Son Preference and Human Capital

Wang, Qingfeng

Nottingham University Business School China

8 January 2018

Son Preference and Human Capital

Qingfeng Wang

University of Nottingham Business School

ABSTRACT

We have developed a theoretical link between the existence of a son preference and the quantity and quality of children. In our model, decisions about the quantity and quality of children are interdependent and are influenced by a son preference. A son preference substantially widens the gap between the quality of female and male children, and increases the possibility of having more children to realize this preference. An increase in the level of son preference lowers the average level of education, especially for female children. For households with a son preference, having a female birth at each birth order significantly increased the probability of having more children, with a magnitude which is much larger than that suggested in Dahl and Moretti (2008). More interestingly, once such households have a son, there is a significant reduction in that household's likelihood of having more children. Furthermore, our findings establish the causal effect of a son preference on the quantity and quality of children within a family as a son preference is found to greatly reduce households' sensitivity to the change in the shadow price of quantity as well as quality.

Keywords: Son Preference; the Quantity and the Quality of Children

JEL: I2, J24, J31

I. Introduction

"A woman's virtue is to have no talent" 女子无才便是德[1]。

There is a large amount of literature on growth models which highlights the importance of gender inequality (as a result of gender preference) in economic growth[2] through the channel of human capital differentiation between different genders. The impacts of gender differences have been observed in many different domains, perhaps most notably in labor market (e.g., Blau and Kahn 2000). Gender inequality (or gender preference) is deeply rooted in countries with a large fraction of its population depends on agriculture for its livelihood, in patrilineal societies and cultures with a tradition of ancestor worship, and it has also found a foothold in some developed economies[3]. One form of gender inequality is perfectly mirrored in the son preference cultures around the world. Son preference is more obvious and more severe in some countries than in others. For example, the son preference culture is much stronger in Asian countries[4] than in Western Europe. We have focused specifically on the son preference because of its impact on the human capital, and hence on economic growth, and because the previous research has little to say about the mechanism underlying the impacts of a son preference on the quantity and quality of children. The quantity and quality of children are often inspired by specific demand. Empirical evidence indicates that children from larger families have lower average educational attainments. However, such evidence does not explain why children within a family may have very different levels of education regardless of family size. It is intriguing, although challenging, to understand why some families choose to have more children but others do not, and more importantly, how the underlying

[1] The original quotation reads: "眉公曰：丈夫有德便是才，女子无才便是德。此语殊为未确。" (出自清·张岱《公祭祁夫人文》).

[2] For example, Lagerlöf 1999; Galor and Weil 1996; Barro 1991; Bloom and Williamson 1998; Taylor 1998; Barro and Lee 1994; Barro and Sala-i-Martin 1995; Dollar and Gatti 1999; Forbes 2000; Lorgelly and Owen 1999; Knowles et al. 2002.

[3] According to a 2011 Gallup survey and a Pew report, a son preference also occurs in the U.S. And, a continuous son preference was observed in Finland among the national majority and the Swedish speaking minority (Andersson et al. 2007).

[4] For instance, the son preference in South Korea was once considered stronger than anywhere else in the world (Goodkind 1999; Park and Cho 1995). All countries with a significantly skewed sex ratio at birth are in Asia.

rationales that bolster decisions to have more children have a cause effect on both the family size and the quality of children.

In his seminal paper on the theory of fertility choice, Gary S. Becker (1960) introduced fertility choice into the realm of economic analysis. Since then, children have been treated as consumer durables. Although Becker acknowledged that preferences may vary across households, the impact of a preference such as gender preference on endowment and on any interactions between the quantity and quality of children was not addressed in his papers, nor in any papers with coauthors. The literature that extends the quantity and quality model introduced by Becker (1960) is quite extensive; despite this, however, they fail to capture the effects of the essential parental characteristics on the quantity and quality of children. We believe that it is crucial to include these parental characteristics in the quantity-quality model as they could affect both family size and children's educational outcome[5].

The family environment is widely believed to be a primary factor in determining children's educational outcomes (Black et al. 2005). In some societies, the son bears the responsibility of continuing the family line, which is a critically important aspect of the "quality" of a child that parents are looking for in these cultures, and males are in general have higher earning power than females, which can also lead to a preference for sons over daughters (Qian 2008). We believe that a son preference is a very important component of a family environment, and it has some causal effects on the quantity and quality of children within a family. But, how does the son preference affect households' decisions on both the quantity and quality of children? We argue that to answer this question, it is essential to account for the fact that parents in societies with a strong son preference culture value sons more than daughters[6].

[5] Family size and children's educational outcome can be endogenously determined by parents. In societies with no compulsory education system, children's education attainment is decided by their parents according to, for example, differentiated labor market returns to education for men vs. women and family's need in labor supply from their children.

[6] Considering children as consumable durable, the justification for those households with a strong son preference place a higher value on a son than on a daughter is comparable with consumers who are strongly attached to a particular brand, in which case they would place higher a preference or value on products from that particular brand than from other brands.

We develop a theoretical link between the existence of a son preference and the quantity and quality of children. In our model, the quantity and quality decisions about children are interdependent and are influenced by a son preference. A son preference substantially widens the gap between the quality of female and male children in households with a son preference, and increases the possibility of having more children to realize this preference. An increase in the level of son preference lowers the average level of education, especially for female children. In this article, we conduct several empirical tests showing the importance of this channel in households' decision in the quality and quantity of children.

Our analysis has yielded the following findings. Foremost, irrespective of how our findings are interpreted, we have shown that children's gender matters in terms of households' decisions about their educational outcomes and about the number of children to have. After controlling for a number of factors, ranging from geographic locations[7] to variables which measure regional economic developments, we found that female children receive significantly lower levels of education than their male siblings. We have taken the following approach to distinguish the causal effect of a son preference on the quality of children. We have compared the quality of children of opposite sexes by using multiple birth samples (where members of twins or triplets from each household were of opposite sexes), while also controlling for family background characteristics. By doing so, we addressed the following two concerns. First, the issue of endogeneity, which is the educational investment in each child is decided by parents, and hence may be related to other unobservable parental and other family characteristics. By using multiple birth data, we were able to control for these characteristics as these twins (triplets) would have concurrently experienced similar changes in the unobservable parental and other family characteristics. Second, using the multiple birth data helped to alleviate concerns that any differences in the quality of male and female children was as a consequence of significant changes in the socioeconomic and educational environment for children of different genders and of a different age or because children from the

same family would have both inherited very different intellectual ability from their parents[8]. Doing so yielded an even wider gap in the quality of female and male children.

Second, we find that for households with a son preference, having a female birth at each birth order significantly increased the probability of having more children, which is much larger than that suggested in Dahl and Moretti (2008). More interestingly, once such households have a son, this reduces significantly the probability of having more children, and this negative effect ranges from 5.6% to 41.8% depending on the birth-order of the son.

Third, the strong existence of a son preference in rural China compared to that found in urban areas can be explained by the paramount differences between the two areas in terms of such factors as the distinctive features of China's welfare system[9] and the fact that job opportunities are more concentrated and diverse in urban areas. Such differences between the two areas have shaped the large differences in people's perceptions of the quality of children of different genders. Based on these clear differences, we were able to quantify the level of son preference in rural China, where the estimated son preference in rural China ranged between 0.20 and 1 across the three waves of CHIPS, where earlier years' rural samples are associated with a stronger son preference.

A final observation is in order. As we have used Chinese data to estimate empirically the impact of a son preference on the quality and quantity of children, we need to highlight the differences between the son preference culture in China and other countries. Although, son preference culture is not unique to China, as shown by Dahl and Moretti (2008) who concluded that parents in the U.S. favor boys over girls, we argue that the nature of China's son preference differs significantly from other countries. Son preference may originate from cultural considerations, as well as factors such as a concern for care of the elderly, but in China, many people continue to endorse the

[8] Both the boys and the girls from pairs of twins (triplets, quadruplets and so on) were free from their parents' gender preference, so they receive unbiased and equal care during pregnancy; and more importantly, twins (triplets, quadruplets and so on) share either some or all of the same genomes, so we have strong reason to believe that they will have inherited very similar intellectual ability from their parents.

[9] China's social welfare system is linked to its hukou system. During our sampling time periods, only those holding non-agricultural hukou status had access to healthcare, employment, retirement pensions, housing, and education programs provided by the government, while most rural residents are generally expected to provide for themselves (Young 2013).

traditional belief that "there are three unfilial things in your life, and to leave no posterity[10] is the worst" (Shi 1982). China has a strong son preference culture in its rural areas (Goodkind 2015), where many villagers' lifetime goal is to have sons (Li 1995; Wasserstrom 1984). The strength of the son preference in China is partially evidenced in Amartya Sen's book *'More than 100 Million Women Are Missing'*. Furthermore, other factors (e.g. economic factors) fail to explain the highly skewed sex ratio at birth (SRB) observed in Asian countries, as none of the developed economies, with the exception of South Korea[11], has a SRB above 1.10. A number of studies has confirmed that the highly skewed SRB in China is largely due to the persistent and strong son preference culture in China (Ebenstein 2010; Chen et al. 2013; Wang 2019). If this is how the son preference works in China, it suggests a very different mechanism from that which seemingly functions in other countries, such as in the U.S.

Our findings are related to three bodies of literature. First, by estimating empirically the level of son preference, our study contributes to the small but growing body of literature on the persistence of this phenomenon (Shi 2009; Ebenstein and Leung 2010; Almond et al. 2013; Hu and Tian 2018; Zhang 2019). Second, our study is related to previous research examining the impact of a son preference on educational attainment and parents' investment in their children (e.g., Rosenzweig and Schultz 1982; Chen et al. 1981; Das Gupta 1987; Ahmad and Morduch 1993; Thomas 1994; Burgess and Zhuang 2001; Park and Rukumnuaykit 2004; Kugler and Kumar 2017; Barcellos et al. 2014). Our findings show that a son preference both lowers significantly the average level of education, and has a very large negative impact on the quality of female children.

Third, our study augments the literature regarding the tradeoff between the quantity and quality of children, as pioneered by Gary S. Becker. What sets our study apart from that of Becker and his coauthors' works, as well as others (e.g. Qian 2009; Black et al. 2005), is that we have incorporated the son preference into the relationship between the quantity and quality of children, and we have found that a son preference exerts significantly more impact than that of the quantity of children

[10] Having no posterity refers to having no son, as ancestor worship is an important part of Chinese culture (Schwartz 1985), and it needs to be performed by male descendants. Having no male descendants means the discontinuation of the line of ancestry.

