Accepted for publication in Review of Development Economics

By Magnus Hatlebakk, CMI, Norway

Son-preference, number of children, education and occupational choice in rural Nepal

Abstract

A unique family survey was conducted in Nepal to investigate the economic consequences of having a first-born girl. Women get more children, but we find no causal effect of number of children on economic outcomes. But independently of the number of children there is a positive effect on boys' education of having a first born sister, who presumably takes care of household work so the boys can focus on school. This indicates a stronger son-preference in Nepal than what is found in studies from neighboring countries.

JEL: D13, I21, J13, O12

Keywords: Fertility, intra-household, gender

1. Introduction

For South-Asia it is well known that son-preference explains selective abortion of girls (at least in India), or additional children if the first born turns out to be a girl[1]. We focus on the potential

[1] The most recent study we know of prenatal sex-selection in India is Hu and Schlosser (2015), and for the role of son-preference on fertility (and education) the most recent study we know is by Kugler and Kumar (2015). For a recent integrated analysis of both issues see Pörtner (2015).

costs of having too many children in Nepal, where, as we shall see, abortion of girls is still not a problem. We investigate whether son-preference leads to more children and thus less investment in each child. Inspired by the literature we started out with the random gender of the first-born as an exogenous instrument for number of children. But we discovered that it is not the number of children that matters for the investments in each child, it is rather the instrument itself that has a direct and positive effect, in contrast to the potential negative effect of number of children. If the first born is a girl, then later born boys get better education. This finding, which has no parallel in the literature as far as we know, is reported together with a number of robustness checks and analysis of additional economic outcomes.

We combine two strands of literature in a unified framework. Angrist and Evans (1998) is the seminal contribution to the literature where the gender of first born children is used as an instrument for the number of children. Their work led to an extensive literature focusing on the effect of family size on female labor supply, see for example Cruces and Galiani (2007) on data from Mexico and Argentina, and Daouli et al. (2009) on Greece. Within this strand there is a more limited literature that looks at other economic outcomes, including Gupta and Dubey (2006) on poverty in India. Our focus is on education, and similarly to us, most studies find no causal effect of number of children on children's education, although there are two recent studies that find a negative effect by Ponczek and Souza (2012) and Kugler and Kumar (2015). Ponczek and Souza also have a good review of this literature.

Within this strand of the literature the seminal theoretical model of the link between number of children and investments in each child is due to Becker (1960, see his footnote 10 for the formal model). Becker describes a quality-quantity trade-off in preferences where the two variables are simultaneously determined, subject to the costs of raising and educating children. If a women for some reason (and we focus on the random event that she gets a girl first) decides to get more children, she may at the same time decide to compensate by investing

less in each child. But that would be a causal effect of having a girl first on both quality and quantity of children. When we below attempt to identify a causal effect of number of children on their education, then we implicitly assume a recursive structure where the women first observe the gender of the first born, then decide on the number of children, and finally decides on their education. This is not an unlikely sequence of events, but as said, we do not find a significant effect in the last part of the chain of decisions, only the first effect of the gender of the first born on number of children is significant. This may reflect that a first born girl simultaneously affects both the number of children and their education. But we focus below on a more direct interpretation of the finding, as it is likely that the first born girl takes care of household work so that her brothers can focus on school. This interpretation is supported by Edmonds' (2006) analysis of time use among children in Nepal where he, in fact, finds that older girls work more than their brothers and even more if they have additional younger siblings.

The second strand of the literature applies birth-order and gender as direct explanatory variables. A robust finding from poor countries seems to be that siblings of first born girls get more education than others. We have, however, an even more specific finding in our data, as only male siblings get more education if the first born is a girl. We thus find a stronger son-preference, since only boys benefit, than in similar studies from neighboring countries where later born children of both gender benefit, see Sawada and Lokshin (2009) on Pakistan, and Ota and Moffatt (2007) on India. This male only effect is the main contribution of the paper. There is a related literature where the focus is only on birth-order and a possible interaction with own gender (but without looking at the gender of the siblings), see for example Ejrnæs and Pörtner (2004) who find that later born boys (in the Philippines) spend more time in school, and Emerson and Souza (2008) who find that daughters (in Brazil) who are first born are less likely to go to school than later born daughters.

We consider it value added to use data from Nepal, where the literature on son-preference is limited, but see Koolwal (2007) who analyzes the link between son-preference and child labor in Nepal. There is, however, a related literature on son-preference elsewhere, in particular Basu and de Jong (2010) on son-preferences in India, Das Gupta et al. (2003) on determinants of son-preference in China, India, and South-Korea, and Edlund and Lee (2009) on theory and evidence for son-preference in South-Korea.

