2008

Birth Spacing Effect on Children's Attainments: Indentification Using Instrument Variables

Jing XIE
Singapore Management University, jing.xie.2007@me.smu.edu.sg

BIRTH SPACING EFFECT ON CHILDREN'S ATTAINMENTS:

INDENTIFICATION USING INSTRUMENT VARIABLES

XIE JING

SINGAPORE MANAGEMENT UNIVERSITY
2008

BIRTH SPACING EFFECT ON CHILDREN'S ATTAINMENTS:

INDENTIFICATION USING INSTRUMENT VARIABLES

XIE JING

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE IN ECONOMICS

SINGAPORE MANAGEMENT UNIVERSITY
2008

Birth Spacing Effect on Children's Attainments: Identification using Instrument Variables

XIE JING

Abstract

In this study, I address the relationship between an often overlooked dimension of family structure—the spacing between children's births—and the degree of children's attainments such as Mathematics, Reading Cognition and Reading Comprehension. Comparing to the results of OLS estimation, 2SLS Estimation using Twin and Catholic as Instrument Variables shows less significant effects on children's attainments. Hausman Test shows that OLS estimators are not consistent with 2SLS estimators, which means there is endogenous problem in OLS estimation. As the result in 2SLS shows the different spacing effects in different spacing groups, it is possible to use nonlinear estimation (quadratic form of birth spacing) and semi-parametric estimation to draw the curve of birth spacing effects, and find the most efficient birth spacing, golden birth spacing. These two estimations, to a large extent, match each other in the range of golden birth spacing.

Contents:

Acknowledgement

I would like to express my gratitude to all those who help me complete this thesis. I want to thank the School of Economics in SMU for giving me the opportunity of starting this thesis in the first place. I am deeply indebted to my advisor Professor Fali Huang who leads me to the world of labor economics, conveys me bountiful research methodologies and encouraged me to go ahead with my thesis. I am also grateful to Professor Peter C.B. Phillips and Professor M. J. Lee, who give a lot of useful suggestions on this thesis. And also I am bound to appreciate Miss Sun Li for helping me a lot in Stata programming and data analysis.

Many classmates from the School of Economics supported me in my research work. I want to thank them for all their help, support, interest and valuable hints. Especially I am obliged to Wei Lin, Baoling Ma, and Leming Lin.

Introduction:

What affects children's attainments has been a central concern of economists and sociologists. Many element that affect children attainments are picked out and get the certain answer of their effects [(Haveman, Wolfe (1995)]. However, there are only a few relevant studies on the children's birth spacing. With highly reliable contraceptive techniques, couples now can plan their family more precisely including how many children they want and when they give birth to them. Correspondingly, the average family size is now quite small and the decline in family size allows for greater variation in how children are spaced, thus making the issue of birth spacing more meaningful. Moreover, high divorce, remarriage, and reconstitution rates produce a situation conducive to unconventional spacing practices. As the previously married remarry, bring their children into new family unites and potentially have additional children, the prospects for every close or very wide birth spacing between sibling sets rise. In addition, the unprecedented influx of women who defer marriage and childbearing may constrain spacing decisions. Although couples who marry at younger ages tend to space their children closely together (Bumpass, Rindfuss & Janosik 1987), closer spacing may also be the most sensible option for women who cannot afford to leave the labor force to care for children (Newman 1983; Wineberg &McCarthy 1989). With this ongoing metamorphosis in the American family, the consequences of child spacing invite greater attention.

Despite impressive data collected on the antecedents of spacing, we know little about its effects. Because of the difficulty on how to define birth spacing (when sibling size is large than 2, the birth spacing is complicated) and the endogeneity (parents can choose when to have the baby exactly) in birth spacing, there are only a few papers study about birth spacing effects on children's attainments and even few of them concern the endogeneity in birth spacing. In this paper, I focus on birth spacing effects on children's attainments, finding proper instrument variables for birth spacing and try to simulate the most efficient birth spacing for children's attainments.

Birth spacing is relevant to the resource dilution hypothesis (Rosenzweig, 1986). If children are born closely together, their parents are hindered from devoting their undivided attention to them, whether at play or furnishing intellectual stimulation. Wider spacing permits parents more "breathing room". This may be especially important during early childhood since these years rearing responsibilities are labor intensive. Close spacing may also strongly diminish financial assets to recover from economic setbacks much harder than wider spacing. It may be easier for family to pay for the expenses incurred by children with wider birth spacing. To illustrate, sending three children to nursery school would be more arduous if the children were each spaced one year as opposed to three years apart. Thus, children spaced shorter may be less likely to be favored with respect to the cultural, social, and economic resources that boost intellectual growth. And also, according to Zajonc and Markus (1975), the closer children are spaced, the lower the quality of the intellectual climate to which they are exposed. That is, to the extent the intellectual milieu is defined as the unweighted average of the intellectual levels of family members, the larger the birth spacing, the higher the intellectual atmosphere.

But on the other hand, a long birth spacing means parent take a longer time to nurture the same amount of children than parents with near-spaced children. These parents have a long time in the "labor intensive period" from the first child to the last and this long period also has negative effects on parents' energy and economic source. This kind of parents cannot get out of nurturing dependent children for a longer time, especially when they enter into mid-ages, which is the indispensable period in their own career development. The negative effects on parents, in reverse, harm children's attainments from the decline of economic resource.

Therefore, the specific objective of this research is to see which part, positive or negative, dominates the birth spacing effects on children's attainments.

Data Description

How to define birth spacing is a complicated problem. For example, if sibling size is 3, the first child has 2 different "birth spacing"—one is age interval with the second child and the other one is age interval with the third child. Here, we assume that nearest sibling affects children most. To simplify analysis, assume the age interval between the child and his/her nearest sibling represent Birth Spacing of the child.