[11] The son preference in South Korea was once believed stronger than in any other country (Goodkind 1999; Park and Cho 1995).

on children's quality. Our findings also suggest that, in societies with a son preference, the quantity and quality of children interacts largely because parents in the process of realizing their son preference need to adjust their decisions about the quantity and quality of their children. A son preference can play a critical role in the interaction between the quality and quantity of children because: 1) it greatly reduces households' sensitivity to the change in the shadow price of quantity as well as quality, and 2) the expected sizable increase in households' marginal utility owing to the realization of a son preference.

The implications from both our theoretical model and empirical findings confirm that the son preference is the driving force behind the gender gap in education in China. Based on our analysis, we have found that a son preference affects the quality of children more than the quantity of children, and it also affects both the shadow price of quality and the shadow price of quantity[12]. The rationale behind this is that in a strong son preference culture, a son is considered to carry the intrinsic quality (or ability) of providing more economic returns to the family than would be the case for a daughter, to continue the family line and to provide care for elderly family members and/or fulfill any inheritance requirements. Therefore, parents with a son preference are more likely to stop producing children only when they have at least one son[13]. When more births are required to realize the son preference, an increase in the quantity of children inevitably lowers the quality of children, especially the quality of daughters. Within their particular budget constraints, having a son is the priority of the household with a strong son preference, while such households are less demanding about the quality of their children compared to those with no clear gender preference, all else being equal. The negative relationship between quantity and quality is robustly confirmed to be negatively correlated across households, however, households with a son preference are found to differentiate strategically on the quality of their children based on their gender rather than to want approximately equal levels of quality for each of their children.

[12] Recall that: an interaction between the quantity and quality that causes significant changes in both shadow price of quantity and of quality is the key element that generates a trade-off between quantity and quality.

[13] Our hypothesis that households possessing a son preference, for the sake of having a son, will very likely continue to have more children, is premised on the following cultural features of Chinese families with a son preference. First, sons are considered as permanent members of the family, and they will maintain lifelong financial relationships with their parents. Second, sons take responsibility for looking after their parents' economic and physical wellbeing, whereas daughters, once married, will contribute mainly to their in-laws' families.

Therefore, the proposition endorsed by Becker 1960; Becker and Lewis 1973; Willis 1973; Becker and Tomes 1976; Rosenzweig and Wolpin 1980, that "the quantity and quality of children interact because parents tend to want approximately equal levels of quality for each of their children" is, in general, not supported among households with a son preference.

Our paper proceeds as follows: the next section describes the data, and Section III outlines the identification strategies in estimating the results. Section IV reports the estimations and discusses the findings, and section V develops the model, and presents the implications suggested by the model. Finally, Section VI presents the conclusion.

II. Construction of the datasets

2.1 The data sample

In estimating the impact of a son preference on the quality and the quantity of children, we made use of three out of the seven waves (1988 to 2013) of the Chinese Household Income Project Survey (CHIPS), namely those from 1988, 1995 and 2002. These are seven years apart, which we consider to be a time period which is neither too short nor too long for the purposes of comparison studies between the different waves. Those individuals surveyed in the first two waves were largely free from the impact of the one-child policy and the 'Nine Years Compulsory Education System'[14], while the one-child policy had more of an impact on the 2002 wave than the other two, the impact was still much milder compared to those in later years. This justifies the use of only the first three waves of survey data. Each survey collected data from households selected randomly from both rural and urban populations, and they covered very large and extensive geographic areas in China. The information of all members of each household is recorded. Each survey contains household information on their income, expenditure, and demographics, as well as geographic information, occupational information, assets and so forth. A more detailed description of each of the seven

[14] By restricting our sample to those children who are above 18 years old, these children's educational outcomes from these two waves of CHIPS are not affected by the 'Nine Years Compulsory Education System'.

waves of the survey is reported on the official website of the China Institute for Income Distribution at Beijing Normal University.

The units of observation are children, to which we link the demographic information of their parents and the household level information from each CHIPS, accordingly. We constructed 18 subsamples in total from the three waves of CHIPS, which were divided equally between both the rural population and the urban population, as well as between the three waves of CHIPS. The size of each sample depends on both the original size of the surveys and the number of missing values of the key variables of interest, and the associated control variables. Some summary statistics of the samples are presented in Table A1 in the Appendix.

2.2 Data to estimate the effect of a son preference on the quality of children

To investigate the impact of a son preference on the quality of children, we constructed data on some of the children included in each of the three waves[15]. As children from rural areas in China are raised and educated in socioeconomic, cultural and educational environments which are very different from those in urban areas, we studied separately the impact of a son preference on the quality of children for rural populations and urban populations.

The key variables of interest are children's education attainment (i.e. their level of education), their gender, and the number of children in a household. The dependent variable is the level of educational attainment. We have used the level of educational attainment as a measure of the quality of a child in relation to a son preference because the number of years a child is in education prior to the introduction of the 'Nine Years Compulsory Education System' in China is largely determined by the parents' willingness to invest in their children's education, and such decisions may be based on gender. The more years in education, the greater the loss of the anticipated income from the labor supply of the children, which discourages parents from providing more education and, more importantly, from providing equal educational opportunities for all children. The earlier their children can begin working, the sooner they can contribute to the family, and this is a

[15] Children with missing values on the key variables, as well as control variables, were dropped from the sample. As the quality of children is measured by their educational attainment, we only included those which were older than 18 years.

convincing strategy to follow when the returns from education are not considered by rural populations to be favorable, and parents are not obliged to send their children to school. However, it should be noted that the educational inequality between genders narrowed considerably alongside China's rapid economic development. In addition, the 'Nine Years Compulsory Education System', which took effect in late 1986, helped both to significantly enlarge the student population, and to reduce the drop-out rate of students, especially that of female students, even in the rural areas (Ding 2012).

The education level[16] (education attainment) is coded from 8 to 1, where 8 represents the highest level of education (university level). For the gender variable, males are coded as 0, while females are 1. In our study, to control for demographic as well as regional effects on the quality of children, we also included the fathers' education attainment[17], a variable called minority[18], the family income, and variables featuring geographic characteristics and locational information.

We extracted a total of 12 subsamples from the three waves of CHIPS, to enable the study of the effect of a son preference on the quality of children. More specifically, for each wave, we compiled two samples for the rural population and two for the urban population. For both the rural and the urban samples, the larger sample contained all children with no missing values on the variables mentioned above, while the smaller sample contained only data on twins, triplets, quadruplets and so on, of opposite sexes[19]. However, we did not distinguish between the different types of twins[20], triplets, quadruplets and so on. The multiple births sample was used to confirm the robustness of the results from using the larger sample.

2.3 Data to estimate the effect of a son preference on the quantity of children

[16] In more detail, the education level was coded as 1 for someone who had never attended formal education, 2 for the lower elementary school (below year 3), 3 for upper elementary school, 4 for junior middle school, 5 for senior middle school (including professional middle school), 6 for technical secondary school, 7 for junior college, and 8 for college/university.

[17] Using mothers' education level did not lead to qualitatively different results.

[18] There are 56 ethnic groups in China, with the Han group considered as the majority. All other ethnic groups are considered as a minority.

[19] For example, we stipulated that they are twins of opposite sexes.

[20] That is, we dis not distinguish between, for example, whether twins were identical (monozygotic) or fraternal (dizygotic).

To investigate the impact of a son preference on the quantity of children, we first used the six larger samples constructed in Section 2.2 respectively for both the rural and urban populations. Here, the number of children in a household taken as the dependent variable (i.e. the quantity of children). We also included the fathers' level of educational attainment, the minority variable, the family income, and the other control variables in the regression analysis. The key variables of interest are gender and quality (i.e. the educational attainment of a child).

The CHIPS data sets we use provide information on each child of a household and contain a panel component. To take advantage of this, we construct six additional samples which incorporated a dummy typed variable, called SonDummy[21]. The SonDummy variable helps to distinguish households with a son preference from those possess no clear son preference in their decisions on whether or not to have another child after the birth of a son. The SonDummy variable, together with the gender of each child sorted by their birth order, were used to capture the impact of the son preference on the quantity of children by considering both the impact of the birth of a son and the sex composition of the current children on the quantity of children. We included in the samples as many children's observations as possible, by including only the following variables in the robustness test, namely the quantity of children, the gender of the children, and the SonDummy[22]. We could do this because both the independent variable gender and the SonDummy have very small correlations with the other independent variables, and the inclusion or exclusion of the other independent variables did not qualitatively change the results. By doing so, we dropped very few children's observations from each wave of CHIPS. In estimating the effect of a son preference on the quantity of children, children of all ages are included in the study.

III. Identification

[21] To construct the SonDummy variable, we first sorted the children of each household according to their birth order, and numbered the first-born child as 1, the second-born child as 2, and so on. The SonDummy takes the value of zero, corresponding to the female children who were born to a household before the first-born male, while it takes value of 1 once a son has been born into the family. The number of observations of the SonDummy equals the number of children in a household.

[22] Using the SonDummy either as an independent variable in the Poisson regression or as a data selection condition in the Poisson regression did not produce qualitatively different results.

The following reduced form analyses were employed to empirically estimate the impact of the son preference on both the quality and quantity of children.

3.1 Using gender to identify the effect of a son preference on the quality of children

We consider the impact of the gender of a child on the quality of children is a natural candidate to use to represent the son preference of a household. The son preference of a household will be reflected in the differentiated quality of children, measured by their educational attainment, with their parents as the 'investors' in their children's education. All else being equal, a household with a son preference is likely to invest more in the education of their sons than in their daughters. The presence of such an effect is confirmed if gender exerts a significantly negative impact on the quality of those female children.

