

IZA DP No. 1323

**Sibling Size and Investment in Children's
Education: An Asian Instrument**

Jungmin Lee

September 2004

Sibling Size and Investment in Children's Education: An Asian Instrument

Jungmin Lee

*University of Arkansas-Fayetteville
and IZA Bonn*

Discussion Paper No. 1323
September 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Sibling Size and Investment in Children's Education: An Asian Instrument*

This study consistently estimates the trade-off between child quantity and quality by exploiting exogenous variation in fertility due to son preferences. Under son preferences, childbearing and fertility timing are determined conditional on the first child's gender. For the sample of South Korean households I find strong evidence of unobserved heterogeneity across households. However, sibling size has adverse effects on per-child investment in education, in particular when fertility is high.

JEL Classification: D13, J13, O12

Keywords: fertility, education, son preference

Jungmin Lee
Department of Economics
University of Arkansas-Fayetteville
Austin, TX 78712
USA
Email: JLee@walton.uark.edu

* I would like to thank Stephen Bronars, Robert Crosnoe, Gordon Dahl, Dan Hamermesh, Gerald Oettinger, Steve Trejo, and seminar participants at the University of Texas at Austin and at the 2003 annual meetings of the Society of Labor Economists.

1 Introduction

High fertility has been pointed out as a major obstacle to improving human capital in developing countries, such as African and southern Asian countries. For example, the World Bank (1994) emphasizes the importance of population policy for poverty reduction and economic growth. The underlying reasoning is very straightforward from the perspective of endogenous economic growth models; the quality of labor force is essential for economic development and, given the scarcity of resource, we can educate each worker better when there are fewer.

The so-called “trade-off between child quantity and quality” is an empirical regularity on cross-country data. For example, figure 1 shows that population growth and per-child investment in education are negatively correlated. The horizontal axis represents the change in log fertility rate over the 1960-1980 period. The vertical axis represents the change in log per-child investment in education over the 1975-1995 period. Per-child investment is the national expenditure in education divided by the number of children aged 0-14. To see the effect of fertility on educational investment, two variables are spaced by 15 years. The graph seems to suggest that slowing down population growth should be a high priority in poor countries. Successful examples are East Asian countries, such as South Korea, Hong Kong, and China. They have dramatically restrained population growth and, at the same time, increased educational investment.

We can also observe the trade-off at the household level. Most of sociological studies succeed in finding that sibling size exerts a negative effect on each child’s educational attainments such as grade completion and test scores, after controlling for parents’ socioeconomic

status. For example, Blake (1989) finds across various samples what is called the “dilution effect”; more siblings dilutes a child’s allocation of parental resources.¹

Economists, however, cast doubt on the empirical regularity in that child quantity and quality should be simultaneously determined by parents.² The observed relationship may be spurious (perhaps exaggerated) due to unobserved heterogeneity across households. One source of heterogeneity is that households differ in their *preferences for child quality*. Those with relatively stronger preferences for child quality would have less children to educate each better rather than have many mediocre ones. As a result, an inverse association is observed, but it simply reflects different preferences across households. To avoid the simultaneity bias, Rosenzweig and Wolpin (1980) and Rosenzweig and Schultz (1987) exploit exogenous variation in fertility due to twinning.

In this study I attempt to estimate consistently the trade-off between child quantity and quality for the sample of South Korean households. South Korea must provide an interesting context in which to examine the current issue. As one of “Asian Tigers,” it has been a showcase for developing countries as a fast track to economic development. As

¹Interestingly, Gomes (1984) finds for the sample of Kenya that children from larger family are more likely to complete grades, because parents in Kenya control their eldest child’s earnings and younger children are benefited from this extra source of family resource. This clearly shows that we need to understand the underlying intrahousehold resource allocation mechanism to find the causal effect of sibling size on children’s educational outcomes.

²On the other hand, economists have tried to develop theoretical explanations for its existence. The major novelty here is that, even without assuming unusual substitutability between child quantity and quality in preferences, the trade-off may exist due to their *interaction* in budget constraint (Becker 1961, Becker and Willis 1971).

seen in figure 1, one might argue that a major factor behind its economic success should be population policy. The government established *Planned Parenthood Federation of Korea* in 1961, which enforced four 5-year plans from 1962 to 1981. It implemented various policies, such as diffusing contraceptive measures, deregulating abortions, and campaigning small-size family norms. The basic question here is how much the decreases in fertility due to these policies can explain the increases in educational investment and, eventually, the economic growth. It is essential to know, rather than recognize a simple correlation, the extent to which high fertility rate adversely affects investment in children's education at the micro level.

There are two primary contributions of this study. First, I use parents' monetary investment in children's education as a quantitative measure of child quality, rather than use such variables as children's schooling, test scores, or earnings, which previous studies have used. It should be a good measure of child quality because it is an important input for educational production. Also, it is more relevant to the economic theory of the trade-off between child quantity and quality. The theory is about parents' choices. It is therefore more appropriate to examine directly parental decisions rather than outcomes. The previous measures (for example, grade completion or earnings) are the outcome variables, which are realized far after parents decide child quantity. They are likely to be tainted with various noisy factors that change child quality but have nothing to do with parents' choice or intention. To test for the theory of the trade-off, it seems more reasonable to examine parents' decisions about child quality at the moment that they decide child quantity.

The second contribution is the use of a new instrumental variable for fertility—the first child's sex. It is well known that people in many Asian countries, including South Korea,

prefer sons to daughters. Preferences for sons lead to some empirical regularities in fertility at the household level, one of which turns out to be very useful for this study. Consider a couple that had the first child and is about to decide whether they have the second child or not. Suppose that they want to minimize the number of children, but prefer to have at least one son. In this case, their decision about the second childbearing depends on the first child's gender. If the first child is not a son, they will try another child. This simple example shows that the first child's sex is a reasonable candidate of the instrumental variable for the actual number of children in the household, if it is uncorrelated with the dependent variable, i.e. parental investment in children's education.³ I call my instrument an Asian instrument since son preferences are quite common in Asian countries, such as China and India.

2 First Child's Sex as the Instrument for Fertility

2.1 Fertility-Stopping Rule

The first child's sex is strongly correlated with the actual number of children in the household. According to the demography literature, there exists a specific fertility-stopping rule under son preferences; parity progression depends not only on the gap between desired and current number of total children, but also on the gap between parents' preferred and present number

³The empirical strategy of this study is motivated by Iacovou (1996) and Angrist and Evans (1998). They exploit preferences for a balanced sex composition of children in Western countries to get exogenous variation in fertility. Both studies find that parents with same-sex children are more likely to have additional children. While they have to focus on the marginal effect of the third child, this study can estimate the marginal effect of the second child. But remember that my approach is not applicable to their samples.

of *sons*. For example, if a couple desired to have one son but failed to have one until now, they should try another child as long as there are not many children. They could have finished fertility if they had a son at the first attempt. The first child's sex predicts well the likelihood of having a second child. The first child's sex is also a good predictor for the existence of a third child because those who had a daughter as their first child are more likely to have two daughters consecutively and to have a third child. In general, those who had a girl at first are more likely to end up with more children than those who had a boy.