To test for the birth spacing effects on children's attainments controlled by other family backgrounds, data are needed on parental characteristics, on the spacing of children, and on specific achievements for every child. The National Longitudinal Survey of Youth 1979 (NLSY79) and NLSY Children & Young Adults are two connective databases which can provide the data we need. Education and age of the mother come from NLSY79, and the children information such as number of siblings, age, attainments and race come from NLSY Children & Young Adults. The NLSY Children & Young Adults child sample is comprised of all children born to NLSY79 mothers. With the primary support from the National Institute of Child Health and Human Development (NICHD), the children of the NLSY79 mothers have been independently followed and interviewed in various ways starting in 1986. Birth interval among siblings is calculated from the age, and birth order of the children. To get the biggest size of sample, the NLSY Children & Young Adults in questionnaire 1996 are the best choice.

One critical aspect is children attainment—Peabody Individual Achievement Test (PIAT) results of each child in NLSY79 Child represent the level of children's attainments. PIAT is a wide-range measure of academic achievement for children aged five and over and is widely known and used in research. It is among the most widely used brief assessment of academic achievement having demonstrably high test-retest reliability and concurrent validity. NLSY79 Children Supplement includes three subtests from the full PIAT battery: Mathematics, Reading Recognition and

Reading Comprehension assessments. PIAT scores are norm percentile scores designed for children enrolled in the first third of the kindergarten year—the closest approximation available to ages 60 to 62 months. These three variables in 1996 survey can explain children's early attainments, controlled by other personally elements such as education and age of mother, and birth order, age, gender and race of the child.

From Haveman et al. (1995), family size (number of siblings) is one of critical elements that determine children's attainments. But also, family size is an endogenous variable that can be chosen by parents. To focus on birth spacing effect, I fixed family size in analysis to simplify the problem. In order to get the largest valid sample size, sibling size equal to 3 is chosen as study sample. And the analysis results of sibling size equal to 2 is consistent with that of sibling size equal to 3. There are 3285 observations in NLSY79 Children whose family size is 3, and the Birth Spacing of these children ranges from 0 to 148 months. Here we define that dummy variable Twins is equal to 1 when Birth Spacing less than 10 months. Figure 1 illustrates the histogram of birth spacing in samples.

Figure 1: Histogram of Birth Spacing

By separating the Birth Spacing into several birth spacing groups, the mean values of children attainments have obvious changes. It seems that narrow spaced children have higher Mathematics, Reading Recognition and Reading Comprehension scores than far spaced children have. Table 1 shows some correlation between birth spacing and children's attainments.

Table 1: Summary Statistics of Children's Attainments

OLS and 2SLS Estimations

OLS Estimations:

To maintain consistently with previous research in Powell & Steelman (1990) and Haveman & Wolfe (1995), in data analysis, we divide birth spacing into 3 groups: birth spacing less than 36 months, 36 to 60 months, and larger than 60 months. Moreover, comparing to other division, this division gets the lowest Akaike Information Criterion (AIC)[1] value than others. 3 dummy variables—d1, d2 and d3 are generated to represent birth spacing groups.

$d1 = 1$(when interval is less than 36 months)

$d2 = 1$(when interval is between 36~60 months)

$d3 = 1$(when interval is more than 60 months)

When birth spacing is multiplied by d1, d2 and d3, these three products capture birth spacing in these three segments, and we call them near-spaced, mid-spaced and

[1] Akaike's information criterion, developed by Hirotsugu Akaike under the name of "an information criterion" (AIC) in 1971 and proposed in Akaike (1974), is a measure of the goodness of fit of an estimated statistical model. In the general case, the AIC is $AIC = 2k - 2\ln(L)$, where k is the number of parameters in the statistical model, and L is the maximized value of the likelihood function for the estimated model. Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best.

far-spaced:

near-spaced=birth spacing*d1

mid-spaced=birth spacing*d2

far-spaced=birth spacing*d3

So the OLS estimation is given by:

$$\text{children's attainments}=\alpha+\beta_1\text{near-spaced}+\beta_2\text{mid-spaced}+\beta_3\text{far-spaced}+\gamma X+\varepsilon$$

Where X is other control variables such as age, age square of child, race, birth order, education level of mother, age of mother and gender of child.

Table 2: OLS Estimation of the Effect of Birth Spacing on Children's Attainments

Table 2 presents OLS estimation results when sibling size is 3. The results of OLS estimation Table 2 suggest a significantly positive effect of near-spaced birth interval on all of the children attainments scores. The effects of mid-spaced birth interval are smaller than near-spaced: in the first column, coefficient of mid-spaced birth interval has 95 percent significance, in the second column, mid-spaced coefficient is insignificant, and the third column shows 90 percent significance of mid-spaced coefficient. All effects of the far-spaced birth spacing are near zero and insignificant. As the results, children's attainments increase as birth spacing increase, but the coefficients of birth spacing decline with birth spacing increase—near-spaced birth interval has the largest coefficient, followed by mid-spaced birth spacing, then far-spaced birth spacing. Another aspect in the estimation is that coefficients of birth spacing are larger in Mathematics than in Reading scores. Other control variables also have significant effects on children's attainments: the increase of children's attainments is associated with the increase of education level of the mother, the decrease of the age when mother gave birth to the child, being white, and the decrease of birth order of children.

It seems that this is the end of this story. However, the correlation between birth spacing and children's attainments is only suggestive, because there is an indispensable point we should not omit—fertility such as birth spacing is most likely to be endogenous.

2SLS Estimations

To solve the possible endogeneity in OLS estimation, proper Instrument Variable should be chosen to adjust Birth Spacing variable. In this section, I try to find a proper instrument variable (IV) which is used to identify the casual effect of birth spacing and children's attainments. In particular, we employ 2SLS estimations that are specified by equations (1) and (2), and compare the results with those of the OLS estimates by Hausman Test.

