Family size, birth order and educational attainment: Evidence from China

Extended abstract

The relationship between family size and economic development has attracted attention of social scientists for decades. One critical gear for economic development is the advancements in the human capital of the population. At family level, as the theoretical framework of Becker and Lewis (1973) and Becker and Tomes (1976) suggests, reduction in the quantity of children tends to increase the resource available for investment per child so that child quality will rise. Numerous empirical studies using data from various countries and over different time periods have attempted to test the quantity-quality tradeoff, and either confirmed the prediction of the negative correlation between family size and child quality (e.g., Rosenzweig and Wolpin, 1980; Hanushek, 1992; Conley and Glauber, 2006; Glick et al., 2007; Lee, 2008; Li et al., 2008; Rosenzweig and Zhang, 2009; Ponczek and Souza, 2012; Liu, 2014; Huang, 2015) or found little evidence (Black et al., 2005; Caceres-Delpiano; 2006; Angrist et al., 2010).

Birth order, a factor strongly related to family size, is also considered an important determinant of child quality in several theories through affecting childhood conditions. Childhood conditions may depend on child's intellectual environment. The confluence model in the psychological literature (Zajonc and Markus, 1975; Zajonc et al., 1979; Zajonc and Mulally, 1997) suggests that child's intellectual environment can be defined as a function of the average of the absolute intellectual levels proxied by age of its family members, which declines with birth order. Childhood conditions may vary with parental time and resources. The dilution model of parental time and resources in the economics literature (Becker, 1981; Behrman, 1997) indicates a negative effect of birth order because later-born children have to share parental time with earlier-born children and the resources available for their human capital investment may have been depleted if parents do not plan the quantity and quality investment simultaneously. However, if parents are financially constrained, the birth order effect may be reversed since parents may be better off later in time when higher-ordered children need investment. The empirical findings in general support the theoretical predictions of a negative birth order effect on educational attainment in developed countries (Black et al., 2005; Conley and Glauber 2006; Kantarevic and Mechoulan 2006; Kristensen and Bjerkedal 2007; Booth and Kee 2009; De Haan 2010), but positive effect in financially constrained families of developing countries (Ejrnæs and Pörtner, 2004; Tenikue and Verheyden, 2010; De Hann et al., 2014).

The present paper examines the effect of family size and birth order on children's educational attainment in China. One important identification issue is that both child quantity and quality are endogenously determined by unobserved parental preferences and family characteristics, because childbearing and child outcome are jointly chosen by parents (Rosenzweig and Wolpin, 1980). One crucial method for tackling endogeneity is to use the exogenous variations in family size introduced without parental control. Widely employed instruments include the occurrence of twin/ multiple births (e.g., Li et al., 2008; Rosenzweig and Zhang, 2009; Angrist et al., 2010), sex composition of first two children across the globe (e.g., Conley and Glauber, 2006; Lee, 2008; Angrist et al., 2010) and one-child policy related measures specifically for China (e.g., Qian, 2009; Liu, 2014; Huang, 2015). These studies typically abstract from birth order effect. Due to its correlation with family size, birth order is also an endogenous variable. The identification of birth order effect in the literature typically relies on regression models including family fixed effects to rule out that estimated birth order patterns are driven by differences in family size or in any other unobserved family characteristic shared among siblings. To avoid confounding effect from family size, some studies estimate the birth order effect separately for each family size (e.g., Black et al., 2005; De Hann, 2010). However, the fixed effects model still cannot deal with the unobservables that are time variant. In this paper, we attempt to treat

both family size and birth order as endogenous variables and use an instrumental variable approach to investigate the effect of family size and birth order on children's educational attainment.