The predictive power of the first child's sex, which is a critical qualification for an instrumental variable, depends on both the strength of son preferences and the social norm about family size. Table 1 presents some numerical examples showing that the first child's sex is strongly correlated with the number of children in the household. We assume that parents prefer a smaller number of children *if other things are equal*, while each household has a specific range of the desirable number of children.

The first panel shows all the possible paths of parity progression for the households that desire one child at least and two children at most (type-1 preferences). There are only three outcomes, B , GB , and GG . It is easy to show that the first child's sex is significantly correlated with the actual number of children that households end up with. By the assumption that parents prefer smaller number of children, those who had a boy (B) at first should stop fertility, while those who had a girl (G) should go on. Assume that the probability of having a son is 0.51 and constant over parity, according to the demography literature. The expected number of children for those who had a girl ($E(n|G)$) is greater than the expected number of children for those who had a boy ($E(n|B)$). $E(n|G) = 2 > 1 = E(n|B)$. Assuming that the entire population has type-1 preferences, the average number of

children per household is 1.49, and the marginal effect of having a girl at first (*First Girl*) on fertility is 1.

Now suppose that households prefer to have at least one child and can have three at maximum (type-2 preferences). There are four possibilities in parity progression, B , GB , GGG , and GGB . Parents who had a girl as their first child should end up with more children than those who had a boy because:

$$\begin{aligned} E(n|G) &= 0.51 \times E(n|GB) + 0.49 \times E(n|GG) \\ &= 2.49 > 1 = E(n|B). \end{aligned}$$

Again assuming that the population consists of only those with type-2 preferences, the average number of children per household is 1.73, and the marginal effect of having a first girl on fertility is 1.49. For those who prefer to have at least two and at most three children (type-3 preferences), it is also true that the event of *First Girl* increases the number of children:

$$\begin{aligned} E(n|G) &= 0.51 \times E(n|GB) + 0.49 \times E(n|GG) \\ &= 2.49 > 2 = E(n|B). \end{aligned}$$

In sum, *First Girl* has positive impacts on the number of children in the household regardless of different preferences for family size. It is also true when households prefer to have more than one son. For those who prefer two sons rather than one son (type-4 preferences), the first child's sex is still correlated with child quantity. Those who had a girl at first have 3 children, while those who had a boy have 2.49 on average.

The examples show that *First Girl* leads to larger sibling size under son preferences. The marginal effect would increase as more parents prefer sons, they have larger desired number

of sons, and they are willing to accept more children. Comparing the outcomes of type-2 and type-3 preferences groups, we conclude that the marginal effect is larger as parents minimize the number of children (or small-size family becomes a social norm).

In sum, the results here show that our instrument should be strongly correlated with the endogenous variable (child quantity). It is competent to serve as an instrument for fertility and most likely to pass the chronic problem of “weak instrument” (Staiger and Stock 1995, Bound, Jaeger, and Baker. 1995).

2.2 Distinguishing Prenatal vs. Postnatal Son Preferences

Now we examine another requirement for consistent instrumental-variable estimation—the “validity” condition. This condition requires that instrumental variables should not *directly* affect the dependent variable of empirical interest (Rosenzweig and Wolpin 2000). For this study, it means that the first child’s sex should not directly affect parents’ investment in children’s education, but only does so indirectly through number of children.

Readers might suspect that investment in children’s education should depend on the first child’s sex, because the latter is correlated with the sex composition of children. Under son preferences, investment in children might depend on the sex composition. If parents prefer sons to daughters, they would invest more in their favored children.

It is, however, important to distinguish prenatal son preference from postnatal son preference. The first means that parents prefer to *have* sons rather than daughters, while the latter implies that they *treat* sons in more favorable ways. The fertility-stopping rule we examined is related to prenatal son preferences. What matters now for the validity condition

is the existence of postnatal son preferences.

Although these two types of son preferences are closely related, they do not necessarily coexist. Parents may prefer to have sons, while they may like to treat them equally once they are born. Fortunately, this is the case in South Korea. There is no strong evidence for the existence of postnatal son preferences. The gender gap in the secondary school enrollment rate has decreased from 15 percent in 1970 to none in 1995 (World Bank 2001). The probability of advancing from high school to college or university is now higher for females (75 percent) than for males (70 percent) in 1995. If there were substantial discrimination against girls, then we should have observed significant differences in educational outcomes between sexes.

Fortunately, even if there existed strong postnatal preferences for sons and, consequently, if the IV estimates were severely biased, we could figure out the direction of bias. They are underestimated, because postnatal preferences for sons suggest that the dependent variable (*total* investment in education) and the instrument (*First Girl*) should be negatively correlated. The true trade-off between child quantity and quality should be weaker than what the estimates imply.⁴

There is another possibility that the instrumental-variable estimations could underestimate the true effect. As Hanushek (1992) pointed out, decreased child spacing acts like

⁴Kim and Lee (2002) find that parents invest in private tutoring more for females. They interpret this finding as a result of the fact that females are more likely to take private tutoring in music and arts, which tend to be more expensive. Their finding suggests that our estimates might be overestimated in absolute term. Thus we experiment with a different dependent variable, total investment in education *except* private tutoring, which does not significantly change our results.

an increase in sibling size because smaller spacing means 1) lower probability of being in small family and 2) less attention and resources from parents. According to the demography literature, under son preferences, child spacing between the first and second child tends to decrease if the first child is a daughter, because parents would hurry to have another child to have a son. As a result, if we do not control for child spacing (also endogenous), we will underestimate the marginal effects in absolute term. This will strengthen my conclusion that the trade-off between child quantity and quality is not as strong as what observed cross-sectional relationships suggest.

2.3 Sex-Selective Abortion

We have shown that the first child's sex should satisfy both the relevance and validity conditions. One might be curious about why only the first child's sex can be a valid instrumental variable. For example, the sex composition of the first two children should predict number of children better than the first child's sex alone for those who have more than one child. If parents had two daughters consecutively, they are more likely to have another child than those who had two sons or those who had a son and a daughter.

It is because of the existence of sex-selective abortion that we do not include the second child's sex as an additional instrumental variable. Sex-selective abortions are quite common in South Korea, in particular in 1990s including the sample period between 1993 and 1998. Table 2 displays the sex ratio at birth (the number of boys divided by the number of girls times 100) by birth order between 1981 and 2000. If children's gender were randomly assigned by nature and infant mortality rate were constant over birth order and between sexes, then

the sex ratio must stay around the normal level (about 106), regardless of birth order. However, table 2 shows that the sex ratio gets imbalanced rapidly as the parity is higher. The increases in the sex ratio over the birth order imply that parents not only prefer to have sons, but also prefer to have small number of children. Thus, female births after a certain number of children are suppressed by sex-selective abortions (Park and Cho 1995).

