased for boys and negatively biased for girls. Additionally, panel b of table 10 performs a reduced form test in the no-first-stage sample. Indicators for preferring a larger quantity of children could be tighter spacing of subsequent births, younger mother at first child birth and one of the parent belongs to ethnic minorities[21]. Unfortunately, the author fails to rule out the direct impact of first child's gender on education outcomes, but the direction of the bias can be inferred. Since there is a negative correlation between girl birth and outcome variables, the true Q-Q trade-off is overestimated.

[21]Ethnic minorities are waved from the One Child Policy. I would expect girl at first birth has less impact on the family size in the non-han sample than the Han-sample.

5.4.2 Heterogeneous Effects

There is some evidence in the literature that the quantity-quality effect differs by families' socioeconomic background. In order to check whether it also applies to this study, I identify potential heterogeneous effects by stratifying the sample by mother's education level. I also test the sensitivity of results to other stratifications of the sample, including rural-urban gap and gender gap. For the ease of presentation, this subsection only reports results based on GMM estimates. In addition, following the logic of bounding strategy (Rosenzweig and Zhang (2009)), GMM estimation based on twins at first birth is also shown in this subsection, despite its problem mentioned above[22].

To identify the heterogeneity in terms of mother's education level, children have been grouped into two sub-samples, depending on whether mothers have completed primary school. The Q-Q trade-off is more evident in families with lower educated mother in 1+ and 2+ samples (table 11). In fact, the big negative effects found in table 5 are mainly driven by households with low educated mothers. There is relative smaller or no significant negative effects of family size on primary school graduation and junior high school enrolment for children of high educated mothers. However, estimation results turn opposite in the case of at least three births family. The Q-Q trade-off is not subject to mothers' education level when the outcome is measured by primary school enrolment, but children in high-educated families turn out to suffer more of the adverse fertility effects than children in low-educated families in the case of primary graduation and junior high school enrolment. Possible explanation could be drawn from mother's labour force participation. I find that mother's employment status keeps stable when family size increase from 2 to 3, but decreases significantly when it increases from 3 to 4. If education level could proxy the wealth of the family, then children of less educated mothers suffer from more severe resource competition, that explains why the Q-Q trade-off is more evident for children of less educated mothers in the 1+ and 2+ sample. However, since mothers reduce labour supply in 3+ sample, then education level is a poor indicator for wealth effect in this case. Conditional on enrolled in primary school, children of high educated mothers have much higher probability of graduation rate and enrolment rate of junior school[23]. However, the

[22]There is caveat using twin at first birth as an instrumental variable. On one hand, twins themselves are not directly comparable to non-twins due to birthweight deficit, therefore, twins at first birth provides the upper bound of Q-Q trade-off. On the other hand, twins at first birth is more likely to be an exogenous shock to the family. Parents have no prior on their children, therefore they are less likely to have reinforcement or compensate behaviour. On the contrary, twins at higher birth order may induce parents who value the quality of children to allocate the resource away from twins, which partly offsets the true Q-Q trade-off.

[23]Conditional on enrolled in primary school, the probability of graduation is 0.69 for low-

adverse effect of a marginal fourth child is bigger conditional on a high-educated family than conditional on a low-educated family.

Rural households are more likely to be financially constrained and the education system in rural areas are relatively poorer. Therefore, it is reasonable to consider that the fertility effects would vary by rural and urban areas. Table 12 presents considerable adverse effects on children's outcome in rural areas. The significant Q-Q trade-off found in the main regression largely attributes to children of rural households, especially rural households account for a large fraction of my sample. The rural-urban gap shows a different pattern for first born and second born children in families with at least three births. There is no significant difference found for the first born children, but a rural-urban gap is observed in the probability of graduating from primary school for the second born children. A marginal fourth child significantly decreases the probability of completing primary school for the second born by 5.1 percentage points in rural areas, while there is no trade-off found for the urban areas.

If families have strong son preferences, then an additional girl birth will affect the performance of older siblings to a less extent than an extra boy birth. I constructed two sub-samples for 2+ sample—one sub-sample excludes households with twin girls at second birth to ensure that one of the twins must be a boy, and the other sub-sample excludes households with twin boys at second birth to ensure that one of the twins must be a girl. The same restriction applies to 3+ sample and the subjects of interest are the average outcomes of first two births. Table 13 shows no significant gender gap by the marginal child, and the magnitude of the effect of an extra boy birth is very similar to an extra girl birth. It could be explained by the possibility that families decide to have larger families have less son preference for gender composition of children. In other words, there is no post-natal gender preference conditional on child birth, parents treat the new born child indifferently of gender. Almond et al. (2013) argue that sex mix is most preferred, conditional on having two children in the family.

I further test Q-Q trade-off by gender composition of older siblings. Table 14 shows that firstborn girls suffer more from an additional child birth comparing to firstborn boys in primary school education. While there is no significant differences observed for secondary school enrolment in 2+ sample, results for 3+ sample show that the rate of junior high school enrolment is reduced by 10 percentage points for boys but no impacts on girls. The adverse effect on the primary school education of girls may imply that families with son preferences suppress the

educated families and 0.84 for high-educated families; the likelihood of enrolling in junior school is 0.48 for low-educated families and 0.76 for high educated families.

education opportunity for girls at an early stage. However, given no significant differences on primary school graduation rate, the enrolment of junior high school may be related with the higher education cost incurred, if boys require higher costs than girls and families are financially constrained, then an additional fourth child affects more negatively on boys.

5.4.3 One Child Policy

The overlap of my sample period and the One Child Policy(OCP) raises concern for the credibility of the estimation. McElroy and Yang (2000) finds that the removal of the One Child Policy would increase family size by one third of a child, and Qian (2009) shows that relaxation of the policy in 1982 for rural families with first born girls lead to an increase in fertility by one fourth of a child. Rosenzweig and Zhang (2009) show that the policy would lead to a less than 4 percent increase in educational attainment, a less than 9 percent increase in the expected college enrolment. Liu (2014) suggests the men's income would be increased by 1.71% and women's income by 1.55% at most by One-Child Policy. Overall, these analysis show that the consequences of the One Child Policy on fertility is modest and on human capital accumulation is negligible.

Although sex-selective abortion is not widely available until late 1980s, it is reasonable to suspect validity of instrumental variables as the ultrasound services to some extent allows parents to decide the gender composition of the children. If this bias exist, I would expect the contamination of estimation is less severe in a pre One Child Policy period. My sample period covers the implementation of the One Child Policy since 1979, a sensitivity test is carried out using 1982 census. If parents are allowed to choose the quantity of children, the willingness to trade quality for quantity would indicate a bigger Q-Q trade-off. However, table 15 shows that the Q-Q trade-off is smaller in the absence of One Child Policy. There could be two candidate explanations. First, the children in 1982 sample were born between 1966 and 1976, which coincided with the cultural revolution period. This revolution mainly disrupted senior high school and higher education, and primary school and junior high school education system were disorganised. Meng and Gregory (2007) describe that primary school students were not taught in standard curricula and they spent most of their school time doing manual work in factories and the countryside. Although surprisingly the primary school graduation rates(around 90%) were higher than the 1990 sample, it is likely that that there was a change in the graduation criterion, which made the trade-off less evident. Second, it could be that those who violated the OCP for an extra birth suffered

from bad economics and socioeconomic situation so that it adversely affected the outcome of children.