China is a vast country with huge discrepancies in the development of socioeconomic and education system across its provinces, and even across counties within a certain province. Therefore, we have included geographic dimensions in capturing factors other than a son preference in the study of the quality of children. The trade-off between the quality and the quantity of children suggests the inclusion of the quantity of children in the study. We therefore propose the following reduced form equation, which will be used to estimate the impact of a son preference on the quality of children:

$$\begin{aligned}\text{Quliaty}_{ijkw} = {} & \alpha_{1k} + \alpha_2 Gender_{ijkw} + \alpha_3 Quantity_{jkw} + \alpha_4 EduFather_{jkw} + \alpha_5 Minority_{jkw} \\ & + \alpha_6 Income_{jkw} + \alpha_7 \left(LocatioVariables_{jkw}\right) + AgeDummy_{ijkw} + \epsilon_{jkw}.\end{aligned} \tag{1}$$

We regress the quality of children (i.e. quality of child i of household j who lives in area k - either rural or urban, and who is taken from the survey data collected in year w) on gender, the quantity of children and other controls. The variables with four digits of subscript are individual level data, and variables with three digits of subscript are household level data. The father's educational attainment is used to control for a household's level of endowment. The ethnic minority groups enjoy preferential treatment from the Chinese government, such as exemption from the population control policies, lower university entrance scores and lower tax thresholds. Consequently, it is

meaningful to distinguish households from ethnic minorities. Minority is a dummy variable, which is equal to zero if a household is Han, and equal to one where it belongs to another ethnic group. Unlike most other countries, China's hukou system clearly defines each individual's resident status as either rural or urban, which distinguishes individuals, for example, in their eligibility for access to public health and other welfare systems. Therefore, there was no difficulty in identifying whether an individual is from a rural or an urban area. Our locational variables comprise information about the geographical features surrounding the areas where households are situated. They also contain information on whether a rural resident is living closer to cities, which helps to differentiate those rural households who live closer to cities than other rural households from the same county but further away from cities. These locational variables and other household level variables allow us to difference out household specific characteristics that are affecting children's education. All explanatory variables are confirmed to either have very small correlations or the absolute value of the correlations was approximately 0.30. To control for cohort effects, we include indicator variables for age groups in the regressions. We estimated equation (1) for both rural and urban samples, and for those samples containing only twins (triplets and so on), and for samples containing all types of households.

3.2 Identifying the effect of a son preference on the quantity of children

To obtain a quantitative estimate of the effect of a son preference on the quantity of children, we first estimated the following reduced form equation.

$$\text{Quantity}_{jkw} = \beta_1 + \beta_2 Gender_{ijkw} + \beta_3 Quality_{ijkw} + \beta_4 EduFather_{jkw} + \beta_5 Minority_{jkw} + \beta_6 Income_{jkw} + \alpha_7\left(LocationVariables_{jkw}\right) + \nu_{jkw}. \quad (2)$$

Quantity_{jkw} is the number of children of household j from area k (either rural or urban) and is taken from CHIPS data in year w. The key independent variables of interest are gender and quality. We regress the quantity of children on children's gender, the quality of children, and other controls. An increase in the quality of children exhausts some of the resources which may otherwise be used in producing more children, therefore, quality is expected to have a negative effect on Quantity_{jkw}. It is based on the rationale that a household with a son preference tends to increase the chance of

having a son by having more children before they have a son, and thus a positive gender effect on Quantity_{jkw} would capture this effect.

Using specially constructed samples, we conducted the following reduced form analysis using Poisson regression, both to validate the earlier findings, and also to understand the households' underlying rationales in determining the number of children.

$$\text{Quantity}_{jkw} = \gamma_1 + \gamma_2 BirthOrderFemale_{ijkw} + \gamma_3 SonDummy_{ijkw} + \eta_{jkw}. \qquad (3)$$

The quantity of children is regressed on the $BirthOrderFemale_{ijkw}$ and $SonDummy_{ijkw}$. The $BirthOrderFemale_{ijkw}$ is a dummy variable which represents the gender of the i^{th} born child in household j from area k and is taken from CHIPS wave w. $BirthOrderFemale_{ijkw}$ is equal to zero if the i^{th} born child is female, and is equal to one if it is male. To differentiate between cases where a son joined different households at a different birth order, a dummy type variable $SonDummy_{ijk}$ was introduced into the regressions. The $SonDummy_{ijk}$ was assigned a value of either zero or one to each child depending on whether a son has been born to a household. All female children born before the first-born male were assigned a value of zero, and once a son had been born to a household all subsequent births, including the first-born male, were assigned with a value of one. The merit of this variable is that it signifies the impact of the first-born male on a household's decision of how many children to have. A significant $SonDummy_{ijk}$ indicates strongly that the level of a son preference is important in determining the household's decision about the number of children to have. Using data selected by a birth order variable[23] for each round of estimations, we were able to uncover the impact of gender on the quantity of children at different birth orders. The gender effect on the quantity of children is validated by 1) both the gender of the child at each birth and the SonDummy or 2) independently by the two. Having no son motivates a

[23] Our samples are unbalanced cross-sectional data where households differ in the number of children. The birth order variable, namely the variable labelled as id, is used to indicate the order of birth of each child in a household and serves as a data selection criteria in the regression. In other words, only data at the same level of birth order is selected for use in the regression. By doing so, we were able to identify the impact of gender on the quantity of children at each birth. The choice of whether to have another child is dependent on budget constraints, so even in a household with a very strong son preference the gender effect on the quantity of children is expected to gradually disappear as its number of children grow. For id=1, the corresponding dataset used in the regression contains households with one or more children, while when id=2, the dataset used in the regression contains households with two or more children, and so on.

household to have an additional child in the hope that their son preference can be realized. Consequently, a child's gender at each birth could have a noticeable influence on the total number of children produced by a household with a son preference.

IV. Estimations

4.1 The effect of a son preference on the quality of children

The children in the 1988 and 1995 samples only include those who were over 18 years old, therefore, the results derived using these two samples are largely free from the impact of the one-child policy and the 'Nine Year Compulsory Education System', and at the age of 18 it is very likely that the individual has completed formal education given that the average education level in China was quite low prior to 1978 (=1995-18). We report the estimation results in both Table 1 and 2[24].

The results concerning rural populations are reported in Table 1. In 1988, on average, girls from rural areas received at least one and a half years [25] less education than boys, and about a half year less in 1995. We interpret these findings as strong evidence that a child's gender matters for their level of education, given that the average schooling was only approximately 7.3[26] years in both 1988 and 1995. What is more striking is that before 2002[27], a child's gender was, in general, found to have played a more decisive role than the quantity of children in determining a child's educational attainment. Although the quantity of children has significant coefficients of between – 0.221 and – 0.037 for most subsamples, these numbers, on average, suggest a much smaller

[24] The estimated results without using the indicators for age groups are presented in the Table A2 and A3 in the appendix. Those results are not qualitatively different from that has been reported in both Table 1 and 2 respectively.
[25] This is calculated as $0.496 \times 3 \approx 1.5$, where 3 is the average number of years for each education level.
[26] 7.3 is converted from the level of education specified in table A1 in the appendix to the number years of schooling.
[27] The reported coefficients in Table 1 are all at mean level. The largest families had eight, six and five children respectively in the 1988, 1995 and 2002 surveys, while the mean number of children is below four, three and two correspondingly. Therefore, the largest impact of the quantity of children on the quality of children is well below the magnitude of the impact of gender on the quality of children, even when we increased the number of children from its mean value to its maximum value.

reduction in the average educational attainment of the children than is the case for gender in the earlier years' samples.

Table 1
The effect of a Son Preference on the Quality of Children for Rural Populations [a] – with indicator variables for age groups

	1988		1995		2002	
	Larger Sample	Multiple Births	Larger Sample	Multiple Births	Larger Sample	Multiple Births
Gender	– 0.496*** (.0271)	– 0.808*** (.0988)	– 0.171*** (.0212)	– 0.218* (.1394)	– 0.062** (.0267)	– 0.203 (.1515)
Quantity	– 0.061*** (.0107)	– 0.099*** (.0372)	– 0.037*** (.0117)	– 0.221** (.0889)	– 0.117*** (.0155)	0.029 (.1350)
Education (Father)	0.165*** (.0083)	0.117*** (.0320)	0.131*** (.0097)	0.015 (.0878)	0.219*** (.0122)	0.1282 (.1080)
Minority	– 0.239*** (.0514)	– 0.158 (.2039)	0.279*** (.0403)	0.996** (.4095)	0.205*** (.0369)	0.325 (.2689)
Family income	0.00003*** (.000005)	0.00005*** (.00001)	0.00001*** (.000002)	0.00001 (.00001)	0.00002*** (.000001)	0.00002 (.00001)
Terrain[28]	– 0.055*** (.0201)	– 0.318*** (.0795)	– 0.046*** (.0145)	– 0.059 (.1438)		
Geographic location[29]	0.050 (.0579)	– 0.085 (.2294)				
Old Revolutionary Areas	– 0.027 (.0423)	– 0.019 (.1325)	– 0.013 (.0259)	– 0.2478 (.2774)		
Suburb	0.183 (.1087)	0.576* (.2962)	– 0.270*** (.0515)	– 0.502* (.2855)		
Impoverished Areas	– 0.274*** (.0393)	– 0.158** (.0677)	– 0.110*** (.0213)	– 0.589*** (.2198)		
Province	– 0.008*** (0.0011)	– 0.004 (0.0044)	– 0.005*** (.0008)	0.009 (.0079)	0.000000001 (.000000009)	– 0.0000003 (.0000007)
N	7,515	697	7,221	119	6,963	105

[a] Data from the Chinese Household Income Project Surveys. Quality of children is the dependent variable. Multiple-birth samples were abstracted from the original larger samples. Standard errors in parentheses. ***, **, * indicates statistical significance at 1%, 5% and 10%, respectively.

[28] Terrain includes the land features: 1. Plain; 2. Plateau; and 3. Mountainous region. Data was not available for the 1995 and 2002 samples.

[29] Geographic location includes: 1. Lakeside regions; 2. Coastal regions; 3. others. Data was not available for the 1995 and 2002 samples.

As the one-child policy abruptly intervened in the process of human reproduction by imposing a limit on the number of children a household could have, the insignificant gender effect identified in the 2002 sample is not necessarily an indication that a significant change had taken place in the son preference culture in rural China, but rather is an indication of the effect of the one-child policy in limiting the number of children, while the 'Nine Year Compulsory Education System' left rural households with little room to exercise their son preference by differentiating their children's level of educational attainment based on their gender.

As a robustness check, we used multiple-birth samples which only contained twins, triplets and so on (e.g. members of each set of twins, triplets, etc., were required to be opposite sexes) to test for the son preference effect on the quality of children. Gender is confirmed to be the single most important factor used by parents in rural China in determining each of their children's level of educational attainment. We believe that the high dropout rates of girls had nothing to do with the level of difficulty of subjects at school, given that most of the children in the 1988 and 1995 samples only completed their education at a level below senior middle school. We conclude that the additional number of years' education provided to male children from rural families was an intentional act by parents at the cost of their female children. The human capital loss of the female children is not an accidental coincidence given that twins, triplets and so on of opposite sexes are, on average, not very different from each other in their inheritable genomes, and hence their ability to complete a basic level of education[30]. The significant loss in human capital, as a result of receiving an inadequate level of education, is very likely to be an adverse consequence of the son preference. The results are remarkably robust and are in line with what we expected.