It is surprising that the imbalance of the sex ratio has been recorded even at the birth order two since the mid-1980s. The second child's sex is not randomly assigned by nature, but chosen by parents. As a result, the second child's sex might be correlated with investment in education on the unobservable. Suppose that those who strongly prefer son and therefore are more likely to abort girls tend to invest more in children's education. If preferences for abortion are somehow related to preferences for child quality, the second child's sex cannot be a valid instrumental variable.

On the other hand, the sex ratio at the birth order one has remained around the normal level during the 1990s. Parents leave the sex of the first child to natural selection. The first child's sex is randomly assigned by nature and orthogonal to parents' preferences.

3 Empirical Analysis and Results

3.1 Data and Descriptive Statistics

The data used in this study are the Korean Household Panel Study (KHPS). The KHPS is the first panel study on South Korean households, conducted by a private institution, Daewoo Economic Research Institute. Our sample spans the entire period over which the

data are available between 1993 and 1998.

The sample is restricted by the following steps. First, I restrict it to those households with at least one child to implement the instrumental-variable estimation method. This does not cause any serious sample selection problem since almost every household has at least one child at some point in the course of life. I also exclude those with more than three children. Very few observations (38 observations, 0.73 percent of the sample size) are excluded, so the problem of sample selection is ignorable. Second, I focus on those households with children in schooling, in particular on those with the first child under age 19. So we estimate the effects of fertility on investment in children's education under the age of college/university entrance. Lastly, those households with missing or unreliable information about crucial variables are deleted. The final sample is an unbalanced panel data set, consisting of 5,180 observations on 1,663 households over 6 years.

Table 3 shows descriptive statistics of crucial variables. It is notable that fertility is quite low, comparable to the United States and even the developed countries in Europe. This low fertility rate is in part because those households included in the sample are quite young and have not completed fertility. However, the average number of births per fertile woman (age 15-49) is only 1.42 in 1998 according to the national statistics based on birth registration. The low fertility rate is inconsistent with the general prediction in the demography literature that son preferences tend to increase fertility because parents continue to have children until they achieve their desirable number of sons. The coexistence of son preferences and low fertility clearly indicates widespread sex-selective abortions (Park and Cho 1995).

The most crucial variable in table 3 is investment in children's education. It measures total monthly household expenditure in such items as tuition, school books and supplies,

reference books, private tutoring, and other school expenses. It is notable that average investment in education increases in trend, except in 1998 when the economy suffered in the midst of the Asian financial crisis. The proportion of investment in total household income steadily increases from 7.7 percent in 1993 to 12.7 percent in 1997 and, despite the crisis, again to 13.1 percent in 1998. Children's education not only accounts for a large portion of household expenditure, but takes a high priority in household resource allocation.

3.2 Testing for the Existence of Son Preferences

There are various reasons for preferences for sons in South Korea. First, sons, in particular the eldest sons, are required to carry traditional family responsibilities, such as serving ancestor worship rites and supporting parents in their old ages.⁵ And preferences for sons are reinforced to exist due to economic reasons. Sons would increase parents' lifetime income more than do daughters given that there exists large gender gap in earnings in the labor market. The gender gap in South Korea is highest among the OECD countries during the mid and late 1990s; in terms of median full-time earnings, females earn less on average by 41 percent than do males.

I first test for the existence of son preference by estimating parity progression models, widely used in the demography literature. The hypothesis to be tested is that, if there were son preferences, those who had daughters in earlier phases of fertility should end up with larger number of children. It is important to note that parity progression models must be tested for those who have completed fertility since households' decision on further

⁵According to Korea Social Survey in 1998, almost 30 percent of people think sons, especially the eldest son, are responsible for their old parents, while only 0.5 percent of people think daughters are so.

childbearing should be fully observed (not right-censored). We assume that mothers aged 40 and older have completed fertility.

Following the convention in the literature, we estimate the logit models in which the dependent variable is the indicators of parity progression (for example, $I[n > 1]$ indicates whether the household has more than one child or not). The key independent variable is the sex composition of children under a certain birth order. Table 4 shows the results. Marginal effects are calculated. In the first column, we find that having a girl as the first child increases the probability of having a second child (9.8 percent). In the second column, having two same-sex children increases the probability of having three or more children (9.9 percent). This positive effect turns out to be entirely due to a large positive effect of having two girls (34.8 percent), when it is entered separately along with the indicator for having two boys in the third column. In fact, having two boys has a negative effect on fertility, implying that parents with two sons are more likely to stop childbearing compared to the reference group of parents of one boy and one girl. South Korean parents do not have preferences for a balanced sex mix of children, which Iacovou (1996) and Angrist and Evans (1998) are based on.

We can also test for the existence of son preferences by hazard models of fertility timing. The hypothesis here is that, given the limited period of fertility, parents should advance the timing of the second birth after the first daughter than they do after the first son. By the same logic, they should try to have a third child sooner, if they had two daughters consecutively. The dependent variable is now the time interval between births. An advantage of hazard models over parity progression models is that we do not have to restrict the sample.

I estimate Weibull hazard models, taking account of right-censoring of birth interval.

Only time-invariant variables, such as parents' education levels and mother's age at first birth, are included to avoid any bias that arises because explanatory variables change during the interval. Relative hazard ratios are calculated in table 5. As expected, having a girl as the first child significantly increases the hazard rate of having a second child by at least 48 percent. Any positive effect of *Same Sex* disappears once we include *BB* and *GG* separately, as with the results of parity progression model. The coefficient of *BB* is estimated less than one, which means that having two boys decreases the hazard rate of having a third child. This implies that parents are likely to delay additional childbearing—probably to delay for good—if they got two boys. On the other hand, *GG* significantly advances the third childbirth.

3.3 Marginal Effects of Fertility on Investment in Education

The purpose of this study is to estimate the causal effect of exogenous change in number of children on investment in children's education.⁶ Suppose that the government succeeds in suppressing population growth by some policies that do not directly affect household investment in children's education. It might tax childbirth or just legally forbid having more than a certain number of children, like China's one-child policy. The question is how the policy-induced changes in child quantity (fertility) affect child quality (investment in education).

First, we start with estimating a simple linear equation:

$$\ln I_{it} = \alpha n_{it} + \mathbf{X}_{it}\beta + u_{1it} \quad (1)$$

$$\text{where } n_{it} = \mathbf{X}_{it}\gamma + \mathbf{Z}_{it}\delta + u_{2it} \quad (2)$$

⁶Strictly speaking, our empirical strategy of instrumental variable allows us to estimate only the effect of *increase* in fertility on the investment. We assume that the effect is symmetric.

where the error term u_1 is assumed to be independent of all the control variables in \mathbf{X} , such as mother's age, parents' years of schooling, natural logarithm of monthly total household income, indicator of living with grandparents, dummies for residential location, and four yearly dummies. Since educational costs differ by levels of schooling, we include average age of children and its square term, as suggested by Browning (1992).⁷ n represents the number of children in the household.