The fact that Nepali women get more children if the first born is a girl is well known, see for example Gudbrandsen (2013), who uses the 2006 demographic and health survey (DHS)[2]. The DHS data has, however, only limited information on economic outcomes. An alternative is the Nepal Living Standards Surveys (NLSS, 1995, 2003, 2010), which are of high quality and contain information on a number of economic outcomes[3]. But with NLSS we have the problem that the unit of observation is a household and not a woman. There are ways to identify the children of each woman, and their birth-order, but there are inconsistencies in the data[4], and one can never be sure that all children are identified since the focus of the surveys is

[2] For more background on the fertility transition in Nepal and South-Asia see KC (2004), Karki and Krishna (2008) and Veron (2008).

[3] For more information and examples of economic analysis using NLSS data see Hatlebakk (2009 and 2011).

[4] In the 2010 NLSS questionnaire there is a maternity section (on page 30), where the person id in principle can link the children to other sections. Despite a column for level of education, it is only filled for children who were not in the rooster on page four. The rooster and the maternity file are inconsistent, some of the children (16%) who were reported in the maternity file as alive and living with the mother are not in the rooster, and some sons and daughters of the mother from the rooster (2%) are not in the maternity file. Furthermore, among the children who can

on household members and not on children who may have left the household. As a result there is no national level survey available that reliably can identify the link from birth-order of children to different economic outcomes in Nepal. As we did not have the resources to conduct a nation-wide family survey, we decided to do a survey in one area of Nepal where we expected to find variation in son-preference and fertility behavior due to the diverse caste and ethnic composition of the area. We interviewed women of age 40-59 as most of them have completed child-bearing and will have adult children for whom we can identify economic outcomes.

In the next section we discuss the suggested link from the gender of the first born, via number of children, to the education of those children and their occupational choice. The third section presents the data, then comes results and finally a discussion of the findings.

2. Theoretical foundation and hypotheses

In a recent report the UN Population Division (2015) argues: "Slower population growth would enable families and Governments to invest more in the health and education of each child, creating a virtuous circle with benefits for the economic, social and environmental dimensions of sustainable development." But why do poor people get many children, when there are potential benefits also at the family level from investing in fewer children? Based on our interviews with women in rural Nepal it appears that son-preference is one of the drivers. Many say they get additional children if the first born is a girl. The too-many-children hypothesis is, however, an imprecise statement, and we will investigate more specific hypotheses utilizing the fact that women get more children if the first born is a girl. More precisely, how will the gender of the first born affect the education and in the last instance the occupational choice of later

---

be matched on their individual id, we find that 3% have changed gender from page four to page 30 in the questionnaire.

born children? The effect is likely to depend on the gender of the later born themselves, so we separate the effects on girls and boys, and as fertility and investments in education may differ between social groups, we will control for social identity. Son-preference is expected to be stronger within particular groups depending first of all on religious beliefs. It is for example considered important for some social groups to have a son who can lit the funeral pyre.

In the analysis we will utilize the natural experiment that the gender of a child is random in Nepal, as indicated by the sex-ratio at birth[5]. Selective abortion of girls was even less likely two decades ago, when the children in our sample were born (they are all born a decade or more before the legalization of abortion in 2002). Thus, women get more children when the first born is a girl, and more children may, in turn, affect other economic decisions. But having a girl first, instead of a boy, may also have additional effects. In particular we shall expect a girl to take over some of the household chores that the mother would otherwise do. This, in turn, means that the mother can work more outside the household bringing home incomes that can be

[5] The most recent DHS (2011) reports (in Table C4) a sex-ratio at birth of 106. It varies (apparently stochastically) according to the year born, so children born during the last five years had a sex-ratio of 105, while the preceding five years had a ratio of 107. These ratios are within the normal range, in UK and US the sex-ratio is 105. This corresponds with the final conclusion of a careful study by Valente (2014) on the causal effect of access to abortion clinics in Nepal. Valente does, however, warn that sex-selective abortion may become a problem in the future as in India, in line with other papers that focus on recent trends in urban Nepal, including Adhikari, Ghimire and Ansari (2008), Frost, Puri and Hinde (2013), and Puri and Tamang (2015). The 2011 census, see Table 3.1 in CBS (2014), also indicates a regional variation in the sex-ratio at birth, with high rates in the Kathmandu valley, and very low rates in the western mountains (which may be a reporting problem).

invested in, for example, education for the younger children. How far can theory help us in understanding these decisions? Although the discussion below is based on economic theory we do not provide a formal model, we instead describe a possible strategy that may be 'optimal' in some sense for a woman living in an economy where girls have fewer opportunities in the labor market than boys.