It is important to note that all independent variables in Table 1 have reasonably small correlations with each other, and more importantly, the gender variable has small absolute correlations of approximately 0.10 with any of the other independent variables. The results concerning the effect made by gender on the quality of children remain unchanged and significant if we drop any or all of the independent variables, which further confirms the robustness of the explanatory power of the gender variable on the quality of children.

[30] It comprises the two stages primary education and lower secondary education.

Table 2
The effect of a Son Preference on the Quality of Children for Urban Populations [a] – with indicator variables for age groups

	1988		1995		2002	
	Larger sample	Multiple Births Sample	Larger sample	Multiple Births Sample	Larger sample	Multiple Births Sample
Gender	– 0.041	– 0.163	0.105***	0.158	0.275**	
	(.0319)	(.1773)	(.0395)	(.3873)	(.0568)	
Quantity	– 0.047**	– 0.047	– 0.128***		– 0.305***	
	(.0196)	(.1298)	(.0294)		(.0532)	
Education (Father)	0.161***	– 0.061	0.199***	0.216	0.229***	
	(.0100)	(.0796)	(.0127)	(.2494)	(.0212)	
Minority	– 0.247***	0.580	0.024		– 0.050	
	(.0285)	(.3543)	(.0891)		(.1155)	
Family income	0.000005	0.0001	0.00002***	– 0.00004	0.00001***	
	(.0000004)	(.0001)	(.000004)	(.00006)	(.000004)	
Residential location [b]	0.017	– 0.474***	– 0.032	– 0.3493	– 0.148***	
	(.0211)	(.1683)	(.0275)	(.6791)	(.0400)	
Province	– 0.031***	0.003	– 0.003**	0.027	– 0.001	
	(.0012)	(.0076)	(.0013)	(.0182)	(.0018)	
N	3,689	106	3,336	38[31]	2,213	Insufficient

[a] Data from the Chinese Household Income Project Surveys.
[b] Residential location is specified as: 1. City center; 2. City; 3. Suburbia; 4. Exurbia.
Quality of children is the dependent variable. Multiple-birth samples were abstracted from the original larger samples. Standard errors in parentheses. ***, **, * indicates statistical significance at 1%, 5% and 10%, respectively.

For the rural sample, the quantity of children was found to exert significantly negative effects on the quality of children, but the magnitude of its effect on the quality of children was overshadowed by the gender effect. Children from households with fathers who had a higher level of educational attainment on average received relatively higher levels of education. The government's preferential treatment of minority ethnic groups took effect on the quality of children in the later waves of CHIPS. Meanwhile, children who lived in mountainous areas, impoverished areas and less economically developed provinces[32] attained lower levels of education compared with other children living in better economically developed areas of China.

[31] For the multiple-birth sample, we intentionally dropped the minority variable to ensure a sufficient number of observations could be used for the regression analysis.
[32] Provinces are ordered in their level of economic development, so the province with higher rank has lower level of economic development.

The results concerning urban populations are presented in Table 2. Contrary to the findings from the rural samples, a son preference was not confirmed to have a significant impact on the quality of children in our urban samples. More specifically, no significant difference between the educational attainment of male children and female children was evident from the 1988 CHIPS data, while female children even obtained slightly higher levels of educational attainment than male children in the 1995 CHIPS data, and in 2002 they had significantly surpassed male children in their level of educational attainment. We believe that this change from females being at a significant disadvantage in their education in rural areas, to females enjoying more equality in urban areas is a result of the shift from a very strong son preference environment to an environment which is more gender equal and/or gender tolerant. These contrasting results between the rural and urban areas sheds some light on the merit of our use of gender to capture the effect of a son preference on the quality of children. The educational attainment of a person who lives in an urban area is more likely to be determined by the child's keenness to study and perform at school, as well as their family's attitude towards education and the household resources. By the comparison of the educational attainment of children of both genders and from rural and urban areas, a significant negative gender effect on the quality of children provides strong evidence of the presence of a son preference in rural China.

4.2 The effect of a son preference on the quantity of children

The total number of children in a household is likely to be determined by its economic wellbeing, the expected quality of the children, as well as its associated demographic, social and cultural factors, and other random factors. The total number of children to have is unlikely to be decided at the outset of a marriage, and the ideal and realized family size may differ significantly from each other. The process, from the first-born to the last-born, will take many years to complete in a changing environment. Our findings are able to shed some light on the potential mechanisms underlying these choices.

Although gender is found to have a positive effect for both the rural and urban samples, this does not necessarily imply that a son preference existed in both areas. This is because, for instance, a mixed sex composition of their children is something preferred by people from very different

cultures, which is not specified in the reduced form model (2) used to estimate the results in Table 3. We argue that the results in Table 3 may be informative about the sign of the relationship between child's gender and the quantity of children, but not about its magnitude, as the elasticity also depends on the birth order of the first-born son.

Table 3

The effect of a Son Preference on the Quantity of Children

	1988		1995		2002	
	Rural	Urban	Rural	Urban	Rural	Urban
Gender	0.220***	0.018	0.175***	0.093***	0.176***	0.085***
	(.0310)	(.0238)	(.0211)	(.0231)	(.0201)	(.0223)
Quality	– 0.061***	– 0.074***	– 0.040***	– 0.021**	– 0.066***	– 0.045***
	(.0128)	(.0097)	(0.0119)	(.0101)	(.0092)	(.0079)
Education	0.034***	0.096***	– 0.041***	0.008	– 0.018*	0.003
(Father)	(.0095)	(.0087)	(.0098)	(0.0077)	(.0096)	(.0086)
Minority	0.359***	0.072***	– 0.174***	0.040	– 0.206***	0.133***
	(.0584)	(.0242)	(.0410)	(.0529)	(.0286)	(.0462)
Family income	0.0001***	0.000001***	0.00002***	– 0.00003	0.00001***	0.000008
	(.000006)	(.0000003)	(.000002)	(.00003)	(0.000001)	(.000001)
Terrain	0.011		0.039***			
	(.0227)		(.0147)			
Old	– 0.118***		0.012			
Revolutionary	(.0476)		(.0264)			
Areas						
Suburb	– 0.092		– 0.085***			
	(.1224)		(.0524)			
Impoverished	0.198***		– 0.104***			
Areas	(.0447)		(.0217)			
Residential		0.040**		– 0.013		0.042***
location		(.0184)		(.0164)		(.0160)
PROVINCE	– 0.013***	– 0.014***	0.005***	0.002**	0.0000008***	0.002***
	(0.0012)	(.0010)	(0.0008)	(.0007)	(.00000007)	(.0007)
N	7,515	4,397	7,221	3,340	6,988	2,230

[a] Data from the Chinese Household Income Project Surveys. Standard errors in parentheses. Quantity of children is the dependent variable. ***, **, * indicates statistical significance at 1%, 5% and 10%, respectively.

Given that the results regarding the gender effect on the quantity of children remain qualitatively unchanged, with or without the other independent variables, this enabled us to only make use of a more parsimonious model (3) compared to that used in Table 3 for the estimation of the gender effect on the quantity of children.

A nonlinear Poisson regression[33] and sorting the sample by birth orders were employed to account for the fact that the dependent variable is a count measure and a probabilistic feature of the decision making faced by each household when deciding on the number of children to have. The results are presented in Table 4. It is important to note that because we have used children of all ages in estimating the results in Table 4, the one child policy itself could have a negative impact on the quantity of children. Therefore, some cautions need to be taken in interpreting the findings.

Each column reports the estimation results specifically based on the observations corresponding to each birth order. The progressive nature of the decision making for the number of children is well captured by both independent variables. For the independent variables, Row1 reports their coefficients, Row 2 reports the clustered robust standard error, and Row 3 reports the marginal effects in a semi-elasticity form. For indicator variables, such as our two independent variables, the marginal effect represents the percentage change in the probability of having more children when the indicator variable moves from zero to one. Overall, the previous results concerning gender on the quantity of children remained qualitatively unchanged. With regard to the urban population, the gender effect for each birth order on the quantity of children is found to be insignificant across the three waves of urban samples (see columns (4), (5) and (6) in Table 4).

[33] The equality of its mean and variance of count data processes is not a reasonable assumption, which is assumed to be true in the Poisson regression. Its presence renders untenable the assumption of a Poisson distribution for the error process. A reasonable alternative is negative binomial regression, which allows the variance to differ from the mean. The estimation results from using the negative binomial regression are not qualitatively different from the Poisson regression.

For rural households where the first-born is a girl, the estimates in column (1) suggest that their probability of having two or more children is significantly higher than that of the first-born son families across the three wave of CHIPS, with these probabilities ranging from 16.9% to 30%. It is interesting to note that the probability is higher in the later years' samples across the three surveys. Three observations are in place: first, the higher probability of rural households having two or more children in the later years' samples, if the first child is a girl, clearly indicates that the level of son preference is not decreasing in China's rural areas; second, a sharp rise in income within a short space of time does not necessarily lead to a sizable reduction in the level of son preference in societies with a traditionally strong son preference; third, the one-child policy has aggravated the urgency of those households with a son preference to have two or more children to realize having a son, even if that meant breaching the one-child policy, if their first child was a girl. This also explains why the one-child policy was not effective in reducing China's fertility rate during its earlier years (e.g. Zhang 2017; Wang 2019).

The estimates in column (2) imply that, for households in which the first two children are girls, the probability of having three or more children is higher than for those families which already had a son. The probability ranges from 0.3% to15.2% across the three waves of CHIPS, where the earlier years' rural samples are associated with a larger probability. The later years' rural samples having a smaller or even statistically insignificant probability of having three or more children is intuitive, because the one-child policy was implemented more strictly in its later years and the authorities kept a close eye on those who had already breached the policy. Severe punishments, ranging from stiff financial fines, the confiscation of property, and even forced abortions (Jimmerson 1990) were imposed on many such "lawbreakers". As a consequence, we suspect that the statistically insignificant probabilities of having three or more children in the later years' sample is a result of the one-child policy. By comparison, using the U.S. census data, Dahl and Moretti (2008) reported a statistically insignificant probability of having two or more children, if the first child was a girl, and reported a probability of having three or more children at only 0.6% higher than the first-born son families[34]. The estimates in column (3) suggest that, in households

[34] Although, they did not report the probability of having three or more children for households in which the first two children were girls, we believe the probability will be increased only slightly, and it is still much smaller than ours. The contrasting difference between our results and those of Dahl and Moretti (2008) can be explained by the fact that

with no son, the probability of having four or more children was higher than for those families in which at least one of their first three children was a boy, with the probability ranging from 9.1% to13.4% across the three waves of CHIPS, although this was only significant statistically in 1988[35].