6 Conclusion

This paper studies the effect of fertility on children's human capital accumulation in the context of China. The main contribution comes from applying a novel GMM approach for testing the existence of Q-Q trade-off. This method is more favourable than conventional two stage least squares because it takes into account the limited dependent variable nature of the data. Thus, this paper not only considers the non-linear fertility effect across parities by examining different birth sample, but also accounts for non-linear distribution within parities. While the conventional approach shows no effect of fertility on education performance of children, GMM estimators give significant Q-Q trade-offs. This finding is consistent with the Becker-Lewis model that an increase in child quantity increases the shadow price of quality. Moreover, the trade-off nonlinearly decreases with family size and shows heterogeneous effects by birth order.

This paper is the first to apply first child's gender as an instrumental variable for testing Q-Q trade-off in the context of China where the son preferences are prevailing. This instrument offers the possibility to test the Q-Q trade-off of a marginal second birth on the first born while typically twins and parental preference for a mixed gender composition estimate the effect of a marginal third or higher birth order child on older siblings.

Although arguably endogeneous, this analysis also employs twins at first birth as an instrumental variable for fertility to provide an upper bound estimation for the Q-Q trade-off. Rosenzweig and Zhang (2009) suggest that the upper bound estimator is negatively biased and the lower bound estimator is positively biased, if not accounting for the birthweight deficit of twins and parental reinforcement behaviour. There is no birthweight information available in the census, but I find evidence of parental reinforcement behaviour in the 2+ sample, by comparing the effect of twinning on twins and non-twins. Therefore, estimator in the 2+ sample is regarded as a positively biased lower bound. In other words, the true Q-Q trade-off are bigger in magnitude for the 2+ sample. The upper bound is provided by using twins at first birth. Additionally, the effect of fertility on the outcome of older siblings are provided for first and second birth separately in the 3+ sample, as I believe that the trade-off would differ by birth order due to differences in sibling interaction and parental resource allocation. Indeed, results

show heterogeneous birth order effects, which are obscured by previous literature.

Several validity tests are conducted to ensure that the estimates are not driven by direct effects of instrumental variables on outcomes, if any, at least the direction of bias is obtained. Twinning may affect the outcome of lower births children through zero spacing(economies of scale) and lower birthweight. I test the existence of these mechanisms and results suggest that some estimators are probably positively biased. Thus the true adverse effects of family size on the schooling performance of children are bigger in magnitude, with the exception of the effect on twin themselves. Son preferences may affect outcome independently through gender-based parental investment, for instance, families with son preferences will tend to allocate resource toward boys and the Q-Q trade-off will be positively biased for boys and negatively biased for girls. There is evidence that gender of firstborn directly affects outcomes, thus Q-Q trade-off is negatively biased. In other words, the true effects of an additional birth induced by a girl birth on quality is weaker than what the estimators imply.

In addition, I explore heterogeneous effects across family socioeconomics background, rural urban areas, gender of older siblings and extra child. The Q-Q trade-off are mainly driven by children of households from rural areas and low-educated parents in the sample of at least one or two births, while the trade-off is more obvious in high-educated parents in a larger family size. Results based on gender confirms the finding of Almond et al. (2013) that parents prefer sex mix than 2 boys, followed by 2 girls.

This study provides technique support for policies aiming at reducing contraceptive costs, controlling population growth and subsidizing families with fewer children. The estimates show that those policies may in fact contribute to better educational outcomes of children. However, population control is unsustainable for long term economic growth especially in ageing societies. If policy maker aims at alleviating population ageing problem by encouraging child birth, then corresponding policy reform is needed to cooperate with the adverse fertility effect. For example, maternal leave, childcare provision, public education and health system should be adjusted accordingly.

References

Almond, D., H. Li, and S. Zhang (2013). Land reform and sex selection in china. *NBER Working Paper No. 19153*.

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

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Becker, G. S. and T. Nigel (1973). On the interaction between quantity and quality of children. *Journal of Political Economy 81*(2), 279–288.

Becker, G. S. and T. Nigel (1976). Child endowments and the quantity and quality of children. *Journal of Political Economy 84*(4), 143–162.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005). The more the merrier? the effect of family size and birth order on children's education. *The Quarterly Journal of Economics*, 669–700.

Das Gupta, M., J. Zhenghua, L. Bohua, X. Zhenming, W. Chung, and B. Hwa-Ok (2003). Why is son preference so persistent in east and south asia? a cross-country study of china, india and the republic of korea. *The Journal of Development Studies 40*(2), 153–187.

Huang, W., X. Lei, and Y. Zhao (2016). One-child policy and the rise of man-made twins. *Review of Economics and Statistics 98*(3), 467–476.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

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

Li, H., J. Zhang, and Y. Zhu (2008). The quantity-quality trade-off of children in a developing country: Identification using chinese twins. *Demography 45*(1), 223–243.

Lin, M.-J., N. Qian, and J.-T. Liu (2008). More women missing, fewer girls dying: The impact of abortion on sex ratios at birth and excess female mortality in taiwan. *National Bureau of Economic Research No. 14541*.

Liu, H. (2014). The quality–quantity trade-off: evidence from the relaxation of china's one-child policy. *Journal of Population Economics 27*(2), 565–602.

McElroy, M. and D. T. Yang (2000). Carrots and sticks: fertility effects of china's population policies. *The American economic review 90*(2), 389–392.

Meng, X. and R. Gregory (2007). Exploring the impact of interrupted education on earnings: the educational cost of the chinese cultural revolution. *IZA discussion paper*.

Mogstad, M. and M. Wiswall (2016). Testing the quantity–quality model of fertility: Estimation using unrestricted family size models. *Quantitative Economics 7*(1), 157–192.

Qian, N. (2009). Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural china. *National Bureau of Economic Research No. 14973*.

Rodriguez, G. (2007). Lecture notes on generalized linear models. *http://data.princeton.edu/wws509/notes/ 4*.

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

Rosenzweig, M. R. and J. Zhang (2009). Do population control policies induce more human capital investment? twins, birth weight and china's "one-child" policy. *The Review of Economic Studies 76*(3), 1149–1174.

Schultz, T. P. (2007). Population policies, fertility, women's human capital, and child quality. *Handbook of development economics 4*, 3249–3303.

Silva, J. S., S. Tenreyro, and K. Wei (2014). Estimating the extensive margin of trade. *Journal of International Economics 93*(1), 67–75.

Wang, W. and F. Famoye (1997). Modeling household fertility decisions with generalized poisson regression. *Journal of Population Economics 10*(3), 273–283.

Willis, R. J. (1973). A new approach to the economic theory of fertility behavior. *Journal of political Economy 81*(2), S14–S64.