2SLS estimation:

(1) birth spacing=$\delta+\phi$IV+φX +u

near-spaced =birth spacing hat*d1
mid-spaced=birth spacing hat*d2
far-spaced=birth spacing hat*d3

Here birth spacing hat is the predict value of birth spacing from equation (1).

(2)

$$\text{children's attainments}=\alpha+\beta_1\text{fitted near-spaced}+\beta_2\text{fitted mid-spaced}+\beta_3\text{fitted far-spaced}+\gamma X+\varepsilon$$

Possible Instrument Variables for Birth Spacing

Fertility is most likely to be endogenous in analysis of children attainments. Birth spacing—one of the important elements in fertility of course can be chosen by parents. When we study the effects of birth spacing on children attainments, one of the critical steps is to find a proper instrument to explain birth spacing exogenously.

In this paper, I find two possible variables that may explain birth spacing exogenously—being one of twins and membership in Rome Catholic.

Before reporting the estimation results, we first discuss the validity of using these two variables as instrument variables. A good IV should be highly correlated with the birth spacing of children in a household, but should not affect the children’s outcome except through birth spacing. That is to say, a valid IV should not be correlated with unobserved parental and household characteristics that are captured by the error term.

The birth of twins is an important source of exogenous variation in fertility that has been used in previous research (Rosenzweig and Wolpin, 1980), and this variable can exogenously narrow the nearest sibling birth spacing to zero. Also, twin is believed to be unlikely depending on other control variables in this analysis. Although the correlation between twin birth and unobserved household attributes is untestable by design, we follow Black et al. (2005) and examine whether the occurrence of twins is associated with certain observed characteristics, such as the education level of parents. Similar to the findings of Black et al., the F-test based on linear probability model suggests that the probability of having a twin birth is uncorrelated with the education level of the mother's education.

Membership in Rome Catholic is a much controversial instrument variable. In recent years a growing number of demographers have been impressed with the impact of an ideological and organizational factor such as membership in the Rome Catholic Church. Their viewpoint is expressed in the following statement: "One of the striking findings is the persistent importance of religious differentials and in particular the higher fertility and the lesser use of appliance contraceptives by Catholics." Because of the opposition to contraception in Catholic doctrine, many people believe that the fertility rate of people who are the member of Rome Catholics is higher than that of non-Catholics. But with the prevalence of AIDS, population and economic problems let many Rome Catholics turn to use contraception to prevent illnesses and control the family size. Although there are controversies that people who are member of Rome Catholics might use contraception as much as non-Catholics, the beliefs of Catholics has significant effects on birth spacing when controlled with other characteristic variables in the data analysis.

Here, I use 2SLS (two stage least square) estimates birth spacing effects to eliminate endogeneity in estimation.

Table 3: The First Step of the 2SLS.

Table 3 shows the results of the first stage in 2SLS. In Table 3, the first column shows how Twin performs as a significant instrument to explain the birth spacing value. Consistent to the assumption, Twin can significantly explain birth spacing with other control variables, and the level of significance is more than 99 percent. On the other side, the explanation of Catholic on birth spacing has a smaller coefficient and less significant level—90 percent, so does the R square. From the table 3 we can see that Twin has a better explanation power than Catholic. This result coincides with the analysis before.

In the second stage of 2SLS estimation, I split the predicted value of birth spacing estimated in the first stage into 3 segments (the same as in OLS estimation):

fitted near- spaced =fitted age spacing*d1
fitted mid-spaced=fitted age spacing*d2
fitted far-spaced=fitted age spacing hat*d3

Here d1, d2, d3 is defined as before.

Table 4: The Second Step of 2SLS when Twin is IV.

Table 4 shows the results of the second stage of 2SLS when the first stage uses Twin as an instrument variable. Only Mathematics score is significantly affected by near-spaced, mid-spaced and far-spaced birth spacing, and other children attainments such as Reading Cognition and Comprehension are not significantly affected by all of the birth spacing groups. In other words, when Twin is instrument variable, the effects of predicted birth spacing decrease (comparing to the results in OLS estimation). If the hypothesis that Twin is a feasible instrument to eliminate endogeneity holds, then, the exogenous birth spacing effects affect much less on children attainments than previous OLS estimation, especially in reading ability. Meanwhile, other control variables have almost the same significant level of explanation on children attainments as in OLS estimation: children's attainments are positively associated with the education level of mother, being white and negatively associated with the birth order of the child and the age of mother when she gave birth to the children.

Table 5: the Second Step of 2SLS when Catholic is IV.

Table 5 shows the results of the second stage of 2SLS when the first stage uses Catholic as an Instrument Variable. None of coefficients of birth spacing variables are significant and the coefficients are much smaller comparing to those in OLS estimation. This means that when we use beliefs in Catholics as an Instrument Variable to eliminate the endogeneity of birth spacing, the coefficients of birth spacing are much smaller. If the hypothesis that Catholic is a feasible instrument variable holds, then the effects of birth spacing on children attainments is much smaller than OLS estimation predicts.

Hausman Test

The Hausman test is a statistical test in econometrics named after Jerry Hausman. The test evaluates the significance of an estimator versus an alternative estimator. In linear model $y = bX + e$, where y is univariate and X is vector of regressors, b is a vector of coefficients and e is the error term, we have two estimators for b: b_0 and b_1. Under the null hypothesis, both of these estimators are consistent, but b_1 is more efficient (has smaller asymptotic variance) than b_0. Under the alternative hypothesis, one or both of these estimators is inconsistent. Deriving the statistic:

$$H = T(b_0 - b_1)' Var(b_0 - b)^{-1}(b_0 - b)$$

where T is the number of observations. This statistic has chi-square distribution with k (length of b) degrees of freedom. If the null hypothesis is rejected, one or both of the estimators is inconsistent.