Table 4
The effect of a Son Preference on the Quantity of Children – Poisson regression

			1988			
	Rural			Urban		
	(1)	(2)	(3)	(4)	(5)	(6)
Birth order	0.063***			0.010		
(of girl)	(.0124)			(.0181)		
1	**0.169**			**0.016**		
		0.051***			0.020	
2		(.0147)			(.0263)	
		0.152			**0.046**	
3			0.034*			0.044
			(.0212)			(.0442)
			0.124			**0.142**
SonDummy		– 0.081***	– 0.061***		– 0.056*	0.039
		(.0178)	(.0212)		(.0309)	(.0513)
		– 0.243	**– 0.227**		**0.130**	**0.126**
Log likelihood	– 13547.88	– 13537.64	– 8183.21	–10201.91	– 4973.82	– 1460
VIF[36]		1.77	1.96	1.00	1.86	1.50
N	9,653	8,212	4,857	7,644	3,469	936
			1995			
Birth order	0.143***			0.011		
(of girl)	(.0124)			(.0262)		
1	**0.295**			**0.012**		
		0.018			-0.002	
2		(.0209)			(.0713)	
		0.045			**-0.004**	

the level of son preference in China is much stronger than in the U.S., and the social environment in China is more tolerant towards gender inequality than in the U.S.

[35] We suspect that the statistically insignificant probability of having four or more children in 1995 and 2002 was a result of the one-child policy.

[36] Variance inflation factor (VIF) is often considered superior to examining the bivariate correlations. A commonly accepted rule of thumb is that VIFs of 10 or higher prompt concerns for high correlations between independent variables. The reported VIFs are much smaller than 10, hence the potential bivariate correlations between the two indicator variables is not a concern for us.

3			0.028			
			(.0299)			
			0.091			
Son Dummy		– 0.169***	– 0.048		– 0.015	
		(.0251)	(.0356)		(.0891)	
		– 0.418	**-0.160**		**– 0.032**	
Log likelihood	– 10991.94	– 7827.63	– 3070.23	– 5662.77	– 710.84	
VIF		1.38	1.43		1.39	
N	7,515	5,301	1,942	5,294	5,37	Insufficient
			2002			
Birth order	0.160***			0.009		
(of girl)	(.0163)			(.0240)		
1	**0.300**			**0.012**		
		0.001			0.0004	
2		(.0215)			(.0467)	
		0.003			**0.0008**	
3			0.04			0.011
			(.0325)			(.1431)
			0.134			**0.033**
SonDummy		-0.134***	0.017		– 0.028	– 0.015
		(.0265)	(.0429)		(.0552)	(.1635)
		– 0.324	**0.056**		**0.059**	**– 0.047**
Log likelihood	– 11767.75	– 7333.53	– 2469.91	– 6465.59	– 1632.09	– 171.75
VIF		1.35	1.36		1.38	1.76
N	8,312	5,010	1,556	5,635	1,211	113

[a] Data from the Chinese Household Income Project Surveys. Standard errors in parentheses. Quantity of children is the dependent variable. Top panel: Poisson regression for 1988 for each birth order and for both rural and urban samples separately. Middle panel: Poisson regression for 1995 for each birth order and for both rural and urban samples separately. Lowest panel: Poisson regression for 1995 for each birth order and for both rural and urban samples separately. ***, **, * indicates statistical significance at 1%, 5% and 10%, respectively.

The estimates in column (2) concerning the SonDummy variable suggest that having a son in the second birth significantly reduces rural households' probability of having three or more children ranges from 24.3% to 41.8%, compare to households with their first two children are girls. The estimates in column (3) suggest that, having a son in their third birth further reduces rural households' probability of having more children ranges from 5.6% to 22.7% compare to

households with their first three children are girls, although it is only statistically significant in 1988 rural sample.

V. Model

The reduced form analysis establishes the responsiveness of important decisions in the quantity and quality of children to changes in sex composition of their children and the birth of the first son. However, empirical findings have little information about the mechanism underlying the decisions of households. The model we have developed allows us to uncover the underlying mechanism of how a son preference affects both the quantity and the quality of children, and how a son preference changes the interactions between the two. To capture analytically the effects of a son preference on the quantity and quality of children, a model is required which is embedded within this heterogeneity of fertility and differentiated quality of children across households with or without a son preference. Therefore, it seems useful to go beyond Becker and his coauthors' models to account for the issues arising out of differences in household choices due specifically to their different levels of son preference.

Similar to Becker and Tomes (1976), we assume that a household has a utility function with the following form:

$$U = U(n, Q, y) \tag{4}$$

where n is the number of children, Q represents quality of each child in the household, and y the aggregate amount of all other commodities. A household faces the following budgetary constraints:

$$I = p_y y + p_q n q, \tag{5}$$

where p_y is the price of y, p_q the average cost of increasing q by one unit, and I is the total family income. Therefore, $p_y y$ is the total expenditure on the aggregate amount of all other commodities, and $p_q n q$ is the total cost of children.

For a household with no son preference, their decision on the number of children is defined as

$$n = \underbrace{\lfloor \bar{n}(\xi) \rfloor}_{(a)} \text{ or } \underbrace{\lceil \bar{n}(\xi) \rceil}_{(b)}.$$

Here, $\bar{n}(\xi)$is the average number of children born into households which possess no son preference and share a similar characteristic ξ, such as geographic location, income and parents' education level, where the floor and ceiling function is defined respectively as $\lfloor x \rfloor = \max\{m \in \mathbb{z} | \, m \leq x\}$ and $\lceil x \rceil = \min\{m \in \mathbb{z} | \, m \geq x\}$.

For households with a son preference, we have assumed their decision about the number of children to have is jointly determined by the sex composition of their children after each birth[37] and other controlling factors (e.g. budget constraints, educational resources). In the following, we have assumed that once a son has been born into households with a son preference, they will behave very similarly in their fertility rate to those households with no gender preference. We propose that such households' decisions about the number of children to have is governed by the following equation, based on the rationale that more births are expected for households with a son preference where no son has already been born,

$$n = \underbrace{\left\lfloor \bar{n}(\xi) \times \left(1 + 1_{Son}\left(\bar{n}(\xi)\right) \times S\right) \right\rfloor}_{(a)} \text{ or } \underbrace{\left\lceil \bar{n}(\xi) \times \left(1 + 1_{Son}\left(\bar{n}(\xi)\right) \times S\right) \right\rceil}_{(b)} \qquad (6)$$

where $1_{Son}\left(\bar{n}(\xi)\right)$ is an indicator function which equals one if none of the children in $\bar{n}(\xi)$ is a boy, and otherwise it equals zero. Therefore, the term $\bar{n}(\xi) \times 1_{Son}\left(\bar{n}(\xi)\right) \times S$ measures the number of additional children[38] that households with a son preference intend to have to realize their son preference compared to those households with no son preference and possess similar characteristic ξ. Here, S signifies the level of a son preference, which is defined as:

[37] For example, the decision whether to have an additional child for households with a son preference depends on whether a son has been born into the households.

[38] The additional number of children is in the range of $[0, \lceil S \times \bar{n}(\xi) \rceil]$, where the upper bound of the interval depends on the households' level of son preference. It is equal to zero when there is already a son born into the family. Assuming that no son has been born into households with a son preference, on average, the additional numbers of children will be translated into a probability of having at least one son, which is the range of $[0.75, 1 - \left(\frac{1}{2}\right)^{\lceil S \times \bar{n}(\xi) \rceil}]$. For illustrative purposes, we assume an equal probability of having a son and of having a daughter. In reality, the probability of having a son and a daughter can differ significantly at different birth orders.

$$S \in (0, \eta); \quad \text{if a household possesses a son preference} \qquad (7)$$

$$S = 0; \quad \text{otherwise.}$$

An individual's level of S is partially determined by the level of intrusiveness of their surrounding environment in personal matters and the social attitude towards genders. The appreciation of the expected future value of a son is formed through learning and interactions with members of social communities.

We modify Becker and Tomes's (1976) assumption that the "the quality of each child is the same" to assume that the quality of each child is either perceived very similarly or very differently by each household, depending on their level of son preference. The households' utilities are, to some extent, determined by the level of their son preference. Following this line of reasoning, we differentiate our definition of the quality of children from Becker and Tomes (1976) by incorporating the effect of a son preference on the perceived quality of a child from the parents' perspective. This is defined as:

$$Q_i = e + \frac{\sum_{j=1}^{n} q_j}{n}(1 + 1_{Son}(i) \times S) = e + q + q \times S \times 1_{Son}(i) \qquad (8)$$

where Q_i is (an element in the set Q) the quality of the i^{th} child, $q\left(=\frac{\sum_{j=1}^{n} q_j}{n}\right)$ captures the households' intention to contribute equal levels of quality to each of their children, and e (i.e. the innate ability – the endowment) is assumed to be fixed for a household and it is considered as exogenous and independent of q. In addition, $1_{Son}(i)$ is an indicator function which equals one if the i^{th} child is a boy, and otherwise it equals zero. Therefore, equation (5) reveals the parents' differentiation in the quality of their children based on their gender, where a female child receives approximately an average level of the households' contribution (which is assumed to be equal to q) towards their quality, but a male child receives preferential treatment subject to the parents' level of son preference. The term $q \times S$ represents a household's perceived additional "quality"[39] of a son compared to a daughter.

[39] This is the quality perceived by parents due to their ability to continue the family line and bringing honor to the family and/or ancestors, and the higher earning power of male children compared to female children.