We assume that the woman (or the family) makes a plan for number of children, investment in their education, as well as her own occupational choice, but a plan that is contingent on the gender of the children. A realistic strategy may be the following, which is based on conversations we have had with rural women in Nepal:

- Girls take care of household chores, which allows the boys to focus on education, and the mother to work outside.
- The woman gets children until she has two boys.
- Both boys and girls are allowed to complete primary school, but boys get more time for school work.
- Children with good school results are allowed further education.

The two-boys target can be an optimal strategy if there are costs of having children, but at the same time there is a higher chance of at least one talented boy if you have more than one. If the women apply this strategy, then we shall expect to observe the following:

1) Women get more children if the first born is a girl (since they want two boys).
2) More boys than girls take education beyond primary level.
3) More boys take education beyond primary level if the first born is a girl.
4) As a result these boys may get a better job than other boys due to a first born sister.

5) The mother works outside the household if the first born is a girl.

Now, one may imagine that even if there is no first-born girl, but for example a second-born girl, she may also take care of household chores, which would have some of the same implications. But if we believe that only girls, or women, take care of household chores, then a family with a first born boy will be forced to find another solution for household work during the first years when there is only a boy around. So we shall not expect a later born girl to have the same strong implications. And since there is already a boy, there will probably be no effect on the number of children of a later born girl. So for these reasons, we start out with the gender of the first born as the natural experiment of interest when we now investigate the hypotheses listed above. As we shall see, we do, however, end up with an empirical analysis where we also utilize the gender of the second born in the empirical strategy.

3. Data and descriptive statistics

The survey was conducted in a rural area north-east of the city of Biratnagar in the plains (terai) of Nepal. The area has a particularly diverse caste and ethnic composition. There is an indigenous (Janajati) population (Tharu and Rajbansi in particular), then there are people of hill origin who over generations have migrated to the plains, in our sample mostly high caste groups, and finally there are people of Indian origin who also have migrated to the area over generations, mostly Dalits in our sample (Bantar and Musahar in particular). As there may be variation in son-preference between different Janajati or Dalit groups, we do not lump them together but rather construct dummy variables for the three largest groups in the sample, and

have all other groups as the reference category[6]. These three groups are Tharu (36% of the sampled women), Bantar (13%) and hill origin Brahmin and Chettri (10%). Compared to national level data the Tharu (about 7% of Nepal's population) and the terai Dalits (about 5% of Nepal's population) are clearly overrepresented, while the hill Brahman and Chettri are underrepresented as compared to the national level of about 30%. We expect our findings to be relevant for other terai villages, in particular in the eastern, so called, Madhes belt, but less relevant for the hills.

The area is relatively peaceful (which was essential given the recent civil war and post-war ethnic unrest in Nepal), and all villages are within 0.5-1.5 hours by bus or bicycle from Biratnagar. Every second household has an extended family member working in Biratnagar (with construction labor being the most common occupation among them). It was also of importance that we have been working in this area for the last 15 years and thus know the economic and social context well. For more information on the eastern terai see Hatlebakk (2007).

We randomly selected five out of the 12 villages (VDCs) in the area with probability given by the population size, and we selected four wards at random in each VDC, so a total of 20 PSUs[7]. Then we selected randomly 24 women of age 40-59, so a total sample of 480. There were three enumerators (two female), and they normally interviewed three women per day, so a PSU was completed in three days and they thus spent two months in the field from mid-November 2011 to mid-January 2012. The enumerators were from the Kathmandu based

[6] The reference category consists of 35 distinct groups that all are separate categories in the national census, with the largest in our sample being Rajbansi (6%), Musahar (5%), Sanyasi (3%), and 2% each of Rai, Jangad, Kewat, Gangai and Mallah.

[7] In all analysis below standard errors are clustered at the PSU level.

research institute New Era and had participated in the recent DHS some months earlier. We intentionally based parts of the survey on the DHS questionnaire to make sure that the questionnaire was familiar to the enumerators. But we added some questions on economic outcomes that were less familiar to the team.

In the sample there were 16 women (3%) with no children and 20 (4%) with a single child. Since we study the implications of the gender of a first born child on the education of later born children these 20 children (and 36 women) are dropped from the analysis. Table 1 still shows the number of all ever born children according to the gender of the first born. There are in total 1813 children (the sum over the products of the first and last columns) born to these 464 women.

Table 1 about here

We note that there are 105 boys to every 100 girls (238 to 226) among the first-born, or 51.3% boys, which is normal and in correspondence with available national level data. The sample is small, and thus the 95% confidence interval for the sex-ratio goes from 46.3% to 56.2%. So although the sex-ratio by coincidence is as expected, we shall not be surprised when the sex-ratio for later born children is 53%. We shall see below that we end up with a relatively small sub-sample of females with a first-born brother. A possible explanation would be that a first-born brother gets priority in terms of food and health care and his sisters thus die early, but in our sample there are in fact fewer girls already at the time of birth, so this is most likely a statistical coincidence.