This test can be used here to check for endogeneity of a variable by comparing IV estimates to OLS estimates. By using Hausman Test to compare OLS estimation and 2SLS estimation with Twin and Catholics as IV, the null hypothesis, both of these estimators are consistent, are rejected. Therefore, OLS estimation is inconsistent with the estimator of 2SLS and has edogeneity.

Approaching to Golden Birth Spacing

Quadratic form of Birth Spacing in Estimation

Even from Hausman Test, OLS estimation has endogeneity and is inconsistent with 2SLS, both OLS and 2SLS estimation generate same trend of birth spacing effects: the near-spaced birth interval has positive effects on children's attainments, but mid-spaced birth interval has smaller positive or insignificant effects, so does far-spaced birth interval. This intuition means, at some point of birth spacing, its positive effects on children's attainments might disappear. This suggestion arises our interest in finding, at least approximate, "golden" birth spacing such that children's attainments get its peak value.

Nonlinear estimation can picture the curve of birth spacing effects so that the golden birth spacing can be illustrated in the picture. From the intuition before, I introduce quadratic form of birth spacing. Assume the estimation model has quadratic form of Birth Spacing as following:

$$\text{children's attainments} = \alpha + \beta_1 \text{birth spacing square} + \beta_2 \text{birth spacing} + \gamma X + \varepsilon$$

Where X is other control variables such as age, age square of child, race, birth order, education level of mother, age of mother and gender of child.

Table 6: Nonlinear Estimation of Birth Spacing Effects Using Quadratic Form of Birth Spacing

Table 6 shows the results of this nonlinear estimation. The negative sign before quadratic form of Birth Spacing in the first row of this table, even some of them are insignificant, illustrates that it is possible to have a peak value of children's attainments when birth spacing changes and other variables are controlled. From table 6, we can pin down the golden birth spacing by maximizing the estimation of children's attainment.

$$\text{children's attainments}=\beta_1 \text{birth spacing}^\wedge 2+\beta_2 \text{birth spacing}+b$$

Here b is other control variables which are constant here. After knowing β_1 and β_2 from Table 6, we obtain that when birth spacing is 46 months, 38 months, and 25 months, Mathematics, Reading Recognition and Reading Comprehension can get their peak value separately. Even though because of the endogeneity in birth spacing may introduce bias in these values, at least we can get some intuition from these values.

Semi-parametric Estimation

For a more generous case in analysis, we do not assume the functional structure of birth spacing, but just assume birth spacing effect is an unknown functional form in partially linear estimation. Partially linear model specifies the conditional mean to be the usual linear regression function plus an unspecified nonlinear component:

$$E[y|x, z] = f(z) + x*b + e$$

where the scalar function $f(z)$ is unspecified. Here $f(z)$ is a smooth function with bounded first derivatives, the function f is known to lie in a particular parametric family, x are control variables that enter this equation linearily, e is the zero-mean innovation error, and b is a vector of parameters. Using partially linear estimation can create a new variable which contains the smoothed value of function f. These values are estimated by the locally weighted regression using lowess[2] (Locally weighted scatterplot smoothing) estimator.

In this analysis, I assume birth spacing effect is an unknown function as $f(z)$, so the estimation function here is:

$$\text{children's attainments} = f(\text{birth spacing}) + X*b + e$$

Where X is other control variables such as age, age square of child, race, birth order, education level of mother, age of mother and gender of child.

Table 7: Semi-Parametric Estimation

[2] Lowess estimator of Cleveland (1979) is a variant of local polynomial estimation that uses the tricubic kernel $K(z) = (70/81)(1-|z|^3)^3 1(|z|<1)$. For a summery see Fan and Gijbels (1996). Lowess is attractive compared to kernel regression as it uses a variable bandwidth, robustifies against outliers, and uses a loval polynomial estimator to minimize boundary problems.

The results of Partially Linear Estimation are shown in Table 7. The first row in table 7 is P value of function birth spacing in partially linear estimation. From this row, functional form of birth spacing is significant when estimates Mathematics and Reading Comprehension. The most attractive point here is by Lowess estimator, we can generate a new variable which contains the smoothed value of function f. Figure 2,3 and 4 picture these curves of birth spacing function when estimates three different children's attainments, Mathematics, Reading Cognition and Reading Comprehension.

Figure 2: Function of Birth Spacing Effects on Mathematics.

Figure 3: Function of Birth Spacing Effects on Reading Recognition.

Figure 4: Function of Birth Spacing Effects on Reading

From figure 2, 3 and 4, showing by the reference lines, the peak value of Mathematics, Reading Cognition and Reading Comprehension obtained when birth spacing is equal to 38 months, 37 months and 30 months separately. Comparing this conclusion to the results of nonlinear estimation with quadratic form of birth spacing, 46 months, 38 months and 25 months separately, even though these numbers are not exactly the same, they share the same trend—golden birth spacing for Mathematics is larger than that for Reading Recognition, and golden value of birth spacing for Reading comprehension is the lest (Showing in Table 8). And the range of these golden birth spacing is not large, from 28 to 46, comparing to the whole range of sample size.

Even though it might be biased because of endogeneity in birth spacing, we can still get some intuition from the results of quadratic form estimation and partially linear estimation. When children's birth spacing fall into the range from 25 months to 46 months, the possibility of getting high children's attainments score is very large.

Conclusion

The OLS estimation suggests a significantly positive effect of near-spaced birth interval on children's attainments. The effects of mid-spaced birth interval and far-spaced birth interval are much smaller than near-spaced. Increasing birth spacing at first has significant positive effects on children's attainments—larger birth spacing induces a better score, but with the space getting larger and larger, the positive effects tend to be less significant.