Windmeijer, F. A. and J. M. Santos Silva (1997). Endogeneity in count data models: an application to demand for health care. *Journal of applied econometrics 12*(3), 281–294.

Winkelmann, R. and K. F. Zimmermann (1994). Count data models for demographic data. *Mathematical Population Studies 4*(3), 205–221.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Yi, Z., T. Ping, G. Baochang, X. Yi, L. Bohua, and L. Yongpiing (1993). Causes and implications of the recent increase in the reported sex ratio at birth in china. *Population and Development Review*, 283–302.

Table 1: Relevant Test: First Child's Gender

	Parity Progression			Weibull Hazard Model of Childbirth Spacing		
	No.children>1	No.children>2	No.children>2	spacing12	spacing23	spacing23
girl1	0.055***	0.095***	0.005	1.564***	1.539***	0.946***
	(0.003)	(0.003)	(0.003)	(0.007)	(0.014)	(0.013)
girl2		0.129***			1.625***	
		(0.003)			(0.015)	
samesex12		0.112***			1.197***	
		(0.003)			(0.011)	
boy12			-0.001			0.736***
			(0.003)			(0.011)
girl12			0.094***			1.948***
			(0.003)			(0.022)
N	60437	60437	58883	566133	222428	222428

Note: All the number contained in the variable name indicates birth order, for instance, spacing12 means spacing between first and second births. Robust standard errors are reported in parentheses.

Table 2: First Stage Effects on Fertility

Panel a: 1+ sample

Sub-sample	Primary enrolment		Primary graduation	
Instrument	**twin1**	**girl1**	**twin1**	**girl1**
No. Children	0.513***	0.275***	0.372***	0.298***
	(0.009)	(0.002)	(0.025)	(0.004)
Wald test	3280.220	23884.040	221.960	5140.730
N	610484	606422	148969	148303

Panel b: 2+ sample

Sub-sample	Primary enrolment	Primary graduation
Instrument	**twin2**	**twin2**
No. Children	0.561***	0.562***
	(0.008)	(0.027)
Wald test	4899.800	425.430
N	429120	137560

Panel c: 3+ sample

Sub-sample	Primary enrolment	Primary graduation
Instrument	**twin3**	**twin3**
First born		
No. Children	0.426***	0.523***
	(0.011)	(0.027)
Wald test	1441.900	389.080
N	153693	69735
Second born		
No. Children	0.482***	0.547***
	(0.013)	(0.032)
Wald test	1411.370	295.600
N	163451	58141

Note: Panel a shows the first-stage effects of twins and girl at first birth on fertility for the 1+ sample. Panel b shows the first-stage effects of twins at second birth on fertility for the 2+ sample. Panel c shows the first-stage effects of twins at third birth on fertility for the 3+ sample.

All the number contained in the variable name indicates birth order, for instance, girl1 indicates girl at first birth and twin2 means twins at second birth.

Regressions for panel a include controls for child's age, gender, mother's age at first child birth, father's age and year of birth, panel b also controls for mother's age at census, birth spacing between first and second birth, panel c additionally controls for sex composition of the first two birth, birth spacing between the second and third birth. Robust standard errors are reported in parentheses and are clustered at household level for the 3+ sample.

Table 3: Descriptive Statistics by Sample 1/2

	1+ sample		2+ sample		3+ sample			
					First born		Second born	
	mean	std dev	mean	std dev	mean	std dev	mean	std dev
Endogeneous variable								
No. children	2.152	0.935	2.531	0.764	3.336	0.629	3.416	0.703
Education outcomes								
Primary Enrolment ($6 <= age <= 13$)	0.857	0.351	0.865	0.342	0.841	0.366	0.841	0.366
N	610484		429120		153693		163451	
Primary Completed ($13 < age <= 16$)	0.752	0.432	0.743	0.437	0.681	0.466	0.653	0.476
Junior Enrolment ($13 < age <= 16$)	0.602	0.490	0.586	0.493	0.485	0.500	0.456	0.498
N	148969		137560		69735		58141	
Family Composition								
twin1	0.006	0.079						
twin2			0.004	0.063				
twin3					0.003	0.053	0.003	0.052
boy1	0.512	0.499	0.482	0.500	0.409	0.492	0.422	0.494
boy2			0.526	0.499	0.420	0.494	0.433	0.496
boy3					0.541	0.498	0.531	0.499
boy12			0.241	0.428	0.198	0.399	0.206	0.404
girl12			0.234	0.423	0.369	0.482	0.351	0.477
boy123					0.087	0.282	0.092	0.289
girl123					0.141	0.348	0.141	0.348
N	759453		566680		223428		221592	

Note: 1+ sample includes households with at least 1 child birth(note here 1+ sample is the sample I use for estimation with twinning at first birth as instrument. When first child' sex is used as instrumental variable, the 1+ sample excludes families with twins at first birth, and all the summary statistics remains largely unchanged); 2+ sample is consists of families with at least 2 child births, 3+ sample is comprised of households with at least 3 births.

For the ease of table presentation, it reports the summary statistics for aggregated sample of primary school enrolment and junior school enrolment. However, in the regression analysis, sample differs by children exposed to the outcome. Specifically, in the case of primary enrolment, the sample only contains children age between 6 and 13; in other cases, sample is comprised of children age above 13 but not older than 16.

All the number contained in the variable name indicates birth order, for instance, twin1 means twins at first birth, spacing12 means spacing between first and second births.

Table 4: Descriptive Statistics by Sample 2/2

	1+ sample		2+ sample		3+ sample			
					First born		Second born	
	mean	std dev	mean	std dev	mean	std dev	mean	std dev
Control Variables								
child's age	10.194	3.110	10.661	3.168	11.240	3.229	11.022	3.037
child's birth year	1979.260	3.143	1978.784	3.198	1978.198	3.261	1978.444	3.076
mother's age at first child birth	23.814	3.025	23.334	2.844	22.920	2.835	22.697	2.792
mother's age at census	34.008	4.143	33.995	4.147	34.160	4.090	36.180	3.734
father's age at census	36.153	4.663	36.196	4.647	36.399	4.645	38.489	4.462
Mother Education level								
illiterate	0.359	0.480	0.426	0.494	0.510	0.500	0.544	0.498
primary school completed	0.311	0.463	0.346	0.476	0.333	0.471	0.335	0.472
junior high school completed	0.221	0.415	0.173	0.378	0.127	0.333	0.100	0.301
senior high completed	0.099	0.298	0.054	0.226	0.030	0.171	0.020	0.141
college completed	0.009	0.095	0.002	0.048	0.000	0.016	0.000	0.021
Father Education level								
illiterate	0.158	0.364	0.183	0.386	0.216	0.412	0.236	0.425
primary school completed	0.320	0.466	0.359	0.480	0.368	0.482	0.403	0.491
junior high school completed	0.344	0.475	0.322	0.467	0.302	0.459	0.271	0.444
senior high school completed	0.153	0.360	0.127	0.333	0.112	0.315	0.086	0.280
college completed	0.025	0.157	0.009	0.095	0.002	0.046	0.004	0.060
mother ethnicity(Han)	0.931	0.254	0.917	0.276	0.884	0.320	0.889	0.314
father ethnicity(Han)	0.930	0.254	0.917	0.275	0.887	0.317	0.891	0.312
mother's birth year	1955.477	4.139	1955.491	4.142	1955.332	4.082	1953.309	3.725
father's birth year	1953.323	4.660	1953.281	4.643	1953.084	4.640	1950.991	4.457
spacing12	37.625	22.703	37.808	22.580	30.525	15.104	29.708	13.583
spacing23					35.338	20.610	38.324	22.938
rural	0.775	0.418	0.894	0.308	0.955	0.208	0.938	0.242
N	759453		566680		223428		221592	

Note: Same note as it stated in table 3.