For a given level of son preference[40] and a fixed p_q, the household chooses the aggregate amount of all other commodities, the number of children n, and the level of quality of children Q, to maximize the utility function in equation (4), subject to the budget constraints (5), son preference conditions (7), and the quality function (8). The maximization yields[41]:

$$\begin{aligned} MU_y &= \lambda p_y = \lambda \pi_y \\ MU_{Q_i} &= \frac{\lambda n p_q}{1+S\times 1_{Son}(i)} = \frac{\lambda \pi_Q}{1+S\times 1_{Son}(i)} = \lambda \pi'_Q \\ MU_n &= \lambda q p_q = \lambda \pi_n \\ MU_S &= MU_n \times \bar{n}(\xi) \times 1_{Son}\big(\bar{n}(\xi)\big) \times 1_B(\beta) + MU_{Q_i} \times q \times 1_{Son}(i) \end{aligned} \tag{9}$$

where MU's are the marginal utilities, λ is the marginal utility of income, π'_Q is the "adjusted" shadow price of the quality of children dependent on the household's level of son preference, and π_n is the shadow price of the quantity of children. It is important to note that our π'_Q can be significantly smaller than the shadow price of the quality of children π_Q $(= np_q)$ defined in Becker and Tomes (1976), and the difference between π'_Q and π_Q depends on the household's level of son preference. The impact of preferences for male offspring to female offspring on sex ratio at birth is obvious in son preference cultures, such as China and South Korea – as evidenced in their heavily skewed sex ratio at birth in China (since 1990) during the one-child policy as well as in South Korea from 1987-2006 (e.g. Wang 2019). This observation of a strong male preference at birth reinforces the importance of the potential discrepancy between π'_Q and π_Q as a result of a son preference.

Although both the shadow price of the quality of children, π'_Q and π_Q, are directly related to the quantity of children, π'_Q is deemed to be less sensitive to the quantity of children than the π_Q derived under the framework of Becker and Tomes (1976). An increase in the quantity of children inevitably increases the expenditure on children, however, to accommodate the increase in the cost of raising more children when more births appear to be a *sine qua non* for having a son, households

[40] The level of son preference can range from it being better to have sons than daughters, to it being essential to have at least one son.

[41] We adopt here the notations from Becker and Tomes (1976) for the purposes of making a comparison between our construction and theirs.

with a strong son preference are keener than other households on redistributing their resources among their children based on their gender. When it becomes economically unsustainable because of the large number of (female) children born before a son, households with a strong son preference may either embrace sex-selective abortions and/or the abandonment of baby girls (when the legal and social environments are tolerant of abortions and the abandonment of baby girls), or substantially reduce investment in their daughters to have one more child for the purposes of having a son. The additional number of children a household plans to have is subject both to their level of son preference and to any budget constraints, all else being equal.

The shadow price of the quantity of children, π_n, is related to the quality of children. The higher the quality of children a household aims to achieve, the higher the cost of an additional child. For households with a son preference, for the purposes of having a son, they may reduce their contribution to the quality of their daughters to offset the increase in cost because of the need to have more children before having a son. Consequently, each daughter in such a household is likely to receive a discounted household contribution towards her quality, where the severity of the discount will depend on the level of son preference of the household. We found that the son preference reduces the social income[42] of a son-preference household by $\frac{Se\pi_Q}{1+S}$, which signifies the potential loss of the household's genetic endowment due to the son preference. Furthermore, it is important to note that the endowment loss of $\frac{Se\pi_Q}{1+S}$ is related to the household's total number of children, its level of genetic endowment and the son preference. For a given level of endowment, an increase in both (either) the number of children and (or) the level of son preference can cause substantial loss in social income as raising the quality of more children becomes more expensive, which further intensifies the unequal distribution of resources between sons and daughters. Considering a society as a whole, its aggregate loss in social income from households with a son preference can be enormous[43]. In addition, as a persistent and strong son preference culture

[42] Household's social income is defined as $\pi_y y + \pi_n n + \pi'_Q Q = I + \pi'_Q Q$. Hence, the difference between the two social incomes with and without son preference equals to the difference between $\pi'_Q Q = \pi_Q(e+q)$ for $S = 0$, and $\pi'_Q Q = \frac{\pi_Q}{1+S}\big(e + q(1+S)\big)$, that is $\pi_Q(e+q) - \frac{\pi_Q}{1+S}\big(e + q(1+S)\big) = \frac{Se\pi_Q}{1+S}$.

[43] That loss would depend on both the number of children born into the households with strong son preference, their level of endowment and the severity of their son preference.

increases the inequality of human capital accumulation among males and females, it is not surprising that this lowers the average level of education and, therefore, economic growth.

For households with a son preference, an increase in the market price p_q has a more profound impact on the quality of children than on the quantity of children. This is because such households are more likely to further reduce their contribution to the quality of their children[44], especially their daughters, to offset the impact due to the increase in the market price p_q on the realization of their son preference, rather than to reduce the number of children before they have a son. It is those households with larger genetic endowments and which possess a strong son preference which experience larger losses in their social income. These losses can be explained by the fact that households with a son preference are very reluctant[45] to invest in their daughters (e.g. in their education). The implications from our model provide further support for empirical findings which suggest that gender inequality in education (i.e. the contribution of a household to the quality of their children) lowers the average level of human capital (e.g. Klasen 2002).

The marginal utility from the realization of a son preference is summarized by $\mathrm{MU_S}$. For a household with no son preference, having a son brings no additional utility to the family than having a daughter. However, for a household with a son preference, the additional utility generated from the realization of their son preference can be extremely large, with its size dependent on their level of son preference, which is well captured by the term $\mathrm{MU_S}$ in equation (9).

Take a two-child household with a son preference having both a son and a daughter as an example, the bias of the parents towards their children is captured by equations (10) and (11):

$$Q_s = e + q(1 + S) \tag{10}$$

$$Q_d = e + q \tag{11}$$

[44] Based on equation (8), we have $\frac{\partial Q}{\partial n} = -q - \frac{q}{n}S < 0$ (assuming the total quality of children Q is a linear function of Q_i), which shows that the existence of a son preference intensifies the sensitivity of the quality of children to the increase in the quantity of children. There are sufficient reasons to believe that this additional amount of negative impact of quantity on the quality of children due to a son preference will be largely passed down to the female children.

[45] Those with a strong son preference uphold the belief that "raising a daughter is like watering your neighbors' garden".

where Q_s and Q_d is the parents' perceived total quality of a son and a daughter, respectively, assuming both a son and a daughter possess the same endowment e, which provides them with the ability to achieve the quality q. Consequently, the difference between Q_s and Q_d is simply the result of the parents' son preference. This difference is translated into the divergence in the marginal utility with respect to the quality of a son and of a daughter, subject to the budget constraints (5):

$$MU_{Q_s} = \frac{2\lambda p_q}{1+S} = 2\lambda p_q' \tag{12}$$

$$MU_{Q_d} = 2\lambda p_q$$

where MU_Q are marginal utilities of quality. Although the market charges equally (p_q) with regard to the cost of adding one unit of increase to the quality, irrespective of whether it is a boy or a girl, a household with a son preference considers that the cost of adding to the quality of a son is "worthier" or "cheaper" than it would be for a daughter[46], based on equation (12). From equation (12), we also have:

$$\frac{MU_{Q_s}}{MU_{Q_d}} = \frac{1}{1+S} < 1. \tag{13}$$

The perceived costs of adding to the quality of a son were negatively related to the son preference, while the desired quality of a son is positively related to the parents' preference. The paradox of value captured in the inequality (13) can be explained by the fact that households with a son preference tend to have higher expectations of sons than daughters, and the inequality (13) is theoretically justified by the law of diminishing returns of utility, given that parents place the quality of sons in a higher location on the utility curve than they do for daughters. It is in the interests of parents with a son preference to invest significantly more in their sons than their daughters. These households' stronger willingness to add more quality to their sons than to their

[46] Our finding differs from Becker and Tomes's (1976), as we do not assume a "child-neutral preference", but rather incorporate the potential impact of a son preference on the subsequent development of the quality of children of different genders. This finding is analogous to consumers who are willing to pay more for the brand of their preference than for other brands with indistinguishable quality in terms of such as functionality (but not in the quality associated with brand image).

daughters inevitably reinforces the difference in the quality of sons and daughters if all else holds equal[47].

Some, but not all, parameters of the model can be estimated using the CHIPS data. The CHIPS data did not permit us to identify the level of son preference at an individual level, but did allow us to estimate the son preference on both a rural and an urban level. We relied on the presumption that the preferential treatment towards the education of males is a strong indication of a son preference, and therefore, we were able to measure the strength of the son preference as the difference between the gender effect on the quality of children in the corresponding rural and urban samples. A son preference can differ drastically between rural and urban areas in China, given that urban populations, for example, enjoy more job opportunities, and much more advanced educational and healthcare resources, and are influenced more by foreign cultures (which promote gender equality) than rural populations. In estimating the level of son preference which exists in rural China, we have taken the effect of gender on the quality of children in the corresponding urban sample as the benchmark (Table 2). The level of son preference for each of the three waves of CHIPS was estimated as follows:

$$S_w = |S_{Rural,w} - S_{Urban,w}| = |\alpha_{2,Rural,w} - \alpha_{2,Urban,w}|.$$

The estimated son preference in rural China ranges from 0.20 to almost 1[48] across the three waves of CHIPS, where earlier years' rural samples are associated with a stronger son preference. For each sample, the average quantity of children $\bar{n}(\xi)$ and the average quality of children q [49], after controlling some of the characteristics ξ, can be easily calculated using the summary statistics in table A1 in the Appendix. However, it is not feasible to estimate λ and p_q using our dataset.

[47] Here, all else being equal means that children of a household inherit equal endowment, and these children put the same effort into improving their own quality.

[48] This is the difference between the effect of gender on the quality of children in a rural sample and its corresponding urban sample.

[49] Though the genetic endowment e is not observable, it is not really vitally important to the main issues studies in the paper, and it can be removed from the equation simply by subtracting it from each individual overall quality Q. As the son preference is really impacting on q then on the overall quality Q, therefore, our main interest lies in the estimated average quality q rather than the overall quality Q. In the table A1 where the average quality q is captured by the variable Edu which represent the level of education rather than years of schoolings.

VI. Conclusion

Our theoretical model sheds some light on the underlying mechanisms of how a son preference affects both the quantity and quality of children. The model shows that a son preference plays a critical role in determining the number of children, and the overall as well as the differentiated quality of male and female children.

The empirical results highlight the fact that a son preference can have a strong influence both on the quantity of children and on the distribution of quality among children based on their gender. Overall, our empirical results are well in line with those predicted by our model. We report that female children born to households with a son preference obtain significantly lower levels of educational attainment than their male siblings, and a not yet realized son preference significantly increases the total number of children born into a household. We have found that a son preference exerts significantly more impact than that of the quantity of children on children's quality. Regardless of how our findings are interpreted, however, the serious negative educational effects for female children and the significantly higher likelihood of having more children because of a son preference are interesting in their own right.