To eliminate endogeneity in birth spacing, I use two possible Instrument Variables—Twin and Catholics to explain birth spacing effects on children's attainments. Twin can significantly (99 percent) explain birth spacing with other control variables, and Catholic can explain birth spacing with 90 percent level of significance. When I use 2SLS to estimate birth spacing effects, most of the coefficients of birth spacing variables decrease and tend to be insignificant comparing to the OLS coefficients. Hausman Test is a statistical test in econometrics which is use to evaluate the significance of an estimator versus an alternative estimator. Hausman

Test shows all the OLS estimations are not consistent with the 2SLS estimations, which means the endogenous problem exists in OLS estimation, and 2SLS estimation has less bias.

Both OLS and 2SLS estimation illustrate a decline birth spacing effects with birth spacing decrease, which require further study by nonlinear estimation with quadratic form of birth spacing and partially linear estimations. These two methods help us picture the curve of birth spacing effects—a concave functional curve of birth spacing effects. The results of these two methods match each other and help us pin down, at least a range of, golden birth spacing, in which children can obtain their highest attainments' score.

Some questions in this thesis are left for further study. This golden birth spacing or golden range of birth spacing is just intuition because some of the coefficients of birth spacing in quadratic form are insignificant, and one of functional form of birth spacing is also insignificant. So mathematical calculation from coefficients or Lowess estimators cannot insures the veracity of this result. Moreover, golden birth spacing in this thesis might have bias because of endogeneity in birth spacing. And the conclusion in this thesis just provides some intuition on best birth spacing in family.

Reference:

1. Becker, Gary S. and H. Gregg Lewis (1973), "On the interaction between the quantity and quality of children," *Journal of Political Economy*, 81(2), 279-288.

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

3. Hall, P., J. Kay, and D. Titterington (1990), "Asymptotically optimal difference-based estimation of variance in non-parametric regression," *Biometrica*, 77, 521-528.

4. Haveman, Robert and Barbara Wolfe (1995), "The determinants of children's attainments: a review of methods and findings," *Journal of Economic Literature,* 33 (4) 1829- 1878.

5. Li, Hongbin and Junsen Zhang, (2006), "The Quantity-Quality Tradeoff of Children in a Developing Country: Identification Using Chinese Twins", Working Paper.

6. Powell, Brain and Lala Carr Steelman (1995), "Feeling the Pinch: Child Spacing and Constraints on Parental Economic Investment in Children," *Social Forces*, 73(4), 1465-1486.

7. Robinson, P. (1988), "Root-n-consistent semi-parametric regression." *Econometrica*, 56, 931-954.

8. Rosenzweig, Mark R. (1986), "Birth Spacing and sibling Inequality: Asymmetric Information within the Family," *International Economic Review*, 27(1), 55-76.

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

10. Rosenzweig, Mark R. and Zhang, Junsen (2006), “Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China’s ‘One Child’ Policy”, Working Paper.

11. Yatchew, A. (1998), “Nonparametric regression techniques in economics.” *Journal of Economic Literature*, 36, 669-721.

Appendices:

Figure 1: Histogram of Birth Spacing

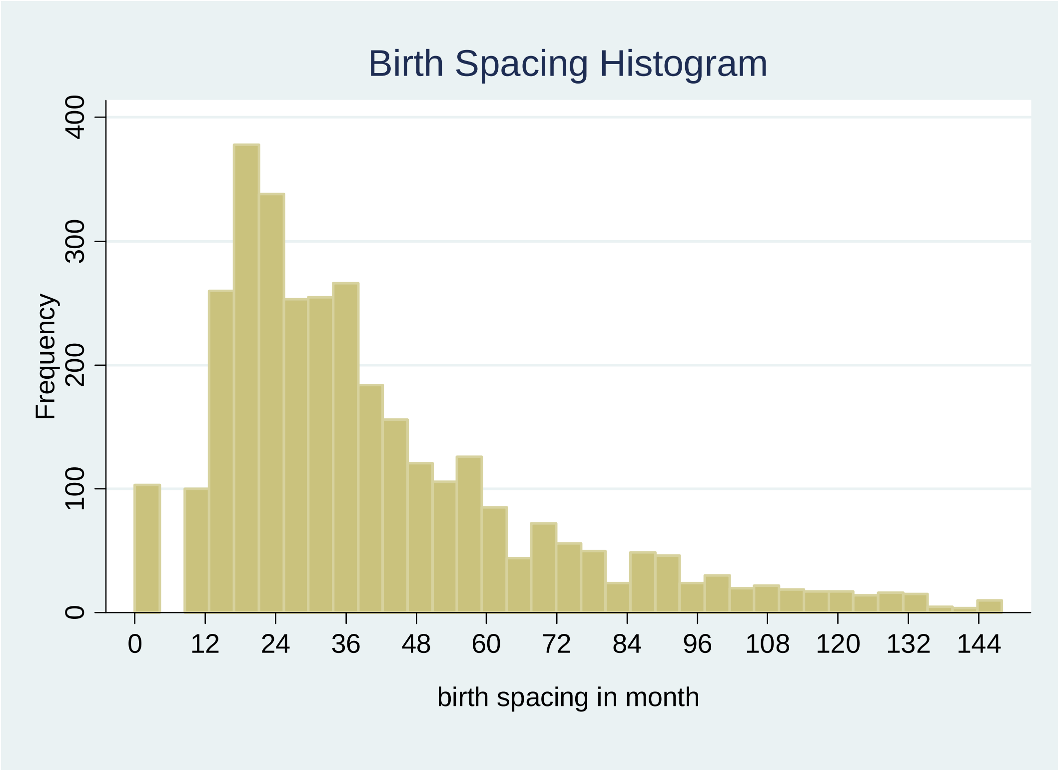

Table 1: Summary Statistics of Children's Attainments (Mean value)