Table 5: The Effect of Family Size on Education Outcome

	(1) OLS	(2) 2SLS[a]	(3) NLIV[b]	(4) GMM[c]	(5) OLS	(6) 2SLS	(7) NLIV	(8) GMM
	Panel a: 1+ sample				**Panel b: 2+ sample**			
	IV: girl at first birth				**IV: twin at the second birth**			
Primary Enrolment								
No. kids	-0.019***	-0.078***	-0.150***	-0.162***	-0.023***	0.005	-0.015***	-0.072***
	-0.022[d]	**-0.091**	**-0.175**	**-0.189**	**-0.027**	**0.006**	**-0.017**	**-0.083**
	(0.001)	(0.003)	(0.003)	(0.002)	(0.001)	(0.008)	(0.004)	(0.004)
N	606422	606422	606422	606422	429120	429120	429120	429120
Primary Graduation								
No. kids	-0.041***	-0.089***	-0.163***	-0.115***	-0.049***	0.045	-0.151***	-0.091***
	-0.055	**-0.118**	**-0.217**	**-0.153**	**-0.066**	**0.061**	**-0.203**	**-0.122**
	(0.001)	(0.007)	(0.007)	(0.009)	(0.001)	(0.030)	(0.009)	(0.011)
N	148303	148303	148303	148303	137560	137560	137560	137560
Junior Enrolment								
No. kids	-0.067***	-0.300***	-0.370***	-0.327***	-0.073***	0.026	-0.203***	-0.091***
	-0.111	**-0.498**	**-0.615**	**-0.543**	**-0.125**	**0.044**	**-0.346**	**-0.155**
	(0.001)	(0.008)	(0.008)	(0.005)	(0.002)	(0.034)	(0.009)	(0.014)
N	148303	148303	148303	148125	137560	137560	137560	137394
	Panel c: 3+ sample							
	First Born				**Second Born**			
	IV: twin at third birth				**IV: twin at third birth**			
Primary Enrolment								
No. kids	-0.025***	-0.001	0.013	-0.011	-0.033***	-0.008	-0.001	-0.046***
	-0.030	**-0.001**	**0.015**	**-0.059**	**-0.039**	**-0.010**	**-0.001**	**-0.055**
	(0.001)	(0.017)	(0.010)	(0.013)	(0.001)	(0.018)	(0.009)	(0.007)
N	153693	153693	153693	153693	163451	163451	163451	163424
Primary Graduation								
No. kids	-0.048***	0.044	-0.057***	-0.033*	-0.053***	-0.022	-0.074***	-0.055**
	-0.070	**0.065**	**-0.084**	**-0.048**	**-0.081**	**-0.034**	**-0.113**	**-0.084**
	(0.002)	(0.035)	(0.016)	(0.017)	(0.003)	(0.041)	(0.015)	(0.022)
N	69735	69735	69735	69710	58141	58141	58141	58141
Junior Enrolment								
No. kids	-0.066***	0.036	-0.045***	-0.009	-0.068***	-0.102**	-0.053***	-0.046**
	-0.136	**0.074**	**-0.093**	**-0.019**	**-0.149**	**-0.224**	**-0.116**	**-0.101**
	(0.002)	(0.040)	(0.007)	(0.019)	(0.002)	(0.044)	(0.014)	(0.021)
N	69735	69735	153693	69556	58141	58141	58141	58054

Note: Panel a shows the effects of fertility on schooling outcome of firstborns in the 1+ sample. Panel b shows the effects of fertility on schooling outcome of firstborns in the 2+ sample, panel c shows the effect on schooling outcome of firstborns and secondborns, respectively, in the 3+ sample.

[a] 2SLS gives the two stage least square estimates. [b] NLIV gives the estimates based on non-linear first-stage. [c]GMM gives the result taking into account non-linear distribution in both first and second stage.

[d] Numbers in bold gives the corresponding transformation of the magnitude of Q-Q trade-off, i.e. **-0.022** means that an extra kid decreases the probability of primary enrolment by 2.2%, which corresponds to 1.9 percentage points decrease in the outcome variable compared to its mean value(0.019/0.857). The same calculation applies to all the estimates.

All the regressions include control variables described in the note of table 2. Robust standard errors are reported in parentheses.

Table 6: Testing the Internal Validity of Instruments — Birth Spacing

Panel a: spacing channel—Black et al.(2005)

	Primary enrolment	Primary graduation	Junior enrolment
spacing23	-0.000***	0.000***	0.000**
	(0.000)	(0.000)	(0.000)
N	207734	207752	207218

Panel b: no-first-stage—Angrist et al. (2010)

	Primary enrolment	Primary graduation	Junior enrolment
	Tight birth spacing sample		
twins at second birth	0.017	0.066	0.019
	(0.014)	(0.054)	(0.054)
N	142381	40431	40394
twins at third birth	-0.018	-0.149	-0.177
	(0.027)	(0.091)	(0.119)
N	27846	6278	6278
	Young mother sample		
twins at second birth	-0.028	0.010	-0.037
	(0.019)	(0.044)	(0.051)
N	104272	46338	46281
twins at third birth	0.064	-0.015	-0.144*
	(0.066)	(0.069)	(0.082)
N	26234	13530	13532

Note: In panel a, sample is restricted to non-twin families with two tightly spaced second and third births.

Panel b performs reduced form estimations in no-first-stage samples as proposed by Angrist et al. (2010). Those samples are families with tightly spaced births, or households with young mothers. Robust standard errors are reported in parentheses and are clustered by household ID for 3+ sample.

Table 7: Testing the Internal Validity of Instruments — Bounding Strategy

Panel a: 2+ sample

	Primary enrolment	Primary graduation	Junior enrolment
twins at second birth	-0.023***	-0.010	-0.047**
	(0.006)	(0.017)	(0.020)
first-birth (non-twin)	0.004***	0.045***	0.064***
	(0.001)	(0.003)	(0.003)
twins at second birth	0.024***	0.037	0.063**
* first-birth (non-twin)	(0.009)	(0.029)	(0.032)
N	802109	258258	258258

Panel b: 3+ sample

	Primary enrolment	Primary graduation	Junior enrolment
twins at third birth	-0.025	-0.073	-0.053
	(0.019)	(0.093)	(0.120)
first two births (non-twin)	0.029***	0.067***	0.059***
	(0.003)	(0.009)	(0.011)
twins at third birth	-0.004	0.053	-0.011
* first two births (non-twin)	(0.028)	(0.093)	(0.120)
N	155408	48768	48705

Note: Panel a applies the bounding strategy (equation 10) proposed by Rosenzweig and Zhang (2009) to 2+ sample, while panel b performs similar methodology on 3+ sample. Robust standard errors are reported in parentheses and are clustered by household ID for 3+ sample.