I represents monthly total investment in education for all children in the household. Formally, I is equal to $p_q(n \cdot q)$, where q is the average quality of children and p_q is the unit price of child quality (Becker and Willis 1973). By the Slutsky theorem and product rule, the coefficient α represents:

$$\begin{aligned} \text{plim}(\alpha) &= \text{plim} \frac{\partial \ln p_q n q}{\partial n} = \text{plim} \frac{1}{n} + \frac{\partial q}{\partial n} \frac{1}{q} \\ &= \frac{1}{E(n)} \cdot (1 + \epsilon_{qn}), \end{aligned}$$

where ϵ_{qn} represents the elasticity of child quality with respect to child quantity. It is therefore possible to recover the elasticity from the consistent estimate of α and the sample mean of n .

Equation (2) specifies the first-stage regression in which fertility is endogenously determined by exogenous explanatory variables (\mathbf{X}) and instrumental variables (\mathbf{Z}) excluded from the second-stage regression. Besides the first child's sex (*First Girl*), there are four additional instrumental variables, including its interactions with some exogenous variables in \mathbf{X} .

First, we expect that the interaction of *First Girl* and *Living with Grandparents* increases

⁷Including time-invariant random effect which is orthogonal to explanatory variables does not make any significant difference.

fertility, but does not directly affect total investment in education. When the household head is the eldest son, he is strongly asked to have at least one son to carry the responsibility for continuing family lineage. Thus those households are more likely to have a second child once they had a girl at first. Unfortunately, there is no direct information about the birth order of household head. But, the indicator of living together with grandparents represents well whether the head is eldest or not. It requires a strong assumption that living arrangement should be exogenous to household decisions about investment for children. But this assumption seems reasonable in the context of South Korea where living with old grandparents is socially enforced.

We also use the quadratic terms of mother's age interacted with *First Girl* as additional instruments. The rationale is that age profile of fertility would differ by the first child's sex, as the results from hazard models imply that those who had a daughter at first are likely to have more children at certain age than those who had a son. Lastly, we include the interaction of mother's education and *First Girl* because lower-educated mothers put greater importance on traditional values, including son preferences. We assume that, after being interacted with the first child's sex, mother's education does not affect total investment in children's education.

Table 6 presents the estimation results for the linear specification. The ordinary least squares (OLS) estimates show that an additional child increases total investment by about 27 percent. If there were no adverse effect of fertility on per-child investment, then the estimate should equal one. The OLS estimate therefore implies the existence of the trade-off between child quantity and quality. The elasticity of per-child investment with respect to fertility is negative (-0.487).

The instrumental-variable (IV) estimates show that the OLS estimates are underestimated by 6 to 10 percentage points. Unobserved heterogeneity accounts for some part of the observed trade-off. The elasticity calculated based on the IV estimates ranges -0.291 to -0.369, smaller in absolute term than that of the OLS estimate. But they are still negative.

The instrumental variables in the first-stage regression show significant explanatory power. In column (2) *First Girl* decreases the probability of having an additional child by 17.7 percent. The partial-R squared amounts to 0.0328, about 30 percent of total R-squared. *F*-statistic for the instrument is 174.8, much larger than the rule-of-thumb threshold of 10 (Bound, Jaeger, and Baker 1995). In column (4) the partial-R squared for multiple instruments is similar, and *F*-statistic is 38.7. The advantage of having multiple instruments is that we can test for the validity of instruments by testing over-identifying restrictions. The results indicate that we cannot reject the validity of instruments (*p*-value for Hansen's *J*-statistic test is 0.53).

We relax the linearity of fertility in equation (1) by substituting n with two dummy variables; $D_2 = 1$ if there are exactly two children in the household and 0 otherwise: $D_3 = 1$ if there are exactly three children and 0 otherwise. Recall that the maximum number of children is three in the sample. We estimate:

$$\ln I_{it} = \alpha_2 D_{2,it} + \alpha_3 D_{3,it} + \mathbf{X}_{it}\beta + u_{1it} \quad (3)$$

$$\text{where } D_{2,it} = \mathbf{X}_{it}\gamma_2 + \mathbf{Z}_{it}\delta_2 + u_{2it} \quad (4)$$

$$D_{3,it} = \mathbf{X}_{it}\gamma_3 + \mathbf{Z}_{it}\delta_3 + u_{3it} \quad (5)$$

The coefficients, α_2 and α_3 , represent the marginal effects of the second and third child, respectively. If there were no trade-off, then we should have $\alpha_2 = 1$ and $\alpha_3 = 2$. Now that

there are two endogenous explanatory variables, we must include at least two instrumental variables for identification.

Table 7 presents the estimation results. Again the OLS estimates imply the existence of the trade-off. I calculate the marginal effects on per-child investment from the coefficients by taking $\frac{1+\alpha_n}{n}$ where $n = 2$ or 3 . For the households with two children per-child investment is only 67.7 percent of the amount of investment that a child could have received if there were only one child. For those with three children it is much smaller, 49.4 percent of that of the reference group.

The OLS estimates exaggerate the trade-off because they ignore the endogeneity of fertility. According to the IV estimates, per-child investment for the households with two children is 74.6 percent, and that for those with three children is 57.6 percent of what a child could have received if he or she were only child. The results suggest that the trade-off exists and, as number of children increases, the trade-off gets stronger. However the adverse effect of fertility on per-child investment is not as large as what the OLS estimates suggest. The comparison is summarized in figure 2.

Furthermore, recall that the IV estimates for α_n and thereby for per-child investment ($\frac{1+\alpha_n}{n}$) are likely to be biased downward because of postnatal son preferences and endogenous birth intervals, as mentioned in the previous section. Thus the true trade-off is likely to be even weaker than what the IV estimates suggest.

Some results in first-stage regressions in columns (2) and (3) are worth noting. As expected, the interacted terms between *First Girl* and *Grandparents* in the first-stage regressions have significantly-positive coefficients. The results are consistent with the prior expectation that those households whose head is the first son have the familial responsibility

of having at least one son.

There are clearly different age profiles of fertility by the first child's sex. The age profiles implied by the estimates are depicted in figure 3. The first graph shows the marginal effects of *First Girl* on the age profile of the probability of having *exactly* two children. The reference group is those with a son as the first child. The effects are graphed separately for three different education groups (middle-school graduates, high-school graduates, and college/university graduates). The age range for each educational group is chosen to guarantee that at least 10 observations are available at each age to avoid small-sample bias.

It is notable that the marginal effects are positive in the beginning and then decrease in age for all three groups. Positive signs are reasonable because having a girl as the first child would increase the likelihood of having a second child. Evaluated at the average age of the second childbirth for each group, the marginal effects are significantly positive. For middle-school graduates, the marginal effect is about 4 to 10 percent at age 26.6, while it is similarly 5 to 9 percent for high-school graduates at age 27.6. The average age of second childbirth is highest (29-30) for college/university graduates. The marginal effect evaluated at the average age is on the other hand lowest, 4 to 7 percent. Overall *First Girl* increases the probability of having two children by 4-10 percent. The age profiles are consistent with the fact that the higher educated women delay childbearing. And they prefer sons to daughters, but their preferences are weaker.

The figure also shows that the marginal effect is decreasing and becomes even negative, which seems unreasonable. However recall that we estimate the probability of having *exactly* two children. As a result, the marginal effect decreases as some of those who have a daughter at first fail to have a son again at the second attempt and now have a third child.