There is, however, a significant difference (at the 99%-level) in the number of children depending on the gender of the first born. We find that if the first born is a girl then the women get on average 4.1 children, while they get only 3.7 children if the first born is a boy, so by that

we have confirmed that women in this area behave as women all over Nepal, and the first hypothesis from the previous section is confirmed. The median is three children if the first born is a boy, and four if the first born is a girl[8]. So we may conclude that if the median woman gets a girl first the child does not count, and the woman gets three more children, the same number she would get in total if the first born was a boy. The median age of the women was 48 years at the time of the survey, while the median for their children was 23 years. So the children were born around 1990. Today women in this area get fewer children.

Some of these children die early, and may be replaced before the woman gets to her ideal number of children. Let us say a boy dies at age five, and the son preference is for adult sons, then she will probably get an additional child and even another one if the next one is a girl. Table 2 thus reports the number of live children at the time of the survey. There are 3.71 live children if the first born is a girl, and 3.17 if the first born is a boy, a significant difference of 0.5 children, while the median is again respectively 4 and 3 children.

Note that the following stopping rule will give 3.75 children if the first born is a girl, and 3.25 children if the first born is a boy (when we simplify by assuming a 50% chance of having a boy): Stop after 3 if you have at least 2 sons, otherwise stop after 4. This seems to be a realistic rule that turns out to give the average number of children that we observe. It will, however, give only 3 and 4 children, while the real distribution is wider. As we shall see later the R-squared in the regressions, where we investigate whether the gender of the first born matters for economic outcomes, also indicates that there is extensive individual unexplained variation. So although there may be some underlying stopping rule of this kind, there will in reality be more variation. Women with only one or two children may have wanted more, but

[8] Among the women with at least one child, 59.7% get four or more children if the first-born is a girl, while the share is 49.6 % (significantly lower at the 99%-level) if the first born is a boy.

some children died, while women with six or more children may have wanted fewer, but were not able to control the number of pregnancies.

Table 2 about here

There are 1587 live children in Table 2, where 386 of the 462 first born are still alive. In the analysis below we focus on outcomes for those children who are not first-born, which gives a sample size of 1201. We lack information on education for some of them, so for education we have a sample of 1187 children. These children vary in age from 5 to 48, with the median being 22 years. In the analysis we omit the youngest children as they have not had a chance to complete their education. From age 15 there is basically no variation in the mean years of schooling, the mean is seven years of schooling up to age 27, while older people have less education. For SLC the critical cutoff seems to be at 19 years. Below 19 there are many more children still in school. We thus restrict the population of interest to adult children of at least 19 years old, which gives a final sample of 787 children[9]. More detailed statistics for this sample will be presented in the next section, but Table 3 gives an overview and also presents the variables that will be used in the regression analysis in tables 9 and 10. Standard deviations are here for the variables themselves, and not the mean, and are thus not clustered.

[9] The sample still includes 98 (12%) people who report to be a student as their primary activity. These are included in the sample to avoid the selection problem that would arise if we excluded them based on their decision to study. The downside is that we underestimate their years of schooling, but we consider this to be a minor problem at the upper end of the distribution. 81% of them have SLC and are now studying at a higher level, so they are recorded with, let us say, 12 years of schooling, while they may end up with 14 years.

Table 3 about here

The first variable shows that this sample of adults has on average seven years of schooling[10], which is high by Nepali standard and reflects that the villages are close to the city of Biratnagar[11]. The second variable, which is a dummy for completed 10-years of schooling, shows that 31% have managed to pass the exam for the school-leaving-certificate (SLC). This is also above the national average. In the sample we have 53% men. Then we have 53% children with a first-born sister, this is also a dummy variable. The interaction variable on the next line is the interaction between the male dummy and the female-first-born dummy. It takes the value one if a boy has a first born sister, otherwise it is zero. Since there are slightly more girls born than boys, the two-female-first-born dummy has a mean of 28.7% of the children, slightly larger than the 25% one would get with equal chance of boys and girls. The women have on average 4.3 live adult children. The final control variable in the analysis below will be social identity. The most numerous social groups in our sample are, as discussed above, the indigenous group of Tharus, the terai Dalit group of Bantars and the hill origin Brahmin and Chettri high caste groups. The reference group is the other ethnic/caste groups combined.

---

[10] We have a coding problem as one possible response to the question on education was 'literate only', meaning that they have not completed class one, but are not illiterate either. We have coded the illiterate with zero years of education, and literate only with one year of education. In an earlier version of this paper it was coded as zero.