Birth Spacing	Mathematics	Reading Recognition	Reading Comprehension
Less than 24 months	41.26	45.04	40.33
24~36 months	44.27	47.31	42.63
36~48 months	38.60	41.90	37.28
48~60 months	38.07	40.68	36.85
More than 60 months	33.00	35.48	32.11

Table 2

OLS Estimation of the Effect of Birth Spacing on Children's Attainments

(Family size is 3)

	Mathematics	**Reading Recognition**	**Reading Comprehension**
near-spaced=birth spacing*d1	0.138***	0.109**	0.116***
	(0.038)	(0.047)	(0.039)
mid-spaced=birth spacing*d2	0.054**	0.037	0.041*
	(0.022)	(0.027)	(0.022)
far-spaced=birth spacing*d3	0.010	0.004	0.007
	(0.012)	(0.015)	(0.013)
age	0.352***	0.392***	0.323***
	(0.010)	(0.012)	(0.010)
age square	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)
race=1 if white	1.144***	2.314***	1.920***
	(0.548)	(0.683)	(0.564)
age of mother	0.292**	0.382**	0.367***
	(0.128)	(0.160)	(0.132)
education of mother	0.887***	1.008***	0.843***
	(0.126)	(0.157)	(0.130)
birth order	-1.360***	-2.136***	-2.146***
	(0.380)	(0.474)	(0.391)
gender=1 if boy	0.054	2.270***	0.894*
	(0.533)	(0.664)	(0.550)
R square	0.706	0.654	0.667
Observation	1153	1148	1132

a. Note: *, **, and *** represent significance levels of 90, 95, and 99 percent.

b. Standard errors are showed in parentheses.

Table 3: The first step of the 2SLS
Using Twin and Catholic as Instruments Variables to estimate Birth Spacing separately
Dependent Variable: birth spacing (nearest sibling spacing of the child)

	Coefficient		**Coefficient**
twin	-46.181***	Catholic	-2.470*
	（3.422）		（1.328）
age	-0.142***	age	-0.140***
	（0.011）		（0.011）
age square	0.001***	age square	0.001***
	（0.000）		（0.000）
race=1 if white	-1.677	race=1 if white	-1.682
	（1.075）		（1.104）
age of mother	1.824***	age of mother	1.717***
	（0.262）		（0.269）
education of mother	-1.702***	education of mother	-1.868***
	（0.233）		（0.239）
birth order	-2.008**	birth order	-2.122***
	（0.786）		（0.812）
gender=1 if boy	0.164	gender=1 if boy	0.582
	（1.028）		（1.055）
R square	0.142	R square	0.096
Observation	3318	Observation	3318

a. Note: *, **, and *** represent significance levels of 90, 95, and 99 percent.

b. Standard errors are showed in parentheses.

Table 4: The second step of the 2SLS (Twin is Instrument Variable)

Dependent Variable: Mathematics, Reading Recognition, and Reading Comprehension scores

	Mathematics	**Reading Recognition**	**Reading Comprehension**
fitted value of near-spaced	0.084**	0.040	0.039
	（0.042）	（0.052）	（0.043）
fitted value of mid-spaced	0.082**	0.009	0.017
	（0.039）	（0.049）	（0.040）
fitted value of far-spaced	0.158*	0.119	0.114
	（0.097）	（0.120）	（0.099）
age	0.370***	0.392***	0.326***
	（0.013）	（0.016）	（0.014）
age square	-0.003***	-0.003***	-0.003***
	（0.000）	（0.000）	（0.000）
race=1 if white	1.319**	2.235***	1.887***
	（0.568）	（0.705）	（0.584）
age of mother	0.122	0.422**	0.372**
	（0.174）	（0.216）	（0.179）
education of mother	1.113***	1.043***	0.906***
	（0.166）	（0.206）	（0.171）
birth order	-1.154***	-2.120***	-2.107***
	（0.396）	（0.493）	（0.407）
gender=1 if boy	0.021	2.316***	0.937*
	（0.538）	（0.669）	（0.555）
R square	0.702	0.652	0.663
Observation	1153	1148	1132

a. Note: *, **, and *** represent significance levels of 90, 95, and 99 percent.

b. Standard errors are showed in parentheses.

Table 5: The second step of the 2SLS (Catholic is Instrument Variable)

Dependent Variable: Mathematics, Reading Recognition, Reading Comprehension scores

	Mathematics	**Reading Recognition**	**Reading Comprehension**
Fitted value of near-spaced	-0.029	-0.090	-0.067
	(0.108)	(0.133)	(0.110)
Fitted value of mid-spaced	-0.074	-0.139	-0.116
	(0.100)	(0.125)	(0.103)
Fitted value of far-spaced	0.114	-0.004	-0.005
	(0.171)	(0.211)	(0.174)
age	0.337***	0.367***	0.301***
	(0.017)	(0.020)	(0.017)
age square	-0.003***	-0.003***	-0.003***
	(0.000)	(0.000)	(0.000)
race=1 if white	0.861	1.916***	1.556***
	(0.576)	(0.715)	(0.592)
age of mother	0.520**	0.720***	0.673***
	(0.208)	(0.257)	(0.213)
education of mother	0.701***	0.698***	0.567***
	(0.212)	(0.262)	(0.216)
birth order	-1.551***	-2.451***	-2.423***
	(0.423)	(0.526)	(0.435)
gender=1 if boy	0.155	2.429***	1.041*
	(0.543)	(0.675)	(0.559)
R square	0.702	0.653	0.665
Observation	1153	1148	1132

a. Note: *, **, and *** represent significance levels of 90, 95, and 99 percent.

b. Standard errors are showed in parentheses.