In particular, for middle-school and high-school graduates, the marginal effects are significantly negative after certain ages (29 for middle-school and 30 for high-school graduates). This suggests that, for these low-educated women, having a daughter at first increases the likelihood of having three or more children relatively, compared to the reference group of households who have a son as their first child.

The bottom graph in figure 3 shows the marginal effects of *First Girl* on the age profile of the probability of having *exactly* three children. The profile for college/university graduates is not drawn because there are few observations. This is not surprising in that few college-educated women have three children (total 35 observations). Contrary to the previous graph, the marginal effects increase in age. The increasing profiles indicate that having a third child happens later in age. For middle-school graduates, *First Girl* increases the probability of having three children by 9 to 11 percent at the age 28.6, which is the average age when those women have a third child. The corresponding effect is 7 to 8 percent for high-school graduates at their average age 29.9. The results confirm that the higher-educated women postpone childbearing and they have weaker preferences for sons.

Some results in column (4) in table 7 are worth noting. First, investment in children's education is a normal good, and the income elasticity is quite inelastic, about 0.4. It confirms that children's education is a priority in household expenditure. Those who live in metropolitan areas, especially in the capital Seoul, invest significantly more in children, while they have smaller number of children. Those in metropolitan areas invest more by about 5 percent, and those living in Seoul invest additionally more by 8 to 9 percent. This might reflect stronger preferences for child quality in these regions or more investment opportunities in cities, such as private tutoring and academic institutes. Not surprisingly, the higher-educated parents

tend to invest more in children's education. Investment is more responsive to mother's education rather than father's education, which is consistent with the notion that child quality depends more on the mother-side characteristics. The higher-education mothers have smaller number of children, but they invest more in children's education. On the other hand, father's education also increases investment, while it tends to increase number of children.⁸

4 Summary

This study examines the effects of fertility on parental investment in children's education to test for the existence of trade-off between child quantity and quality. The main contribution is that this study exploits exogenous variation in fertility due to son preferences to identify the causal effects of child quantity on quality. A consistent estimate of the effect of fertility on total investment in children's education is important for policy makers in developing countries. It is important to know the extent to which population policy aiming at lowering fertility rate is effective in fostering human capital investment in future labor force.

The results for South Korean households suggest that the endogeneity of fertility plays a significant role in the observed inverse relationship between child quantity and quality. The observed trade-off exaggerates the true relationship. However, there still exists the trade-off in a significant way. Furthermore, as number of children increases, the trade-

⁸I experiment with a different instrument—the occurrence of twinning at the first birth. The validity condition here seems questionable since the costs of educating children would differ for two singletons and twins. Having two children at the same time may decrease per-child investment, in particular when the household faces borrowing constraints. The estimate turns out to be significantly negative. The sign is very difficult to justify by economic theories other than econometric bias.

off gets stronger. Restricting population growth would be effective in increasing per-child investment in education, especially when fertility is high.

The findings suggest that successful population policies in South Korea can explain in part why per-child investment in education has increased in past decades. However the magnitude of the causal effect of fertility on educational investment seems too small. Per-child investment in education in terms of the proportion of household expenditure has increased by four times between 1981 and 2001. The fertility rate has decreased from 2.6 to 1.3, but the decrease seems too small to explain the increase in investment based on our estimates. There must exist other factors that decrease fertility rate and, at the same time, increase investment in education—for example, increasing parental concern about children’s education or increases in the returns to human capital associated with technical changes.

References

- [1] Angrist, Joshua and Evans, William. 1998. “Children and Their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size.” *American Economic Review*, 88(3):450-477.
- [2] Becker, Gary and Lewis, Gregg. 1973. “Interaction between Quantity and Quality of Children.” *Journal of Political Economy*, 81(2):S279-S288.
- [3] Blake, Judith. 1989. *Family Size and Achievement*. Los Angeles, CA: University of California Press.

- [4] Bound, John, Jaeger, David and Baker, Regina. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*, 90(430):443-450.
- [5] Browning, Martin. 1992. "Children and Household Economic Behavior." *Journal of Economic Literature*, 30(3):1434-1475.
- [6] Chun, Hyunbae and Oh, Jeungil. 2002. "An Instrumental Variable Estimate of the Effect of Fertility on the Labor Force Participation of Married Women." *Applied Economics Letters*, 9(10):631-634.
- [7] Gomes, Melba. 1984. "Family Size and Educational Attainment in Kenya." *Population and Development Review*, 10(4):647-660.
- [8] Hanushek, Eric. 1992. "The Trade-off between Child Quantity and Quality." *Journal of Political Economy*, 100(1):84-117.
- [9] Haveman, Robert and Wolfe, Barbara. 1995. "The Determinants of Children's Attainments: A Review of Methods and Findings." *Journal of Economic Literature*, 33(4):1829-1878.
- [10] Iacovou, Maria. 1996. "Fertility and Female Labor Force Participation." *Working Paper*, University College London.
- [11] Kim, Sunwoong and Lee, Ju-Ho. 2002. "Private Tutoring and Demand for Education in South Korea." *mimeo*, 2002.

- [12] Park, Chai Bin and Cho, Nam-Hoon. 1995. "Consequences of Son Preference in a Low-Fertility Society: Imbalance of the Sex Ratio at Birth in Korea." *Population and Development Review*, 21(1):59-84.
- [13] Rosenzweig, Mark and Schultz, T. Paul. 1987. "Fertility and Investments in Human Capital: Estimates of the Consequence of Imperfect Fertility Control in Malaysia." *Journal of Econometrics*, 36:163-184.
- [14] Rosenzweig, Mark and Wolpin, Kenneth. 2000. "Natural "Natural Experiments" in Economics." *Journal of Economic Literature*, 38(4):827-874.
- [15] Rosenzweig, Mark and Wolpin, Kenneth. 1980. "Testing the Quantity-Quality Fertility Model: the Use of Twins as a Natural Experiment." *Econometrica*, 48:227-240.
- [16] Staiger, D. and Stock, J.H. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65(3):557-586.
- [17] World Bank. 1994. *Population and Development: Implications for the World Bank.*, Washington D.C..