[11] Note that NLSS tend to report years of schooling in this range at the national level, but then only for those who have actually attended school, while we also includes the zeros.

4. Findings

We have already found support for the first hypothesis, as discussed in relation to Table 1, that women get significantly more children if the first born is a girl. We now go on to the other hypotheses. Table 4 shows the level of education for six groups, female and male adult children with either a first-born sister or brother in the four first columns (with the four subsamples adding up to the sample in Table 3), while the last two columns show the first-born themselves. We have already commented on the fact that the sub-sample in the fourth column is, apparently by coincidence, smaller.

Table 4 about here

Each column shows the distribution of a sub-sample according to years of schooling with summary statistics at the bottom, where we show the median and the mean, followed by the percentage of illiterates and the percentage with completed SLC after 10 years of schooling. For (later born) men there is a significant difference in the level of education (along all three dimensions) depending on the gender of the first born sibling. These differences are found by comparing the summary statistics in the first and second columns. In particular we find that if a boy has a first-born sister he will have on average 8.3 years of schooling, while if he has a first-born brother he will have 7.2 years of schooling. He will also have a much higher chance of completing the SLC exam (41.3% as compared to 26.6%), and a lower chance of being illiterate (5.3% as compared to 10.1%). There is no such difference for girls, as we can see from the next two columns where both groups have 6.2 years of schooling. So later born girls have the same chances independently of the gender of their first born sibling, while boys are much better off if their first born sibling is a girl. We thus have support for the third hypothesis. There is also support for the second hypothesis, boys get significantly more education in general,

except for SLC where only boys with a first born sister are doing better than the other groups. We also see, from the summary statistics in the two last columns, that a first born girl gets less education than a first born boy.

We thus conclude that if the first born is a girl, it appears that she will more often than a first born boy have to work at home, instead of going to school, and it is her brothers who benefit from this. Furthermore, the mother will get more children than otherwise in an attempt to get one or more boys, who will ultimately benefit from their sister's work.

Economic outcomes

We now go on to the two last hypotheses, where we attempt to identify economic consequences later in life of having a first-born sister. Since all these children are adults we have information on their occupation, as well as the occupation of the mother, see Table 5 for the children and Table 7 for the mothers.

Table 5 about here

We see that even though the gender of the first born has consequences for the level of education for boys, it appears to have no consequence for the type of work they end up with. For basically all occupations we find about the same number of people independently of the gender of their first born sibling, but of course with differences according to own gender. The exception is again education, where there seems to be more students (the second last category) among men with a first-born sister. In contrast, there seems to be more 'other labor' among men with a first born brother, which indicates that boys that do not have a first born sister more often end up as laborers.

Although they, for the most part, select the same occupation independently of the gender of the first born, there may be a difference in pay due to the difference in education. Approximately 50% of the people have occupations where they work for a daily wage or a monthly salary, we report these wages and salaries in Table 6.

Table 6 about here

There is only one significant finding when it comes to regular incomes for the non-first born adult children, women earn a smaller salary if their first-born sibling is a girl (5761 rupees on average as compared to 10230 if she has a first born brother). So the difference we found for education among men has no implications for salaries or wages, while there is a difference for women instead. This difference is for a very small sub-sample, there are only 13 women with a first-born sister, and 10 with a first born brother, so although significant, the sample (and thus probably the corresponding population) is small. The lower paid women work basically as teachers in private schools, while the higher paid women work in government schools, and for some reason the women with a first-born sister end up in private schools, while women with a first-born brother end up in government school, maybe because those brothers have better networks.

The main finding here is, however, that for the larger sample of men there is no difference in pay, in contrast to the findings for education. This may reflect that education has limited value in Nepal beyond the intrinsic value and the value of not having to do household or farm work while you go to school. But these men are still relatively young, and some of them are still students, so there may be an expectation that they will get higher incomes in the future.

With respect to the mother's occupation and earnings, the hypothesis was that a first-born girl will take care of household chores so that the mother can work outside the household.

We do not have information on her occupation when the children were small, but if she did work at that time, she may have a better chance of having a good job today, and this is what we investigate in Table 7. There are so few women with a salaried job that we do not report on salaries. Similar to the children, the distribution over occupations is almost identical for the two sub-samples, so apparently no difference according to the gender of the first born. In conclusion, it appears that there is no long-term effect of having a first-born daughter who helps out with household-work.