Our data are derived from the fifth wave of the Chinese Household Income Project survey (henceforth CHIP2013) collected in 2014. The sample was obtained from a larger sample regularly used by the National Bureau of Statistics of China to produce official statistics for China, and the sampling followed a multi-stage procedure: 31 provincial-level administrative units except Hong Kong, Macao and Taiwan were divided according to their geographical location into eastern, central and western units, and 14 were selected including Anhui, Beijing, Chongqing, Gansu, Guangdong, Henan, Hubei, Hunan, Jiangsu, Liaoning, Shandong, Shanxi, Sichuan, and Yunnan. Within each provincial-level unit, counties in rural areas and cities in urban areas were then selected according to their income level and economic development. We match the information on parents and all children of the respondent's family of origin to construct a within-family sibling dataset and clean the data following the rules commonly employed by other studies in the literature. After selection we are left with 36,039 children born aged 25 to 65 from 10,312 families. Educational attainment is measured by years of schooling, family size is defined as the reported number of children in a family from one to eight plus (including eight, nine, ten and eleven), and birth order of each child is identified from the information on year of birth.

We follow the recent empirical literature to specify an equation with child's education as the dependent variable and measure family size and birth order both as a continuous variable and as indicators to capture their non-linear effect. We measure family size and birth order both as a continuous variable and as indicators to capture their non-linear effect. Besides family size and birth order we include a vector of child characteristics, including age, gender, ethnic group, and *Hukou* at birth, and a set of parental attributes, including age and years of education. We estimate this equation first by ordinary least squares (OLS). Since family size and birth order tend to be endogenous and the OLS results may merely suggest a correlation, we identify a causal effect by using two-stage least squares (2SLS) where family size and birth order are both measured as continuous variables.

Our instruments are constructed using mother's exposure to various phases of China's family planning policies that can be broadly defined by the stringency of the birth quota after the People's Republic of China was founded in 1949. The first phase (1949-1970) is captured by no or leniently and narrowly implemented birth quota, the second phase (1971-1979) by strongly and broadly implemented birth quota, and the third phase by the one-child policy (1980-2013) (Wang, 2016). An indicator is constructed for whether a mother at each age during primary fertile age range 15-49 is exposed to each of the policy phase, equal to one if the mother's condition is met, zero if not. We then compute the probability of childbearing at each of mother's age between 15 and 49, which is defined as the number of children born to mother of each age divided by total number of children in the sample. Then two policy variables are constructed for phases two and three, each of which is a summation of the exposure indicator weighted by the probability of childbearing at each age between 15 and 49. Since China's family planning policies differ by *Hukou* type (urban/rural) and ethnicity (Han/non-Han), the two policy variables are interacted with an urban dummy and a Han dummy. Therefore, we have four instruments for two endogenous variables, and the model is overidentified.

Besides the instrumental variable approach, we also estimate a within family fixed effects (FE) model to study the effect of birth order on educational attainment. Birth order is also measured both as continuous variable and as indicators.

Table 1 reports the OLS (columns (1)-(3)) and fixed effects (columns (4)-(5)) regression results. When using OLS, we find a significantly negative correlation between family size and children's education when family size is measured as a continuous variable. When measured as dummies with single-child families as the reference, the coefficients are all negative, statistically significant, and decreasing in family size. These results are consistent with the predication from the quantity-quality tradeoff model. In addition, we find a significantly positive correlation between birth order and children's education when birth order is measured as a continuous variable. When measured as dummies with first-borns as the reference, the third child starts to have significantly more education than the first-born, and the gap increases with birth order. Furthermore, when estimating birth order effect by fixed effects model, we find that the results are very similar to the OLS results. These findings support the financial constraints theory and are in line with those obtained using developing country data. Table 2 reports the 2SLS estimation results. Family size has a negative effect and birth order has a positive effect on children's education. Although the OLS results hold, the magnitude of both family size and birth order coefficients are multi-folded. Test of overidentifying restrictions, test of weak instruments and test of endogeneity of the concerned variables are all passed. Our results are robust to excluding single-child families and using a larger sample by including all matched siblings sample.