Table 8: The lower bound of using twin births as instrument

	(1) OLS	(2) 2SLS[a]	(3) NLIV[b]	(4) GMM[c]
IV: twin at the first birth				
No. childnum	-0.017***	-0.018*	-0.226***	-0.203***
	-0.020[d]	**-0.021**	**-0.264**	**-0.237**
	(0.001)	(0.009)	(0.005)	(0.007)
N	610484	610484	610484	610484
No. childnum	-0.039***	-0.185***	-0.380***	-0.262***
	-0.052	**-0.246**	**0.505**	**-0.348**
	(0.001)	(0.044)	(0.017)	(0.007)
N	148969	148969	148969	148969
No. childnum	-0.059***	-0.212***	-0.563***	-0.343***
	-0.098	**-0.352**	**0.935**	**-0.570**
	(0.001)	(0.045)	(0.022)	(0.031)
N	148969	148969	148969	148791

Note: Panel a shows the effects of fertility on schooling outcome of firstborns in the 1+ sample. Panel b shows the effects of fertility on schooling outcome of firstborns in the 2+ sample, panel c shows the effect on schooling outcome of firstborns and secondborns, respectively, in the 3+ sample.

[a] 2SLS gives the two stage least square estimates. [b] NLIV gives the estimates based on non-linear first-stage. [c]GMM gives the result taking into account non-linear distribution in both first and second stage.

[d] Numbers in bold gives the corresponding transformation of the magnitude of Q-Q trade-off, i.e. **-0.020** means that an extra kid decreases the probability of primary enrolment by 2%, which corresponds to 1.7 percentage points decrease in the outcome variable compared to its mean value(0.017/0.857). The same calculation applies to all the estimates.

All the regressions include control variables described in the note of table 2. Robust standard errors are reported in parentheses.

Table 9: Testing the Internal Validity of Instruments — Economies of Scale

	Economies of scale					
	Primary Enrolement		Primary Graduation		Junior Enrolment	
	same-sex twin	mixed-sex twin	same-sex twin	mixed-sex twin	same-sex twin	mixed-sex twin
No. children	-0.194***	-0.210***	-0.271***	-0.210***	-0.402***	-0.393***
	(0.003)	(0.012)	(0.002)	(0.005)	(0.004)	(0.004)
N	609332	607574	148765	148507	148587	148329

Note: Two sub-samples are used to identify the economies of scale channel. The same-sex twin sample consists of singleton boy(girl) and boy-boy(girl-girl) twinning at first birth. The mixed-sex twin sample is comprised of singleton boy(girl) and boy-girl twinning at first birth. Robust standard errors are reported in parentheses.

Table 10: Testing the Internal Validity of Instruments — Son Preferences

Panel a: education outcome by number of boys and family size

	No. boys			
		Primary enrolment		
No. births	0	1	2	3
1	0.822	0.847	.	.
2	0.878	0.885	0.887	.
3	0.838	0.853	0.858	0.865
		primary graduation		
1	0.879	0.858	.	.
2	0.841	0.807	0.791	.
3	0.767	0.732	0.719	0.715
		Junior enrolment		
1	0.814	0.792	.	.
2	0.728	0.683	0.679	.
3	0.587	0.539	0.538	0.548

Panel b: no-first-stage

	Primary enrol	Primary grad	Junior enrol
		Tight spacing	
girl at first birth	-0.027***	-0.018***	-0.085***
	(0.001)	(0.004)	(0.004)
N	142381	40431	40394
		Young mother	
girl at first birth	-0.041***	-0.047***	-0.118***
	(0.002)	(0.004)	(0.004)
N	121039	47722	47662
		Ethnic minorities	
girl at first birth	-0.059***	-0.055***	-0.082***
	(0.003)	(0.008)	(0.008)
N	41656	10966	10809

Note: In panel a, average education outcomes of children are listed by number of boys and family size.
In panel b, sample is restricted to families with tightly spaced births, or households with young mothers, or families with Non-Han ethnicity(exempted from the One Child Policy). Robust standard errors are reported in parentheses.

Table 11: Robustness Check: Heterogeneity—Q-Q Trade-off by Mother's Education

	Primary Enrolement		Primary graduation		Junior Enrolment	
	low-edu	high-edu	low-edu	high-edu	low-edu	high-edu
Panel a: 1+ sample						
IV: girl at first birth						
No. children	-0.124***	-0.060***	-0.111***	0.032***	-0.343***	-0.018
	(0.004)	(0.007)	(0.006)	(0.012)	(0.009)	(0.013)
N	321624	107496	115425	22135	115259	22135
IV: twins at first birth						
No. children	-0.065***	-0.052***	-0.087	-0.012	-0.096***	-0.037
	(0.005)	(0.008)	(0.067)	(0.012)	(0.034)	(0.031)
N	323856	109326	115867	22359	115701	22359
Panel b: 2+ sample						
IV: twins at second birth						
No. children	-0.042***	-0.044***	-0.065***	-0.003	-0.069***	-0.019
	(0.006)	(0.008)	(0.014)	(0.013)	(0.016)	(0.016)
N	321624	107496	115425	22135	115259	22135
Panel c: 3+ sample						
IV: twins at third birth						
No. children	-0.040	0.000	-0.002	-0.089*	0.003	-0.116**
	(0.027)	(0.025)	(0.067)	(0.048)	(0.026)	(0.059)
N	175914	32424	175914	3454	175378	3454

Note: High educated mother is defined as completed primary school. Robust standard errors are reported in parentheses and are clustered by household ID for 3+ sample.

Table 12: Robustness Check: Heterogeneity—Q-Q Trade-off in terms of Rural-Urban Gap

	Primary Enrolement		Primary graduation		Junior Enrolment	
	rural	urban	rural	urban	rural	urban
Panel a: 1+ sample						
IV: girl at first birth						
No. children	-0.131***	-0.096***	-0.136***	0.027*	-0.402***	0.018
	(0.002)	(0.009)	(0.009)	(0.015)	(0.006)	(0.015)
N	466549	139873	118935	29368	118935	29368
IV: twins at first birth						
No. children	-0.210***	-0.084***	-0.301***	-0.065***	-0.339***	-0.092***
	(0.004)	(0.009)	(0.003)	(0.009)	(0.010)	(0.011)
N	469207	141277	119345	29624	119345	29624
Panel b: 2+ sample						
IV: twins at second birth						
No. children	-0.082***	-0.018*	-0.110***	-0.021***	-0.098***	-0.048***
	(0.005)	(0.011)	(0.014)	(0.008)	(0.017)	(0.014)
N	392259	36861	114267	23293	114267	23293
Panel c: 3+ sample						
IV: twins at third birth – average outcome of first two births						
No. children	-0.054***	-0.061***	-0.015	-0.073	-0.053	-0.029
	(0.012)	(0.021)	(0.030)	(0.066)	(0.037)	(0.069)
N	200862	7476	30166	3584	30166	3584
IV: twins at third birth						
1st birth						
No. children	-0.081***	-0.064**	-0.008	-0.032	0.020	0.002
	(0.011)	(0.032)	(0.021)	(0.031)	(0.022)	(0.040)
N	148754	4939	64529	5206	64529	5206
2nd birth						
No. children	-0.079***	-0.048***	-0.051**	-0.015	-0.039	-0.017
	(0.010)	(0.017)	(0.022)	(0.028)	(0.025)	(0.033)
N	155822	7629	51932	6209	51932	6209

Note: Robust standard errors are reported in parentheses and are clustered by household ID for 3+ sample.