Table 7 about here

Number of children and education

The main finding above is that boys get more education if they have a first-born sister. And we also confirmed that women get more children if the first born is a girl. We now go on to investigate whether the effect of a first born girl on her brothers' education is a direct effect (as she takes care of household chores), or whether the effect goes via the number of children. If there are more children available, then the family will probably select the most talented one, maybe only among the boys, for further education. If so, then we may not see any difference between the children (at least not between boys) in primary school, but later they may prioritize one child. If this is the case, then there are two conflicting mechanisms. There will be more children with only primary school in large families, which will pull down the average number of years in school for large families. This effect will probably dominate, but there is a counteracting effect as the average talent of the best student will increase with the number of children, and thus the investments put into that child in terms of years of schooling may increase with the number of children. In addition, if there are more children around they may help each

other out at home and in school, and thus do better in school, which in turn may lead to more investment in education. So it is not clear what will be the total effect of number of children on the average level of education of those children. Furthermore, we would like to separate the number of children effect from the gender of the first born effect.

We start with descriptive statistics, before we go on to discuss causality. In the descriptive statistics we separate the total effect shown in Table 4 into the direct effect and the number of children effect by splitting the table according to number of children. Table 8 reports on the 413 men from the two first columns of Table 4 as this is where we have a difference that appears to be determined by the gender of the first born. Very few of these men are from families with only one live child, or more than six children, so we categorize number of children into the five categories shown. We then show the average level of schooling (using all three indicators from Table 4) for the ten categories of men determined by the five family sizes and the gender of the first born.

Table 8 about here

The higher level of education for men with a first born sister is consistently significant over all three indicators for the three-children category, while the higher proportion of SLC is significant also for the four-children category and the higher years of schooling is significant for the five-children category. These sub-group findings add up to the aggregate findings from Table 4 where men benefit from having a first-born sister. But when it comes to the number of children, and we control for the gender of the first born by comparing the outcomes within each column, then there is not much variation.

There are two exceptions that we will argue are special cases. First we note that in the two children case, where both children are male, the second born has better education than other

groups (48% have SLC for example). This is probably a case where the mother or some other women takes care of the household work. The second exception is the variation in literacy rates according to the number of children. There seems to be a U-shaped function, at least if the first-born is a boy. But if we restrict ourselves to only significant findings, then this conclusion is based on a single number, the low illiteracy rate for men with a first-born brother in four children families. So the finding may be a statistical coincidence, but we do see a U-shaped function also for the first-born sister group, although not significant.

Causality

Our initial strategy for identification was to follow the standard approach in the literature and use the (random) gender of the first born as an instrument for number of children to identify the potential costs of having many children. But when the descriptive statistics, which would also be the reduced form model, showed that the gender of the first born matters only for the male siblings of the first born, then we were no longer convinced that the number of children was the main factor. In the reduced form the only significant finding is that if the first born is a girl then later born boys get a better education. The reduced form thus indicates the opposite sign of what we expected. If the first-born is a girl, we know there are more children around, but it turns out that this leads to more education, although only for boys. So it seems we have to reject the number of children hypothesis and look for an alternative explanation, which in fact jumps out of the data. A better education for boys when the first-born sibling is a girl is probably due to the direct effect of that girl taking care of household chores so that her younger brothers can focus on school. So this is what we report as the main finding.

But let us now revisit the original hypothesis. If we control for the main effect, can we still find an additional effect of the number of children? In identifying a useful instrument for the number of children, we can no longer use the gender of the first born. However, while

controlling for the gender of the first born, we have found that a dummy for two first-born girls appears to be the best instrument for number of children[12]. The instrument will measure the additional effect on number of children of having a second born girl subject to having a first born girl[13]. If there is a son-preference we expect the dummy to have a strong effect on number of children, which is already documented by Gudbrandsen (2013).

The dummy is problematic to use as an instrument if a second born girl also contributes to the education of later born boys. To investigate this we checked the educational level of third born boys to see whether it varies according to the gender composition of the two first born. We find that third born boys have 8.1 years of schooling if there are two first-born girls, which is not significantly different from the 9.2 years of schooling if only the first-born is a girl (and the second is a boy). This indicates that it is only the first born girl who matters. The finding is further strengthened when we shift the order, if the first born is a boy (and the second is a girl)

---

[12] We have systematically investigated different combinations of the gender of the first born, but this is the best predictor for number of children. This systematic testing of different combinations of gender is similar to the first stage regressions in Angrist, Lavy and Sclosser (2010). As in that paper we may also control for ethnic dummies also in the first stage but that does not essentially change our results. Gupta and Dubey also use a dummy for two first born girls as the instrument, and similarly to us, and Angrist, Lavy and Sclosser (2010), they find no effect of number of children on the economic outcome (poverty in their model). But as far as we know we are the first to simultaneously estimate the direct effect of the gender of the first born.