Table 1. Effect of family size and birth order on children's education - OLS and FE

	OLS			FE	
Dependent variable: Children's education	(1)	(2)	(3)	(4)	(5)
Family size	-0.291***	-0.298***			
	(0.021)	(0.021)			
Family size indicators					
Two-child family			-0.615***		
			(0.114)		
Three-child family			-1.018***		
			(0.115)		
Four-child family			-1.341***		
			(0.122)		
Five-child family			-1.544***		
			(0.128)		
Six-child family			-1.816***		
			(0.145)		
Seven-child family			-1.996***		
			(0.185)		
Eight+ -child family			-2.629***		
			(0.268)		
Birth order	0.087***			0.096***	
	(0.021)			(0.030)	
Birth order indicators					
Second		-0.053	-0.023		0.001
		(0.040)	(0.040)		(0.046)
Third		0.065	0.113**		0.131*
		(0.054)	(0.053)		(0.067)
Fourth		0.177**	0.211***		0.234**
		(0.072)	(0.071)		(0.095)
Fifth		0.364***	0.349***		0.370***
		(0.094)	(0.092)		(0.123)
Sixth		0.412***	0.377***		0.412**
		(0.121)	(0.121)		(0.160)
Seventh		0.714***	0.664***		0.610***
		(0.180)	(0.178)		(0.212)
Eighth or later		0.560*	0.753**		0.838***
		(0.296)	(0.299)		(0.324)
Individual characteristics	Yes	Yes	Yes	Yes	Yes
Parental characteristics	Yes	Yes	Yes	No	No
Family fixed effects	No	No	No	Yes	Yes
R-squared	0.302	0.302	0.303	0.095 [1]	0.096 [1]
No. of observations	36,039	36,039	36,039	36,039	36,039
No. of families	10,312	10,312	10,312	10,312	10,312

Notes: All regressions are based on respondent sample. Individual characteristics for OLS estimations include indicators for age, gender, ethnicity and Hukou status. Individual characteristics for FE estimations include indicators for age and gender. Parental characteristics include mother's age, mother's education, father's age, and father's education. Robust standard errors allowing for heteroskedasticity and correlation within family are reported in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

[1] FE R-squared is within-family R-squared.

Table 2. Effect of family size and birth order on children's education - 2SLS

Second stage		First stage		
Dependent variable	Child's education (1)	Dependent variable	Family size (2)	Birth order (3)
Family size	-1.539***	Policy 2 * Urban	-0.400**	-0.640***
	(0.486)		0.169	0.104
Birth order	1.286***	Policy 3 * Urban	0.145	-0.093
	(0.498)		0.099	0.061
		Policy 2 * Han	-0.991***	-1.094***
			0.138	0.088
		Policy 3 * Han	-1.768***	-1.496***
			0.137	0.094
Test of endogeneity: F-statistic[1]	3.766**			
Test of weak instruments: Minimum eigenvalue statistic[2]	13.306			
Test of overidentifying restrictions: Sargan chi2 [3]	2.072			
Individual characteristics	Yes		Yes	Yes
Parental characteristics	Yes		Yes	Yes
Family fixed effects	No		No	No
R-squared	0.141		0.268	0.480
First stage F-statistics			110.25***	282.74***
No. of observations	36,039		36,039	36,039
No. of families	10,312		10,312	10,312

Notes: All regressions are based on respondent sample. Individual characteristics include indicators for age, gender, ethnicity and Hukou status. Parental characteristics include mother's age, mother's education, father's age, and father's education. Robust standard errors allowing for heteroskedasticity and correlation within family are reported in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% levels, respectively.

[1] Endogeneity test determines whether endogenous regressors in the model are in fact exogenous. Rejection of the null hypothesis of exogeneity indicates endogeneity of the concenred regressors.

[2] Minimum eigenvalue statistic measures the relevance of the excluded exogenous variables. A statistic greater than the chosen critical value indicates rejection of the null hypothesis of weak instruments. See Stock and Yogo (2005) for critical values. The 2SLS relative bias at 5% level is 11.04.

[3] Overidentification test performs the test of overidentifying restrictions. Rejection of the null hpothesis of valid instruments indicates that the instruments may be invalid.