Table 13: Robustness Check: Heterogeneity—Q-Q Trade-off by Gender of the Extra Sibling

	Primary Enrolement		Primary graduation		Junior Enrolment	
	boy	girl	boy	girl	boy	girl
Panel a: 2+ sample						
IV: twins at second birth						
No. children	-0.077***	-0.073***	-0.096***	-0.093***	-0.098***	-0.095***
	(0.006)	(0.004)	(0.013)	(0.010)	(0.014)	(0.014)
N	428547	428559	137443	137432	137277	137266
Panel b: 3+ sample						
IV: twins at third birth						
No. children	-0.052***	-0.053***	-0.010	-0.004	-0.049	-0.038
	(0.009)	(0.009)	(0.026)	(0.025)	(0.032)	(0.031)
N	208168	208146	33720	33706	33678	33664

Note: Robust standard errors are reported in parentheses and are clustered by household ID for 3+ sample.

Table 14: Robustness Check: Heterogeneity—Q-Q Trade-off by Gender of Older Sibling

	Primary Enrolement		Primary graduation		Junior Enrolment	
	male	female	male	female	male	female
Panel a: 2+ sample						
IV: twins at second birth						
No. children	-0.023***	-0.060***	-0.023	-0.040***	-0.008	-0.010
	(0.006)	(0.005)	(0.024)	(0.015)	(0.014)	(0.018)
N	204453	224667	68618	68942	68536	68858
Panel b: 3+ sample						
IV: twins at third birth						
No. children	-0.008	-0.035**	-0.093	0.010	-0.100**	-0.009
	(0.018)	(0.017)	(0.406)	(0.044)	(0.051)	(0.047)
N	40416	80346	7076	10606	7058	10616

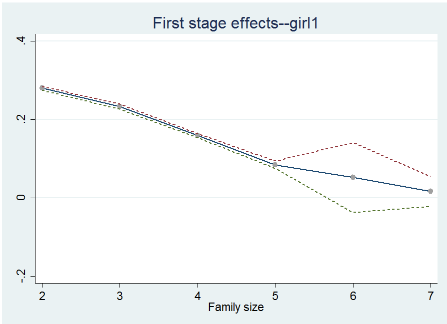
Note: Robust standard errors are reported in parentheses and are clustered by household ID for 3+ sample.

Table 15: Robustness Check: Heterogeneity—the Effect of Family Size on Education Outcome (1982 Census)

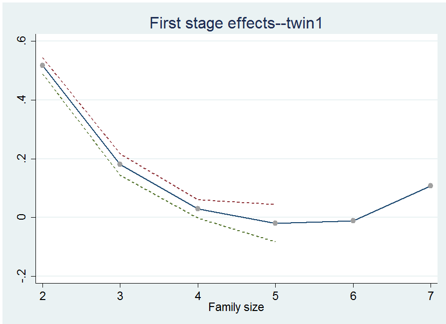
	First stage	Primary graduation OLS	2SLS[a]	NLIV[b]	GMM[c]
Panel a: 2+ sample					
IV: twins at second birth					
twin2	0.533***				
	(0.032)				
No. childnum		-0.023***	0.007	-0.213***	-0.041***
		-0.025[d]	**0.008**	**-0.236**	**-0.045**
		(0.001)	(0.022)	(0.013)	(0.010)
Wald test	280.160				
N		104487	104487	104487	104487
Panel b: 3+ sample					
IV: twins at third birth					
1st born					
twin3	0.576***				
	(0.025)				
No. childnum		-0.032***	0.015	-0.091***	-0.019**
		-0.036	**0.017**	**-0.102**	**-0.021**
		(0.001)	(0.017)	(0.010)	(0.008)
Wald test	531.440				
N		90233	90233	90233	90021
2nd born					
twin3	0.605***				
	(0.035)				
No. childnum		-0.033***	-0.005	-0.130***	-0.023*
		-0.037	**-0.006**	**-0.146**	**-0.026**
		(0.001)	(0.022)	(0.013)	(0.013)
Wald test	299.29				
N		76734	76734	76734	76734

Note: Robust standard errors are reported in parentheses.
[a] 2SLS gives the two stage least square estimates. [b] NLIV gives the estimates based on non-linear first-stage. [c]GMM gives the result taking into account non-linear distribution in both first and second stage.
[d] Numbers in bold gives the corresponding transformation of the magnitude of Q-Q trade-off, i.e. **-0.025** means that an extra kid decreases the probability of primary enrolment by 2.5%, which corresponds to 2.3 percentage points decrease in the outcome variable compared to its mean value(0.023/0.904). The same calculation applies to all the estimates.

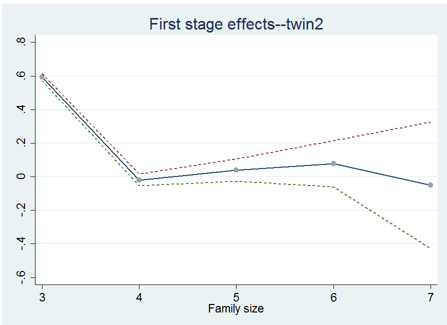
Figures



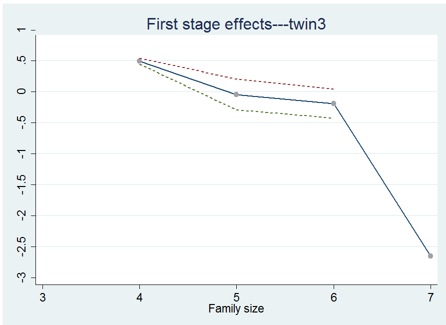
(a) Girl at first birth



(b) Twins at first birth



(c) Twins at second birth



(d) Twins at third birth

Figure 1: First Stage Effects on Fertility

Note: It plots the first stage effects of different instrumental variables on fertility, with number of children on x-axis and the first stage effects on y-axis. For example, the first dot in panel (a) tells that a girl at first birth increases the probability of fertility rises from 1 to 2 by 0.29. The dashed lines are estimated confidence intervals, note that at higher fertility range, the confidence interval is not shown due to the lose of significant power.