Conclusions may be drawn on two fronts regarding the son preference effects on the quantity of children. Firstly, mixed gender preference is an important consideration in determining the number of children to have, and therefore, it is important to distinguish a son preference from a mixed gender preference. Secondly, it is difficult to estimate the effects of a son preference on the quantity of children as a household's decision about the number of children to have is a progressive and dynamic process conditional on, for example, the gender at each birth order and whether a male has been born into a household. Both considerations suggest that great care should be taken when identifying of how a son preference impacts on the quantity of children. In order to do that in our empirical analyses, we first split the data by birth order, then regressed the quantity of children on the children's gender at each birth, and a dummy variable which indicates whether a son has been born into a household. The combinational use of the two variables enabled us to distinguish a son preference from a mixed gender preference. For households with a son preference, having a female

birth at each birth order significantly increased the probability of having more children, with a magnitude much larger than that suggested in Dahl and Moretti (2008). More interestingly, once such households have a son, that household's likelihood of having more children reduces significantly.

References

Ahmad, A., and J. Morduch. 1993 "Identifying Sex Bias in the Allocation of Household Resources". *Harvard Institute for Economic Research* DP 1636.

Almond, D., and L. Edlund. 2008. "Son biased sex ratios in the U.S. 2000 census." *Proceedings of the National Academy of Sciences* 105(15): 5681-5682.

Almond, D., L. Edlund, and K. Milligan. 2013. "Son preference and the Persistence of Culture: Evidence from South and East Asian Immigrants to Canada." *Population and Development Review* 39(1): 75-95.

Andersson, G., K. Hank, and A. Vikat. 2007. "Understanding parental gender preferences in advanced societies: lessons from Sweden and Finland." *Demographic Research* 17(6): 135-156.

Barcellos, S. H., Carvalho, L. S., and A. Lleras-Muney. 2014. "Child gender and parental investments in India: Are boys and girls treated differently?" *American Economic Journal: Applied Economics* 6(1): 157-189.

Barro, R. 1991. "Economic Growth in a Cross-Section of Countries." *Quarterly Journal of Economics* 106(2): 407-443.

Barro, R., and J. Lee. 1994. "Source of Economic Growth." *Carnegie-Rochester Series on Public Policy* 40: 1-46.

Barro, R., and X. Sala-i-Martin. 1995. *Economic Growth*. New York: McGraw-Hill.

Becker, Gary S. 1960. *An Economic Analysis of Fertility. Demographic and Economic Change in Developed Countries*. Princeton: Princeton University Press.

Becker, G. S. and H. G. Lewis. 1973. "On the Interaction between the Quantity and Quality of Children." *Journal of Political Economy* 81(2): S279-S288.

Becker, G. S. and N. Tomes. 1976. "Child Endowments and the Quantity and Quality of Children." *Journal of Political Economy* 84(4): S143-S162.

Black, S., P. Devereux & K. 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.

Bloom, D., and J. Williamson. 1998. "Demographic Transition and Economic Miracles in Emerging Asia." *World Bank Economic Review* 12(3): 419-455.

Burgess, R., and J. Zhuang. 2001. "Modernisation and Son Preference," *London School of Economics Working Paper*.

Chen, L., E. Huq., and S. D'Souza. 1981. "Sex bias in the family allocation of food and health care in rural Bangladesh." *Population and Development Review* 7(1): 55-70.

Chen, Y., H. Li., and L. Meng. 2013. "Prenatal sex selection and missing girls in China: evidence from the diffusion of diagnostic ultrasound." *Journal of Human Resource* 48(1):36-70.

Dahl, G. B., and E. Moretti. 2008. "The Demand for Sons." *Review of Economic Studies* 75(4):1085–120.

Das Gupta, M. 1987. "Selective Discrimination against Female Children in Rural Punjab, India," *Population and Development Review* 13(1): 77–100.

Ding, Y. 2012. "The Problems with Access to Compulsory Education in China and the Effects of the Policy of Direct Subsidies to Students." *Chinese Education & Society* 45(1):13-21.

Dollar, D., and R. Gatti. 1999. "Gender Inequality, Income, and Growth: Are Good Times Good for Women?" Policy Research Report, Engendering Development, Working Paper No. 1. Washington, D.C.

Ebenstein, A. 2010. "The missing girls of China and the unintended consequences of the one child policy." *Journal of Human Resource* 45(1):87-115.

Ebenstein A., and S. Leung. 2010. "Son preference and access to social insurance: evidence from China's rural pension program." *Population and Development Review* 36(1): 47-70.

Forbes, K.J. 2000. "A Reassessment of the relationship between Inequality and Growth." *American Economic Review* 90(4): 869-887.

Galor, O., and D. N. Weil. 1996. “The Gender Gap, Fertility, and Growth.” *American Economic Review* 86(2): 374-387.

Goodkind, Daniel. 1999. “Do parents prefer sons in North Korea?” *Studies in Family Planning* 30(3): 212-218.

_____ 2015. “The claim that China's fertility restrictions contributed to the use of prenatal sex selection: A skeptical reappraisal.” *Population Studies*, 69(3), 269-273.

Hu, A., and F. Tian. 2018. “Still under the ancestors' shadow? Ancestor worship and family formation in contemporary China.” *Demographic Research* 38(1): 1-26.

Jimmerson, J. 1990. “Female infanticide in China: An examination of cultural and legal norms.” *Pacific Basin Law Journal* 8(1), 47-79.

Knowles, S., P.K. Lorgelly, and P.D. Owen. 2002. “Are Educational Gender Gaps a Brake on Economic Development? Some Cross-Country Empirical Evidence.” *Oxford Economic Papers* 54(1):118-149.

Klasen, S. 2002. “Low Schooling for Girls, Slower Growth for All? Cross‐Country Evidence on the Effect of Gender Inequality in Education on Economic Development.” *The World Bank Economic Review* 16(3): 345-373.

Kugler, A.D., and S. Kumar. 2017. “Preference for boys, family size, and educational attainment in India.” *Demography* 54(3): 835-859.

Lagerlöf, N. 1999. “Gender Inequality, Fertility, and Growth.” University of Sydney, Department of Economics, Australia.

Li, J. 1995. “China's Family Planning Program: How, and How Well Has it Worked? A Case Study of Hebei Province, 1979-88.” *Population and Development Review* 21:563-85.

Lorgelly, P. K., and P. D. Owen. 1999. “The Effect of Female and Male Schooling on Economic Growth in the Barro-Lee Model.” *Empirical Economics* 24(3): 537-557.

Park, Chai Bin., and Nam-Hoon Cho. 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.

Park, A., and P. Rukumnuaykit. 2004. "Eat Drink Man Woman: Testing for Gender Bias in China Using Individual Nutrient Intake Data." *University of Michigan Working Paper*.

Qian, N. 2008. "Missing women and the price of tea in China: the effect of income on sex imbalance." *Quarterly Journal of Economics* 123(3):1251-1285.

Qian, N. (2009). *Quantity-quality and the one child policy: The only-child disadvantage in school enrollment in rural China. NBER Working Paper 14973*. Cambridge, MA: National Bureau of Economic Research.

Rosenzweig, M. R., and T. P. Schultz. 1982. "Market Opportunities, Genetic Endowments, and Intrafamily Resource Distribution: Child Survival in Rural India." *American Economic Review* 72(4): 803-815.

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

Schwartz, B. I. 1985. *The world of thought in ancient China*. Cambridge: Harvard University Press.

Sen, A. 1990. *More than 100 Million Women are Missing*. New York Review of Books.

Shi, L. 2009. "Little quilted vests to warm parents' hearts: Redefining the gendered practice of filial piety in rural north-eastern China." *China Quarterly* 198: 348-363.

Taylor, A. 1998. "On the cost of Inward-Looking Development." *Journal of Economic History* 58(1): 1-28.

Thomas, D. 1994. "Like Father, Like Son; Like Mother, Like Daughter: Parental Resources and Child Height." *Journal of Human Resources* 29(4): 950-988.

Wang, Q. 2019. "Missing Women, Gender Imbalance and Sex Ratio at Birth: Why the One-Child Policy Matters." *Working Paper*.

Wasserstrom, J. 1984. "Resistance to the One-Child Family." *Modern China* 10:345-74.

Willis, R. J. 1973. “A New Approach to the Economic Theory of Fertility Behavior.” *Journal of Political Economy* 81(2): S14-S64.

Young, J. 2013. *In China's Hukou System: Markets, Migrants and Institutional Change (pp. 27-64)*. New York, NY: Palgrave MacMillan.

Zhang, C. 2019. “Family support or social support? The role of clan culture.” *Journal of Population Economics* 32(2): 529-549.

Zhang, J. S. 2017. “The Evolution of China's One-Child Policy and Its Effects on Family Outcomes.” *Journal of Economic Perspectives* 31(1): 141-160.

Appendix

Table A1

Summary statistics of key variables for each sample

1988 (Age>15): Data used in Table I, II and III												
			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	3,737	3,823	3,692		707	757	504	216	84	26	6
	Edu	4.38	4.62	4.13		4.51	4.44	4.34	4.24	4.34	4.19	3.76
		(1.268)	(1.085)	(1.389)		(1.210)	(1.179)	(1.273)	(1.340)	(1.475)	(1.438)	(1.432)
Rural					**Birth Order**							
						1th	2nd	3rd	4th	5th	6th	7th
	N					4,737	2,402	1,073	301	78	22	2
	M/F		1.035			1.352	0.806	0.805	0.720	0.857	0.294	
			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	2,744	2,287	2,110	1,483	932	279	42	8			
	Edu	4.59	4.62	4.57	4.66	4.57	4.51	4.63	4.69			
		(1.464)	(1.504)	(1.418)	(1.760)	(1.356)	(1.208)	(0.906)	(0.821)			
Urban					**Birth Order**							
						1th	2nd	3rd	4th	5th	6th	7th
	N					2,744	1,263	330	51	9		
	M/F		1.084			1.229	0.844	1.075	0.759	0.500		
			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	336	337	360		70	46	48	25	11	5	1
	Edu	4.28	4.71	3.87		4.21	4.41	4.42	4.26	3.97	3.94	4.25
Multi-		(1.476)	(1.184)	(1.602)		(1.590)	(1.417)	(1.419)	(1.319)	(1.689)	(1.501)	(1.753)
birth					**Birth Order**							
-Rural						1th	2nd	3rd	4th	5th	6th	7th
	N					336	336	14	11			
	M/F		0.936			0.988	0.888	1.333	0.571			
			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	53	50	56		30	14	8	1			
Multi-	**Edu**	4.91	5.00	4.82		4.92	5.07	4.62	4.50			
birth		(0.991)	(1.195)	(0.765)		(0.996)	(1.152)	(0.619)	(0.707)			
-Urban					**Birth Order**							
						1th	2nd	3rd	4th	5th	6th	7th
	N					53	53					
	M/F		0.893			0.893	0.893					