Table 1. First Child's Sex and Actual Number of Children

<i>Fertility Preferences</i>	<i>Desired Number of Children</i>		<i>Parity Progression</i>	<i>Probability</i>
	<i>Min</i>	<i>Max</i>		
Type-1	1	2	<i>B</i>	0.51
	1	2	<i>GB</i>	0.49 X 0.51
	1	2	<i>GG</i>	0.49 X 0.49
Type-2	1	3	<i>B</i>	0.51
	1	3	<i>GB</i>	0.49 X 0.51
	1	3	<i>GGB</i>	0.49 X 0.49 X 0.51
	1	3	<i>GGG</i>	0.49 X 0.49 X 0.49
Type-3	2	3	<i>BB</i>	0.51 X 0.51
	2	3	<i>BG</i>	0.51 X 0.49
	2	3	<i>GB</i>	0.49 X 0.51
	2	3	<i>GGB</i>	0.49 X 0.49 X 0.51
	2	3	<i>GGG</i>	0.49 X 0.49 X 0.49
Type-4	2	3	<i>BB</i>	0.51 X 0.51
	2	3	<i>BGB</i>	0.51 X 0.49 X 0.51
	2	3	<i>BGG</i>	0.51 X 0.49 X 0.49
	2	3	<i>GBB</i>	0.49 X 0.51 X 0.51
	2	3	<i>GBG</i>	0.49 X 0.51 X 0.49
	2	3	<i>GGB</i>	0.49 X 0.49 X 0.51
	2	3	<i>GGG</i>	0.49 X 0.49 X 0.49

Table 2. Sex Ratio at Birth by Birth Order

<i>Birth Order</i>	<i>One</i>	<i>Two</i>	<i>Three</i>	<i>Four and Higher</i>	<i>Total</i>
1981	106.3	106.7	107.1	112.9	107.2
1982	105.4	106.0	109.2	113.6	106.8
1983	105.8	106.2	111.8	120.0	107.4
1984	106.1	107.2	116.9	128.1	108.3
1985	106.0	107.8	129.2	146.8	109.4
1986	107.3	111.2	138.6	149.9	111.7
1987	104.7	109.1	134.9	148.8	108.8
1988	107.2	113.2	164.4	181.7	113.3
1989	104.1	112.4	181.4	198.7	111.7
1990	108.5	117.0	188.8	209.2	116.5
1991	105.7	112.4	179.5	193.5	112.4
1992	106.2	112.4	191.1	212.1	113.6
1993	106.4	114.7	202.1	234.8	115.3
1994	106.0	114.1	202.2	223.5	115.2
1995	105.8	111.7	177.2	203.7	113.2
1996	105.3	109.8	164.0	183.4	111.6
1997	105.1	106.3	133.5	153.5	108.2
1998	105.9	108.0	144.7	153.2	110.2
1999	105.6	107.6	141.8	153.8	109.6
2000	106.2	107.4	141.7	166.9	110.2

* Korea National Statistical Office, Vital Statistics

Table 3. Descriptive Statistics

<i>Year</i>	<i>1993</i>	<i>1994</i>	<i>1995</i>	<i>1996</i>	<i>1997</i>	<i>1998</i>
Fathers						
Age	36.42 (4.26)	36.87 (4.41)	37.11 (4.40)	37.73 (4.59)	38.22 (4.63)	38.82 (4.71)
Schooling	12.46 (2.78)	12.44 (2.64)	12.64 (2.60)	12.66 (2.62)	12.73 (2.63)	12.78 (2.58)
Mothers						
Age	33.12 (3.65)	33.53 (3.81)	33.79 (3.86)	34.38 (4.03)	34.91 (4.13)	35.46 (4.25)
Schooling	11.46 (2.43)	11.54 (2.32)	11.76 (2.33)	11.83 (2.40)	11.93 (2.35)	12.02 (2.26)
Households						
Seoul	0.31 (0.46)	0.33 (0.47)	0.32 (0.47)	0.31 (0.46)	0.32 (0.47)	0.28 (0.45)
Large Cities	0.28 (0.45)	0.27 (0.44)	0.25 (0.43)	0.26 (0.44)	0.25 (0.44)	0.28 (0.45)
Grandparents	0.07 (0.25)	0.08 (0.28)	0.07 (0.26)	0.07 (0.25)	0.08 (0.27)	0.09 (0.29)
Mean Child Age	7.09 (3.63)	7.37 (3.80)	7.45 (3.78)	7.97 (3.89)	8.37 (3.98)	8.81 (4.04)
Middle School	0.08 (0.21)	0.11 (0.25)	0.11 (0.25)	0.13 (0.26)	0.13 (0.27)	0.14 (0.27)
High School	0.02 (0.12)	0.03 (0.13)	0.03 (0.15)	0.05 (0.16)	0.07 (0.20)	0.09 (0.24)
Number of Children	1.91 (0.52)	1.93 (0.51)	1.92 (0.50)	1.93 (0.48)	1.92 (0.50)	1.95 (0.50)
Monthly Investment in Education	0.18 (0.12)	0.20 (0.14)	0.24 (0.16)	0.27 (0.18)	0.29 (0.18)	0.24 (0.16)
Monthly Household Income	2.00 (1.16)	2.13 (1.15)	2.35 (1.19)	2.59 (1.32)	2.60 (1.15)	2.12 (1.14)
<i>N</i> =	1,136	967	874	849	800	554

* Monthly Investment in Education and Monthly Household Income are denoted in 100,000 Korean won and real value.

Table 4. Logit Models of Parity Progression

<i>Dependent Variable</i>	<i>I[n > 1]</i>	<i>I[n > 2]</i>	<i>I[n > 2]</i>
First Girl	0.098 (0.019)	0.286 (0.035)	0.036 (0.046)
Second Girl		0.251 (0.033)	
Same Sex		0.099 (0.035)	
<i>BB</i>			-0.151 (0.044)
<i>GG</i>			0.349 (0.045)
Pseudo R squared	0.158	0.151	0.151
<i>N</i> =	4,259	3,373	3,373

* Robust standard errors are displayed in parentheses. Equations include Mother Age, Mother Age squared, Mother Age at First Birth, Grandparents, Seoul, and Large Cities.

Table 5. Weibull Hazard Models of Childbirth Spacing

<i>Dependent Variable</i>	<i>AGE1 - AGE2</i>	<i>AGE2 - AGE3</i>	<i>AGE2 - AGE3</i>
First Girl	1.476 (0.111)	2.559 (0.504)	0.704 (0.211)
Second Girl		3.452 (0.693)	
Same Sex		2.450 (0.479)	
<i>BB</i>			0.710 (0.227)
<i>GG</i>			8.457 (1.972)
Pseudo R squared			
<i>N</i> =	7,149	5,363	5,363

* Hazard ratios are presented. Robust standard errors are displayed in parentheses. Equations include Mother Age, Mother Age at First Birth, and Mother Education. AGE_i represents the age of child of birth order *i*.