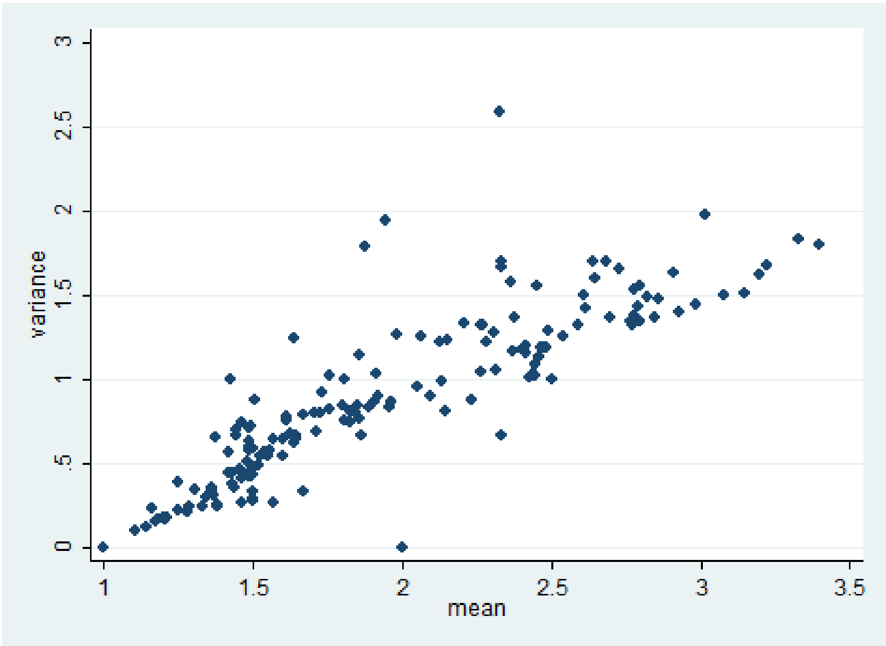
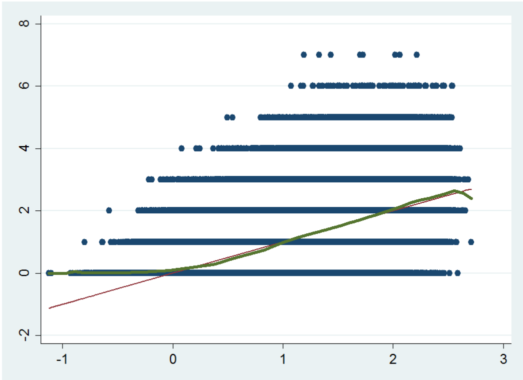
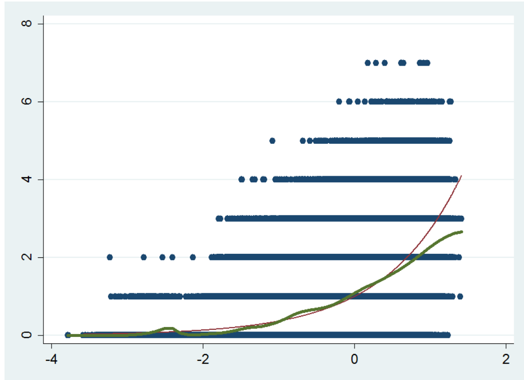


Figure 2: The Mean-variance Relationship for Families with Children

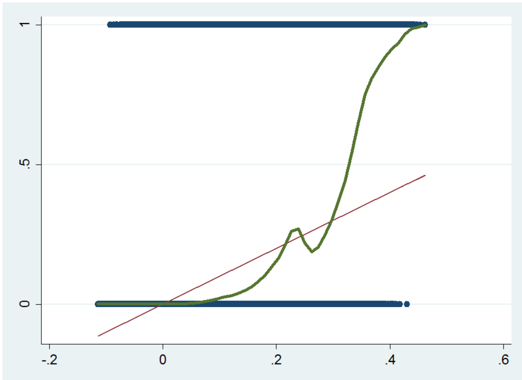
Note: It plots the variance against the mean conditional on the education level of mother and age of giving first birth.



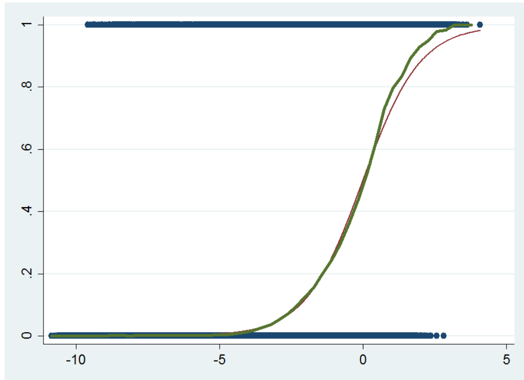
(a) first stage given by OLS



(b) first stage given by Poisson



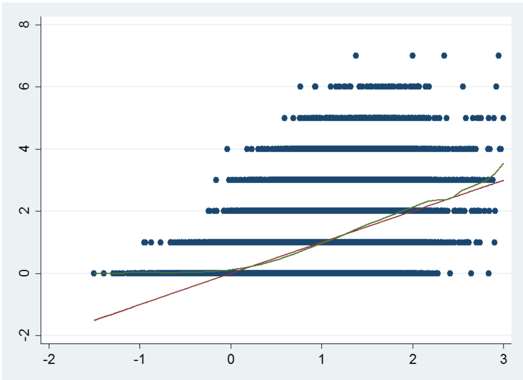
(c) second stage given by OLS



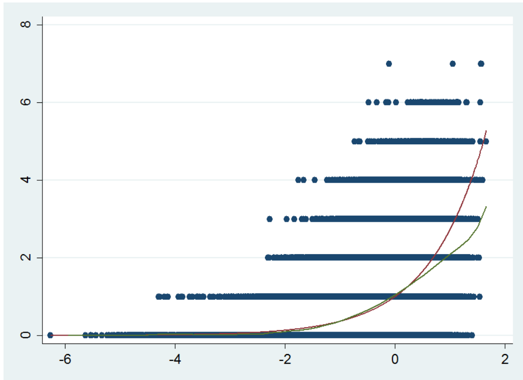
(d) second stage given by Logit

Figure 3: Model Fitting in Two Stages 1/2

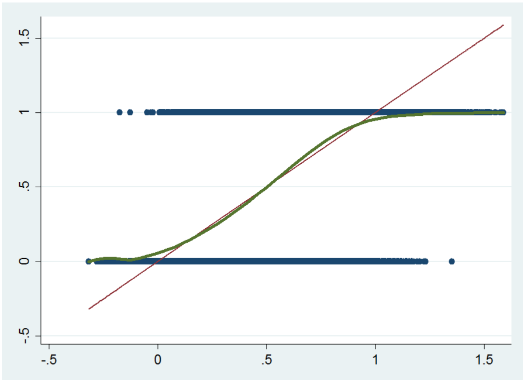
Note: Sample: at least 1 birth household; Endogenous variable: fertility; Outcome variable: the probability of completing primary school; Instrumental variable: gender of the first birth. The dots plot dependent variable Yi or Ni versus estimated linear indexes. The red line shows parametric fit of dependent variable $E(Yi|Xi)$ or $E(Ni|Xi)$ versus estimated linear indexes. The green thicker line represents a kernel regression of Yi or Ni on the fitted index for each model.