Table 6: Nonlinear Estimation of Birth Spacing Effects Using Quadratic Form of Birth Spacing

	Mathematics	Reading Recognition	Reading Comprehension
square of birth spacing	-0.0007***	-0.0004	-0.0002
	0.0002	0.0003	0.0002
birth spacing	0.0644**	0.0312	0.0098
	0.0296	0.0367	0.0304
age	0.3532***	0.3930***	0.3242***
	0.0098	0.0122	0.0101
age square	-0.0033***	-0.0030***	-0.0028***
	0.0003	0.0003	0.0003
race=1 if white	1.1225**	2.2793***	1.8888***
	0.5504	0.6842	0.5669
age of mother	0.3090**	0.3910**	0.3699***
	0.1291	0.1606	0.1331
education of mother	0.9013***	1.0224***	0.8591***
	0.1264	0.1569	0.1307
birth order	-1.4058***	-2.1582***	-2.1590***
	0.3816	0.4757	0.3935
gender=1 if boy	0.1123	2.3221***	0.9267*
	0.5348	0.6660	0.5524
R square	0.7035	0.6523	0.6634
Observation	1153	1148	1132

a. Note: *, **, and *** represent significance levels of 90, 95, and 99 percent.

b. Standard errors are showed in parentheses.

Table 7

Semi-parametric Regression: Effect of Birth Spacing on Children's Attainments

<table>
<tr><th></th><th>Mathematics</th><th>Reading Recognition</th><th>Reading Comprehension</th></tr>
<tr><td>Function of birth spacing</td><td>P>|V| = 0.000</td><td>P>|V| = 0.606</td><td>P>|V| = 0.017</td></tr>
<tr><td>age</td><td>0.371***</td><td>0.403***</td><td>0.328***</td></tr>
<tr><td></td><td>(0.019)</td><td>（0.024）</td><td>（0.019）</td></tr>
<tr><td>age square</td><td>-0.002***</td><td>-0.002***</td><td>-0.002***</td></tr>
<tr><td></td><td>(0.001)</td><td>（0.001）</td><td>（0.001）</td></tr>
<tr><td>race=1 if white</td><td>1.170*</td><td>2.547***</td><td>1.706**</td></tr>
<tr><td></td><td>(0.663)</td><td>（0.867）</td><td>（0.692）</td></tr>
<tr><td>age of mother</td><td>0.289*</td><td>0.380</td><td>0.238</td></tr>
<tr><td></td><td>（0.152)</td><td>（0.200）</td><td>（0.160）</td></tr>
<tr><td>education of mother</td><td>0.908***</td><td>1.097***</td><td>0.932***</td></tr>
<tr><td></td><td>（0.151）</td><td>（0.197）</td><td>（0.158）</td></tr>
<tr><td>birth order</td><td>-1.316***</td><td>-1.831***</td><td>-1.940***</td></tr>
<tr><td></td><td>（0.454）</td><td>（0.595）</td><td>（0.476）</td></tr>
<tr><td>gender=1 if boy</td><td>0.371</td><td>2.739***</td><td>0.729</td></tr>
<tr><td></td><td>（0.610）</td><td>（0.802）</td><td>（0.645）</td></tr>
<tr><td>R square</td><td>0.5082</td><td>0.4339</td><td>0.4631</td></tr>
<tr><td>observation</td><td>1152</td><td>1147</td><td>1131</td></tr>
</table>

a. Note: *, **, and *** represent significance levels of 90, 95, and 99 percent.

b. Standard errors are showed in parentheses.

Figure 2: Function of Birth Spacing Effects on Mathematics.

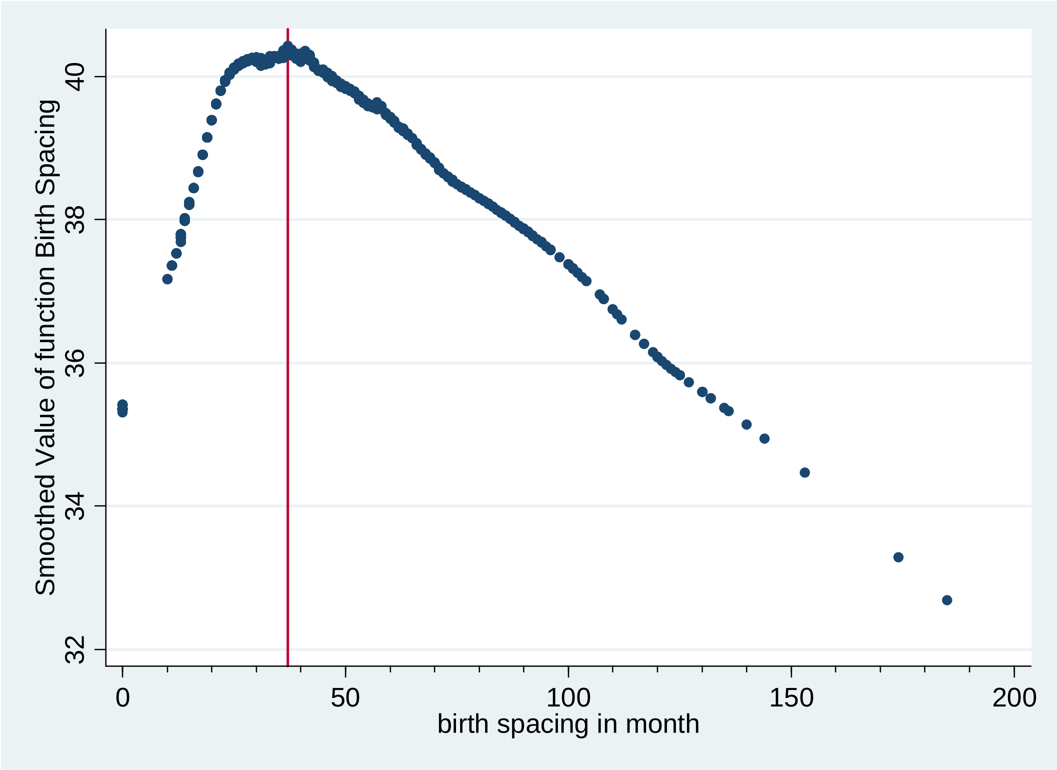

Figure 3: Function of Birth Spacing Effects on Reading Recognition.