[13] The model can be set up as N = f(gender of the first born, gender of the second born) and E = h(gender of the first born, N), where the gender of the second born is the instrument that affects the number of children N, but not directly the education E of later born children.

then the third boys have 6.2 years of schooling, which is significantly lower than 9.2 years. Similarly if there are two boys, the third boy will have 5.8 years of schooling. So we conclude that it is only a first-born girl that matters for her brothers' education. This confirms the descriptive findings in Table 4, but in addition it supports the use of the dummy for two girls as an instrument for number of children as it appears that there is no direct effect of that second girl on the education of later born boys.

Before we go on to the results, note that we can include two-children families in the analysis, but we have to be careful in interpreting the reduced form model: The dummy for two first-born girls in a two children family automatically means that the sampled child is a girl, and thus on average has less education. So when we find no effect of this dummy in the reduced form, this can, in principle, be part of the explanation. In reality, however, this is not a problem, we get the same insignificant parameter if we restrict the analysis to families with three children or more. Similarly the IV-estimates do not change essentially if we restrict the sample to families with more than two children.

The IV regression (with the two-children families included) as well as OLS regressions with different controls, including the reduced form model, are reported in tables 9 and 10.

Table 9 about here

The purpose of our research was to investigate whether families with many children invest less in the children's education, which may be expected given a potential quantity-quality trade-off as indicated by the OLS-regression in the first column of tables 9 and 10, where we find that children have less education the more children (higher N-alive) their mother has.

However, the descriptive statistics (and the reduced form model in the third columns of tables 9 and 10) indicate that the gender of the first born only matters for male siblings of the

first born. Boys with a first born sister, and only this group, get more education than any other group. And we have a positive effect, not a negative as we should expect if a first born girl led to more children and thus a higher cost of educating each of them. So we need an alternative explanation, and the immediate interpretation is the one discussed above, where the boys benefit directly as the first born girl takes care of household work, while the boys focus on school. This finding, that boys benefit from a first-born sister, is significant in the IV regression in the second column, as well as in the reduced form model in the third column.

The dummy for two first born girls instrument is highly significant in the first stage of the IV. Then in the second stage we find that number of children does not matter for the education of each child. Subject to this finding we also report a regression without the apparently redundant (and endogenous) number of children variable in the fourth column, and the results are still the same, and we add social identity as controls in the fifth column, without changing the results. When we control for caste, however, the R-squared increases, which indicates that preferences for investment in children vary between social groups. As expected, the hill origin high castes are better educated than others, while the large Dalit group of Bantar have less education. The indigenous Tharu group have more years of schooling than the reference group.

We also note that boys in general get more years of schooling, while for SLC there is no significant direct gender effect. This may be because a girl with a good education may enter the high status marriage market, and also will have some probability of getting a well paid job.

Table 10 about here

The main finding of this paper is that boys get a better education if they have a first born sister. For the US, Butcher and Case (1994) find the opposite result, women raised with brothers

get a better education, while men's education does not depend on the gender of their siblings. For South-Asia, Sawada and Lokshin (2009) find a similar effect to ours in Pakistan, but for both gender, so in Pakistan even girls benefit from having older sisters. Ota and Moffatt (2007) replicate this finding on data from India[14]. In conclusion, our findings indicate an even stronger son-preference in Nepal than in neighboring India and Pakistan, as only boys appear to benefit from having an older sister.

5. Conclusions

We find that the number of children does not matter for how much education each child gets. In looking for a variable that could establish an exogenous variation in the number of children, we did, however, find new evidence for son-preference in Nepal. We knew from before that if the first born is a girl, then women get more children. When we control for the number of children, we find that a first born girl has a major direct effect on education of later born children, but only for boys. Thus boys benefit from having a first born sister, independently of how many siblings they have, with the most likely explanation being that the sister takes care of household work while the boys can focus on school. This is a strong evidence of son-preference, and a kind that we have not seen in earlier studies from South-Asia.

References

Adhikari, N, Ghimire, A. and Ansari, I. (2008). Sex preference in urban Nepal. *Journal of Institute of Medicine. 30(2)*, 19-23.

[14] Although their descriptive statistics in Table 2 may indicate that boys benefit more, but this effect may not be significant?

Angrist, J.D. and Evans, W.N. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*. *88*(3), 450-477.

Angrist, J., Lavy, V. and Schlosser, A. (2010). Multiple Experiments for the Causal Link between the Quantity and Quality of Children. *Journal of Labor Economics*. *28*(4), 773-824.

Basu, D. and de Jong, R. (2010). Son Targeting Fertility Behavior: Some Consequences and Determinants. *Demography*. *47*(2), 521-536.

Becker, G. S. (1960). An Economic Analysis of Fertility. In Becker, G. S. (Ed.), *Demographic and Economic Change in Developed Countries* (pp. 209-231). Princeton University Press.

Benporath, Y. and Welch, F. (1976). Do Sex Preferences Really Matter? *Quarterly Journal of Economics*. *90*(2), 285-307.