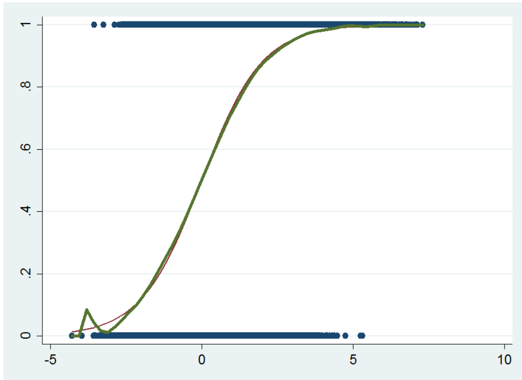
(a) first stage given by OLS



(b) first stage given by Poisson



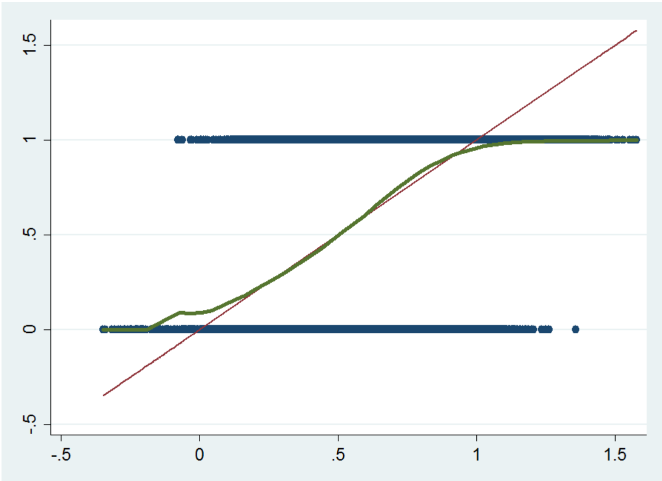
(c) second stage given by OLS



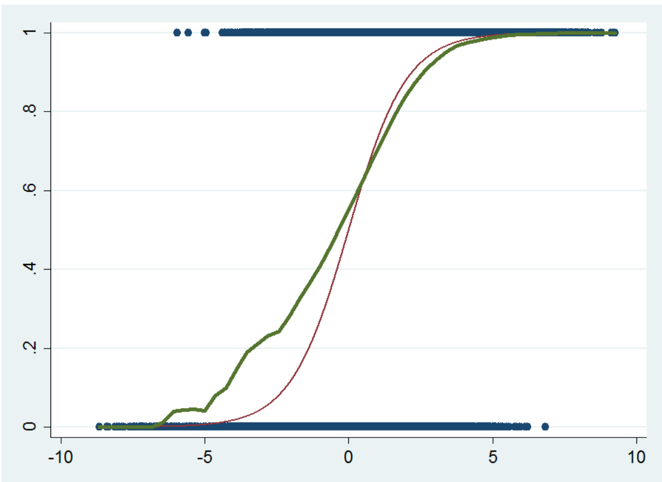
(d) second stage given by Logit

Figure 4: Model Fitting in Two Stages 2/2

Note: Sample: at least 2 birth households; Endogenous variable: fertility; Outcome variable: the probability of entering junior high school; Instrumental variable: twins at the second birth. The dots plot dependent variable Yi or Ni versus estimated linear indexes. The red line shows parametric fit of dependent variable $E(Yi|Xi)$ or $E(Ni|Xi)$ versus estimated linear indexes. The green thicker line represents a kernel regression of Yi or Ni on the fitted index for each model.



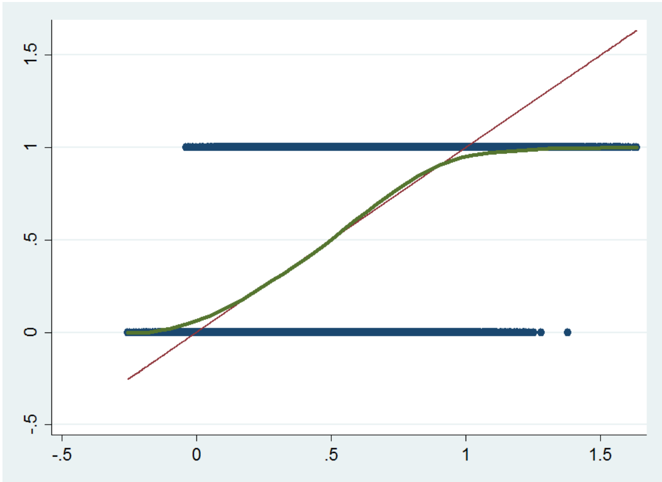
(a) 2SLS estimation



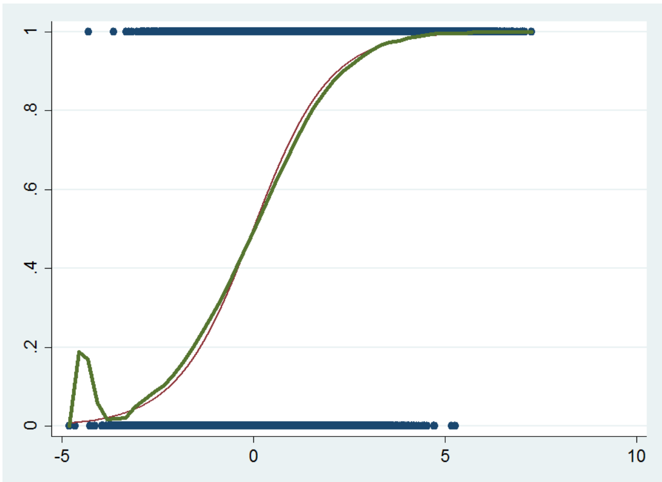
(b) GMM estimation

Figure 5: Model Fitting in One Step 1/2—2SLS VS. GMM

Note: Sample: at least 1 birth household; Endogenous variable: fertility; Outcome variable: the probability of completing primary school; Instrumental variable: gender of the first birth. The dots plot dependent variable Yi versus estimated linear indexes. The red line shows parametric fit of dependent variable $E(Yi|Xi)$ versus estimated linear indexes. The green thicker line represents a kernel regression of Yi on the fitted index for each model.



(a) 2SLS estimation



(b) GMM estimation

Figure 6: Model Fitting in One Step 2/2—2SLS VS. GMM

Note: Sample: at least 2 birth households; Endogenous variable: fertility; Outcome variable: the probability of entering junior high school; Instrumental variable: twins at the second birth. The dots plot dependent variable Yi versus estimated linear indexes. The red line shows parametric fit of dependent variable $E(Yi|Xi)$ versus estimated linear indexes. The green thicker line represents a kernel regression of Yi on the fitted index for each model.