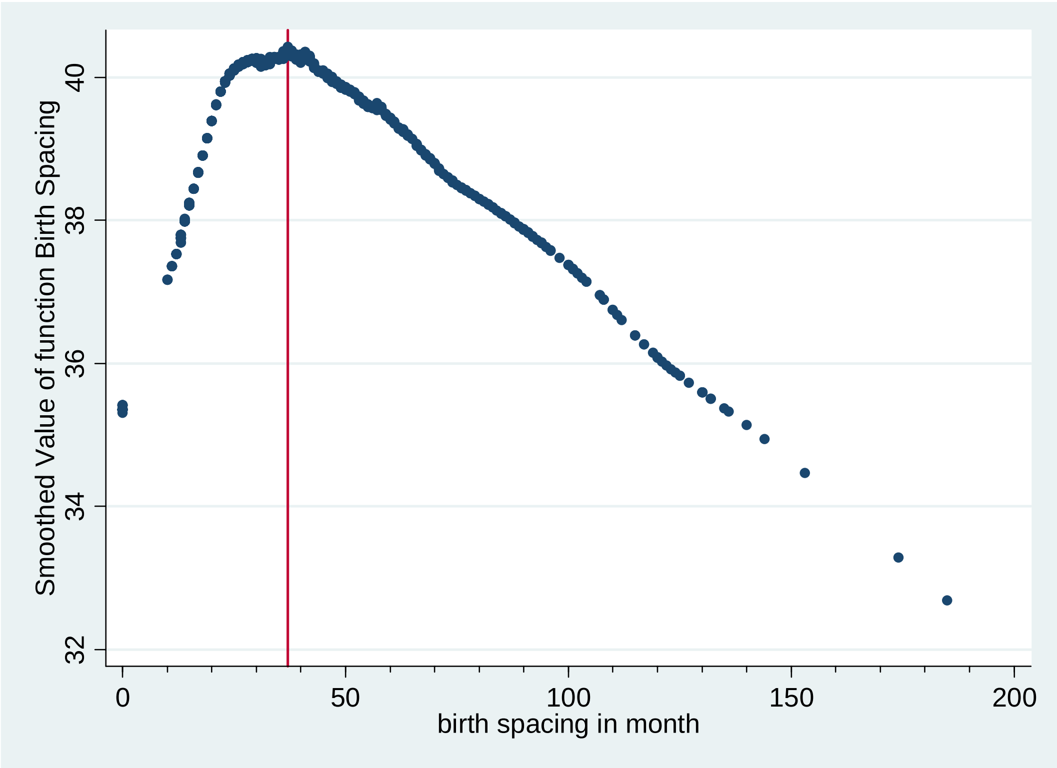

Figure4: Function of Birth Spacing Effects on Reading Comprehension

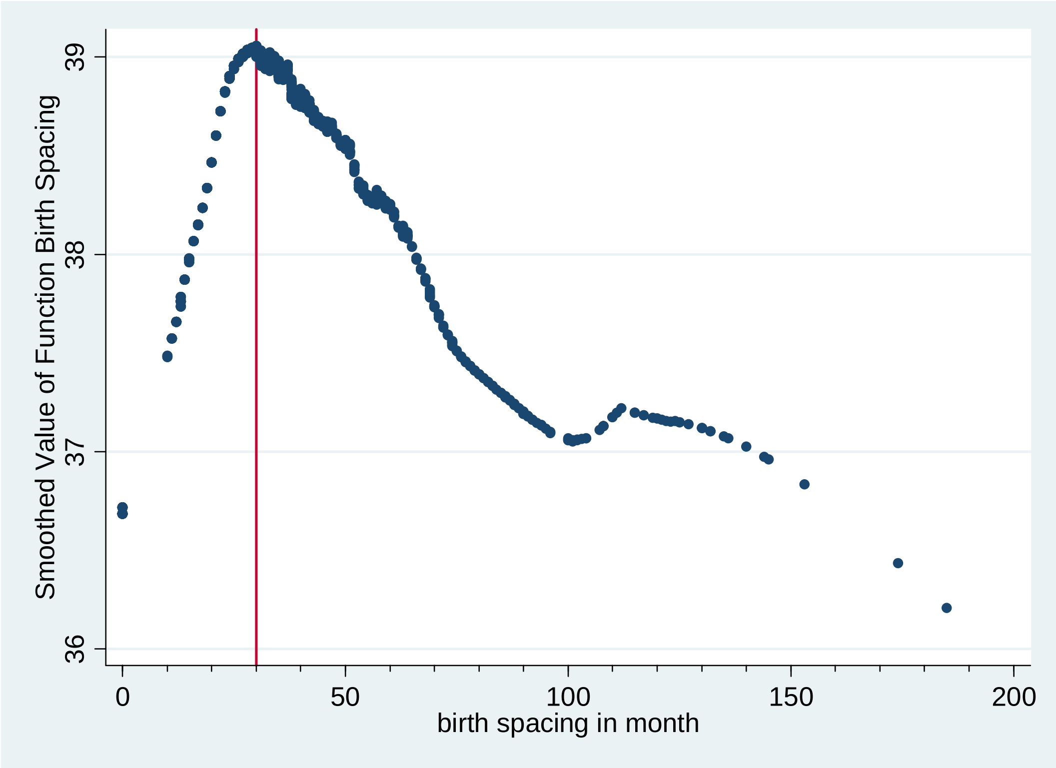

Table 8: Golden Birth Spacing.

	Mathematics	Reading Cognition	Reading Comprehension
quadratic estimation	46 months	37 months	25 months
Partially linear estimation	38 months	37 months	30 months