Table 6. Linear Model of Investment in Children's Education

	(1)	(2)	(3)	(4)	(5)
	<i>OLS</i>	<i>IV</i>		<i>IV</i>	
<i>Regression Stage</i>		<i>First</i>	<i>Second</i>	<i>First</i>	<i>Second</i>
<i>Dependent Variable</i>	<i>ln I</i>	<i>n</i>	<i>ln I</i>	<i>n</i>	<i>ln I</i>
Number of Children	0.267 (0.017)		0.369 (0.093)		0.328 (0.088)
Household Income	0.422 (0.022)	0.061 (0.016)	0.416 (0.023)	0.060 (0.016)	0.419 (0.022)
Seoul	0.075 (0.020)	-0.118 (0.016)	0.088 (0.023)	-0.116 (0.016)	0.083 (0.023)
Large Cities	0.044 (0.020)	-0.032 (0.016)	0.047 (0.020)	-0.031 (0.016)	0.046 (0.020)
Grandparents	0.001 (0.031)	0.080 (0.025)	-0.006 (0.032)	0.029 (0.033)	-0.003 (0.032)
Father Education	0.013 (0.004)	0.002 (0.003)	0.013 (0.004)	0.002 (0.003)	0.013 (0.004)
Mother Education	0.031 (0.005)	-0.027 (0.004)	0.034 (0.006)	-0.022 (0.005)	0.033 (0.006)
Mother Age	0.051 (0.033)	0.141 (0.026)	0.038 (0.035)	0.160 (0.035)	0.044 (0.035)
Mother Age squared	-0.001 (0.000)	-0.002 (0.000)	0.000 (0.000)	-0.002 (0.001)	-0.001 (0.000)
Mean Child Age	0.278 (0.018)	0.099 (0.012)	0.268 (0.020)	0.098 (0.012)	0.272 (0.020)
Mean Child Age squared	-0.013 (0.001)	-0.008 (0.001)	-0.012 (0.001)	-0.008 (0.001)	-0.012 (0.001)
Middle School	0.262 (0.064)	0.461 (0.053)	0.214 (0.078)	0.460 (0.052)	0.233 (0.076)
High School	0.641 (0.119)	0.881 (0.093)	0.551 (0.144)	0.880 (0.093)	0.587 (0.142)
Year 1994	0.033 (0.026)	0.010 (0.021)	0.032 (0.026)	0.010 (0.021)	0.033 (0.026)
Year 1995	0.142 (0.027)	-0.013 (0.022)	0.144 (0.027)	-0.013 (0.022)	0.143 (0.027)
Year 1996	0.187 (0.027)	-0.012 (0.022)	0.188 (0.027)	-0.011 (0.022)	0.187 (0.027)
Year 1997	0.252 (0.028)	-0.032 (0.023)	0.255 (0.028)	-0.030 (0.023)	0.254 (0.028)
Year 1998	0.038 (0.033)	0.017 (0.025)	0.037 (0.033)	0.019 (0.025)	0.037 (0.033)

Table 6. (Continued)

<i>Regression Stage Dependent Variable</i>	<i>OLS</i>	<i>IV</i>		<i>IV</i>	
	<i>ln I</i>	<i>First n</i>	<i>Second ln I</i>	<i>First n</i>	<i>Second ln I</i>
First Girl		0.177 (0.013)		0.573 (0.789)	
X Grandparents				0.125 (0.051)	
X Mother Education				-0.013 (0.006)	
X Mother Age				-0.022 (0.046)	
X Mother Age squared				0.000 (0.001)	
Constant	-5.483 (0.559)	-0.701 (0.437)	-5.439 (0.561)	-1.022 (0.578)	-5.456 (0.560)
R squared	0.424	0.104	0.420	0.106	0.423
Partial R squared		0.033		0.036	
F test		p < 0.01		p < 0.01	
Overidentification test				p = 0.53	
N =	5,180	5,180	5,180	5,180	5,180

Table 7. Nonlinear Model of Investment in Children's Education

	(1)	(2)	(3)	(4)
	<i>OLS</i>	<i>IV</i>		
<i>Regression Stage</i>		<i>First 1</i>	<i>First 2</i>	<i>Second</i>
<i>Dependent Variable</i>	<i>ln I</i>	<i>D₂</i>	<i>D₃</i>	<i>ln I</i>
Two Children	0.354 (0.024)			0.492 (0.282)
Three Children	0.481 (0.034)			0.727 (0.212)
Household Income	0.423 (0.022)	0.003 (0.015)	0.029 (0.009)	0.416 (0.023)
Seoul	0.071 (0.020)	0.004 (0.014)	-0.060 (0.009)	0.086 (0.023)
Large Cities	0.043 (0.020)	0.003 (0.015)	-0.017 (0.009)	0.046 (0.020)
Grandparents	0.001 (0.031)	-0.032 (0.030)	0.030 (0.019)	-0.008 (0.032)
Father Education	0.013 (0.004)	-0.005 (0.003)	0.004 (0.002)	0.013 (0.004)
Mother Education	0.030 (0.005)	-0.003 (0.004)	-0.009 (0.003)	0.033 (0.006)
Mother Age	0.041 (0.034)	0.216 (0.032)	-0.028 (0.020)	0.024 (0.046)
Mother Age squared	0.000 (0.000)	-0.003 (0.000)	0.000 (0.000)	0.000 (0.001)
Mean Child Age	0.281 (0.018)	-0.001 (0.011)	0.049 (0.007)	0.269 (0.021)
Mean Child Age squared	-0.013 (0.001)	0.000 (0.001)	-0.004 (0.000)	-0.012 (0.001)
Middle School	0.267 (0.064)	0.072 (0.048)	0.194 (0.030)	0.208 (0.087)
High School	0.659 (0.118)	0.069 (0.085)	0.406 (0.053)	0.549 (0.157)
Year 1994	0.032 (0.026)	0.014 (0.019)	-0.002 (0.012)	0.030 (0.026)
Year 1995	0.139 (0.027)	0.024 (0.020)	-0.018 (0.012)	0.141 (0.028)
Year 1996	0.181 (0.027)	0.041 (0.020)	-0.026 (0.013)	0.182 (0.028)
Year 1997	0.248 (0.028)	0.020 (0.021)	-0.025 (0.013)	0.252 (0.028)
Year 1998	0.034 (0.033)	0.037 (0.023)	-0.009 (0.014)	0.032 (0.034)

Table 7. (Continued)

<i>Regression Stage Dependent Variable</i>	<i>OLS</i>	<i>IV</i>		
	<i>ln I</i>	<i>First 1 D₂</i>	<i>First 2 D₃</i>	<i>Second ln I</i>
First Girl		3.486 (0.721)	-1.457 (0.453)	
X Grandparents		0.107 (0.046)	0.009 (0.029)	
X Mother Education		0.018 (0.005)	-0.015 (0.003)	
X Mother Age		-0.219 (0.042)	0.098 (0.027)	
X Mother Age squared		0.003 (0.001)	-0.001 (0.000)	
Constant	-5.115 (0.561)	-2.793 (0.529)	0.386 (0.332)	-4.919 (0.642)
R squared	0.428	0.015	0.098	0.422
Partial R squared		0.012	0.049	
F test		12.35	53.08	
Overidentification test		p < 0.01	p < 0.01	p = 0.39
N =	5,180	5,180	5,180	5,180