1995 (Age>15): Data used in Tables I, II and III

			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	4,152	4,133	3,088	1,900	1,585	537	112	16	2		
	Edu	3.92	3.98	3.86	3.99	3.94	3.86	3.85	3.64	4.17		
		(0.930)	(0.898)	(0.966)	(0.888)	(0.913)	(0.964)	(1.006)	(1.275)	(0.389)		
Rural												
						Birth Order						
						1th	2nd	3rd	4th	5th	6th	7th
	N					4,152	2,252	667	130	18	2	
	M/F		1.34			1.658	1.047	0.928	0.940	0.636		

			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	2,453	1,833	1,507	1,477	784	86	18				
	Edu	5.39	5.36	5.43	5.43	5.39	5.22	5.38				
		(1.199)	(1.205)	(1.192)	(1.209)	(1.190)	(1.201)	(1.128)				
Urban												
						Birth Order						
						1th	2nd	3rd	4th	5th	6th	7th
	N					2,453	797	83	7			
	M/F		1.216			1.316	0.963	1.128	1.333			

			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	57	62	57		23	14	17	2	1		
Multi-	**Edu**	3.73	3.82	3.63		3.91	3.79	3.50	3.00	4.00		
birth		(0.880)	(0.820)	(0.938)		(0.784)	(0.820)	(0.961)	(1.414)	(0.000)		
-Rural												
						Birth Order						
						1th	2nd	3rd	4th	5th	6th	7th
	N					57	57	3	2			
	M/F		1.088			0.965	1.192					

			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	19	19	19		16	3					
Multi-	**Edu**	5.29	5.21	5.37		5.44	4.50					
birth		(1.206)	(1.134)	(1.300)		(1.190)	(1.049)					
-Urban												
						Birth Order						
						1th	2nd	3rd	4th	5th	6th	7th
	N					19	18					
	M/F		1.000			0.727	1.375					

2002 (Age>15): Data used in Tables I, II and III

			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	4,736	4,372	2,616	2,956	1,382	332	58	8			
	Edu	4.34	4.34	4.35	4.45	4.32	4.15	4.08	4.27			
		(1.103)	(1.086)	(1.133)	(1.101)	(1.127)	(1.146)	(0.981)	(0.679)			
Rural												
						Birth Order						
						1th	2nd	3rd	4th	5th	6th	7th
	N					4,736	1,780	398	66	8		
	M/F		1.671			2.073	1.070	1.287	0.692	3.00		
			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	2,015	1,256	974	1,696	296	21	2				
	Edu	6.07	5.91	6.27	6.16	5.85	5.63	3.75				
		(1.394)	(1.416)	(1.337)	(1,354)	(1.490)	(1.286)	(0.463)				
Urban												
						Birth Order						
						1th	2nd	3rd	4th	5th	6th	7th
	N					2,015	193	20	2			
	M/F		1.290			1.335	1.053	0.333				
			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	53	52	54		31	14	6	2			
Multi-	**Edu**	4.02	4.12	3.92		3.98	4.04	4.18	4.00			
birth		(0.796)	(0.855)	(0.730)		(0.820)	(0.331)	(1.471)	(0.000)			
-Rural												
						Birth Order						
						1th	2nd	3rd	4th	5th	6th	7th
	N					53	53					
	M/F		0.963			1.944	0.472					
			Children		*No. of Children*							
		Household	***M***	***F***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	N	4	4	4		4						
	Edu		5.75	4.75		5.25						
Multi-			(0.957)	(2.217)		(1.669)						
birth												
-Urban												
						Birth Order						
						1th	2nd	3rd	4th	5th	6th	7th
	N					4	4					
	M/F		1.000			1.000	1.000					

1988 : Data used in Table IV

			Children		Birth Order							
		Household	***M***	***F***	1th	2nd	3rd	4th	5th	6th	7th	8th
Rural	***N***	9,655	13,588	12,462	9,653	8,212	4,857	2,197	795	254	66	15
	M/F		1.090		1.260	0.970	1.048	1.001	1.087	0.924	1.062	1.143
Urban	***N***	7,644	6,298	5,963	7,644	3,469	936	180	29	3		
	M/F		1.056		1.110	0.951	1.013	1.250	0.706	2.000		

1995 : Data used in Table IV

		Household	***M***	***F***	1th	2nd	3rd	4th	5th	6th	7th	8th
Rural	***N***	7,515	8,507	6,819	7,515	5,301	1,942	450	93	21	4	
	M/F		1.247		1.385	1.146	1.072	1.228	0.979	0.909	3.000	
Urban	***N***	5,635	3,683	3,286	5,635	1,211	113	10				
	M/F		1.121		1.164	0.928	1.306	1.000				

2002 : Data used in Table IV

		Household	***M***	***F***	1th	2nd	3rd	4th	5th	6th	7th	8th
Rural	***N***	8,312	8,940	6,435	8,312	5,010	1,556	395	87	14	1	
	M/F		1.389		1.578	1.147	1.301	1.351	1.900	2.500		
Urban	***N***	5,294	3,121	2,733	5,294	537	23					
	M/F		1.142		1.148	1.131	0.353					

Note: N represents the number of observations. M is for male, F female. The figure in the cells under M and F is the number of male children and female children in total, respectively. M/F is the number of males per female. The figure in each cell under "No. of Children" is the number of households having the corresponding number of children. The figures under "Birth Order" specify the number of children born in each birth order. Edu represents the level of education, where 8 is the highest education level, while 1 is illiterate. We reported the mean level of education, and its standard deviation is included in brackets.

Table A2
The effect of a Son Preference on the Quality of Children for Rural Populations [a]

	1988		1995		2002	
	Larger Sample	Multiple Births	Larger Sample	Multiple Births	Larger Sample	Multiple Births
Gender	– 0.501***	– 0.830***	– 0.151***	– 0.215*	– 0.005	– 0.191
	(.0267)	(.1008)	(.0212)	(.1297)	(.0261)	(.1505)
Quantity	– 0.047***	– 0.068*	– 0.039***	– 0.163**	– 0.110***	0.074
	(.0099)	(.0361)	(.0116)	(.0805)	(.0154)	(.1027)
Education (Father)	0.154***	0.151***	0.143***	0.005	0.235***	0.180**
	(.0082)	(.0304)	(.0095)	(.0681)	(.0121)	(.0855)
Minority	– 0.268***	– 0.194	0.268***	1.395***	0.194***	0.111
	(.0392)	(.2005)	(.0404)	(.3496)	(.0371)	(.2035)
Family income	0.00003***	0.00005***	0.00001***	0.00001	0.00002***	0.00001
	(.000005)	(.00001)	(.000002)	(.00001)	(.000001)	(.00001)
Terrain[50]	– 0.053***	– 0.299***	– 0.049***	– 0.169		
	(.0200)	(.0792)	(.0145)	(.1268)		
Geographic location[51]	0.060	– 0.152				
	(.0578)	(.2306)				
Old Revolutionary Areas	– 0.024	– 0.049	– 0.012	– 0.340		
	(.0418)	(.1302)	(.0260)	(.2176)		
Suburb	0.162	0.587*	0.281***	– 0.191		
	(.1075)	(.2976)	(.0515)	(.2617)		
Impoverished Areas	– 0.268***	– 0.127*	0.104***	0.507***		
	(.0392)	(.0684)	(.0214)	(.2081)		
Province	– 0.008***	– 0.007***	– 0.005***	0.003	0.000000001	– 0.0000007
	(0.0011)	(0.0042)	(.0008)	(.0067)	(.000000009)	(.0000006)
N	7,515	697	7,221	119	6,988	105

[a] Data from the Chinese Household Income Project Surveys. Quality of children is the dependent variable. Multiple-birth samples were abstracted from the original larger samples. Standard errors in parentheses. ***, **, * indicates statistical significance at 1%, 5% and 10%, respectively.

[50] Terrain includes the land features: 1. Plain; 2. Plateau; and 3. Mountainous region. Data was not available for the 1995 and 2002 samples.

[51] Geographic location includes: 1. Lakeside regions; 2. Coastal regions; 3. others. Data was not available for the 1995 and 2002 samples.

Table A3
The effect of a Son Preference on the Quality of Children for Urban Populations

	1988		1995		2002	
	Larger sample	Multiple Births Sample	Larger sample	Multiple Births Sample	Larger sample	Multiple Births Sample
Gender	– 0.015	– 0.158	0.069*	0.158	0.355***	
	(.0469)	(.1908)	(.0398)	(.3873)	(.0562)	
Quantity	– 0.210***	– 0.013	– 0.062**	– 1.156**	– 0.334***	
	(.0276)	(.1266)	(.0297)	(.5639)	(.0531)	
Education (Father)	0.127***	– 0.012	0.202***	0.067	0.236***	
	(.0147)	(.0689)	(.0128)	(.1043)	(.0214)	
Minority	– 0.180	0.430	0.041		– 0.078	
	(.0407)	(.3370)	(.0909)		(.1168)	
Family income	0.000003***	0.000006	0.00002***	– 0.00004	0.00002***	
	(.0000006)	(.0002)	(.000004)	(.00005)	(.000004)	
Residential location[52]	0.005	– 0.355	– 0.041	– 0.483	– 0.147***	
	(.0310)	(.1465)	(.0281)	(.3146)	(.0405)	
Province	– 0.028***	0.010	– 0.004***	0.005	– 0.001	
	(.0016)	(.0069)	(.0013)	(.0115)	(.0018)	
N	3,689	106	3,340	38[53]	2,213	Insufficient

[a] Data from the Chinese Household Income Project Surveys. Quality of children is the dependent variable. Multiple-birth samples were abstracted from the original larger samples. Standard errors in parentheses. ***, **, * indicates statistical

[52] Residential location is specified as: 1. City center; 2. City; 3. Suburbia; 4. Exurbia.

[53] For the multiple-birth sample, we intentionally dropped the minority variable to ensure a sufficient number of observations could be used for the regression analysis.