Figure 1. The Trade-off between Child Quantity and Quality across Countries

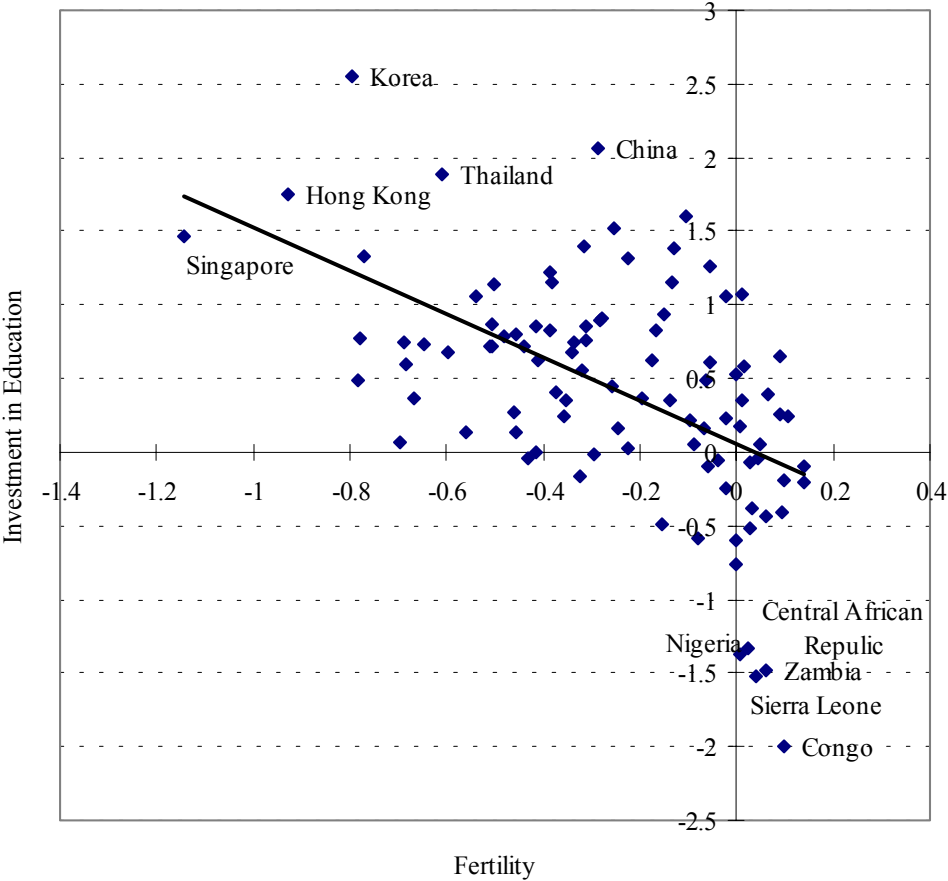


Figure 2. Per-child Investment in Education by Sibling Size

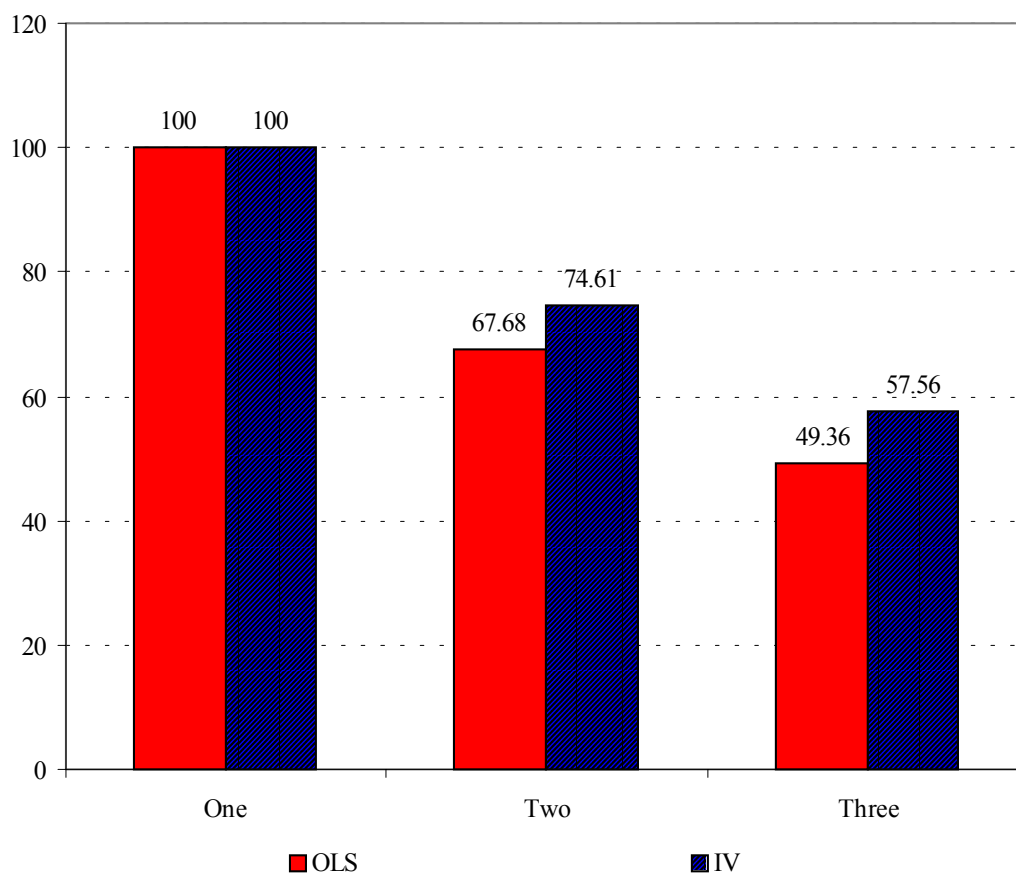
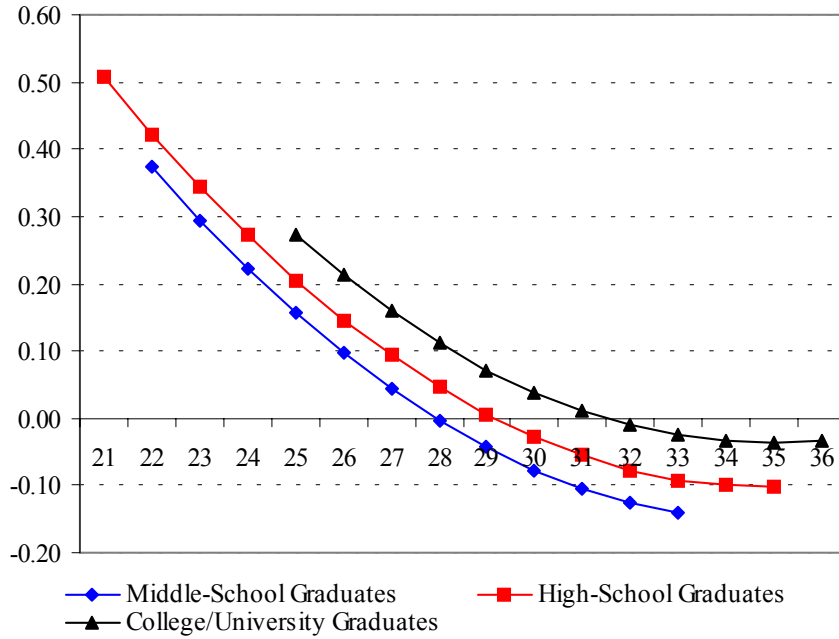


Figure 3. Age Profile of Fertility by First Child's Sex

The Marginal Effects of First Daughter
on the Probability of Having Two Children over Mother's Age



The Marginal Effects of First Daughter
on the Probability of Having Three Children over Mother's Age