Butcher, K.F. and Case, A. (1994). The Effect of Sibling Sex Composition on Women's Education and Earnings. *Quarterly Journal of Economics*. *109*(3), 531-563.

CBS (2014). *Population Atlas of Nepal*. Central Bureau of Statistics. Kathmandu, Nepal.

Cruces, G. and Galiani, S. (2007). Fertility and Female Labor Supply in Latin America: New Causal Evidence. *Labour Economics*. *14*(3), 565-573.

Das Gupta, M., Zhenghua, J, Bohua, L., Zhenming, X., Chung, W. and Hwa-Ok, B. (2003). Why is Son Preference so Persistent in East and South Asia? A Cross-Country Study of China, India and the Republic of Korea. *Journal of Development Studies*. *40*(2), 153-187.

Daouli, J., Demoussis, M. and Giannakopoulos, N. (2009). Sibling-Sex Composition and its effects on Fertility and Labor Supply of Greek Mothers. *Economics Letters*. *102*(3), 189-191.

DHS (2011). *Nepal Demographic and Health Survey*. Available at: http://dhsprogram.com/.

Edlund, L. and Lee, C. (2009). *Son Preference, Sex Selection and Economic Development: Theory and Evidence from South Korea*. Columbia University. Department of Economics. Discussion Paper no.: 0910-04

Edmonds, E.V. (2006). Understanding Sibling Differences in Child Labor. *Journal of Population Economics*. *19*(4), 795-821.

Ejrnaes, M. and Portner, C.C. (2004). Birth order and the intrahousehold allocation of time and education. *Review of Economics and Statistics*. *86*(4), 1008-1019.

Emerson, P.M. and Souza, A.P. (2008). Birth order, child labor, and school attendance in Brazil. *World Development*. *36*(9), 1647-1664.

Frost, M.D., Puri, M. and Hinde, P.R.A. (2013). Falling sex ratios and emerging evidence of sex-selective abortion in Nepal: evidence from nationally representative survey data. *BMJ Open*. 3:e002612.

Gudbrandsen, N.H. (2013). Female autonomy and Fertility in Nepal. *South Asia Economic Journal*. *14*(1), 157-172.

Gupta, N.D. and Dubey, A. (2006). Fertility and the Household's Economic Status: A Natural Experiment using India Micro Data. *Journal of Development Studies*. *42*(1), 110-138.

Hatlebakk, M. (2007). *Economic and social structures that may explain the recent conflicts in the Terai of Nepal*. CMI. Report for the Norwegian Embassy in Kathmandu.

Hatlebakk, M. (2009). Capacity-constrained Collusive Price Discrimination in the Informal Rural Credit Markets of Nepal. *Review of Development Economics*. *13*(1), 70-86.

Hatlebakk, M. (2011). Triadic power relations in Rural Nepal. *Journal of Development Studies*. *47*(11), 1739-1756.

Hu, L. and Schlosser, A. (2015). Prenatal Sex Selection and Girls' Well-Being: Evidence from India. *Economic Journal. 125*(587), 1227-1261.

Karki, Y.B. and Krishna, R. (2008). *Factors Responsible for the Rapid Decline of Fertility in Nepal – An Interpretation: Further Analysis of the 2006 Nepal Demographic and Health survey*. Calverton, Maryland: Macro International Inc.

K.C., Bal Kumar (2004). Migration, poverty and development in Nepal. *Asian and Pacific Migration Journal*. *13*(2), 205-232.

Koolwal, G.B. (2007). Son Preference and Child Labor in Nepal: The Household Impact of Sending Girls to Work. *World Development*. *35*(5), 881-903.

Kugler, A. and Kumar, S. (2015). *Preference for Boys, Family Size and Educational Attainment in India*. April-2015. NBER-WP-21138.

NLSS (1995). *Nepal Living Standards Survey Report 1996. Main findings*. Central Bureau of Statistics. National Planning Commission Secretariat.

NLSS (2003). *Nepal Living Standards Survey 2003/2004. Statistical report*. Central Bureau of Statistics. National Planning Commission Secretariat.

NLSS (2010). *Nepal Living Standards Survey 2010/2011. Statistical report*. Central Bureau of Statistics. National Planning Commission Secretariat.

Ota, M. and Moffatt, P. (2007). The within-household schooling decision: a study of children in rural Andhra Pradesh. *Journal of Population Economics*. *20*(1), 223-239.

Ponczek, V. and Souza, A.P. (2012). New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country. *Journal of Human Resources*. *47*(1), 64-106.

Puri, M. and Tamang, A. (2015). *Understanding Factors Influencing Adverse Sex Ratios at Birth and Exploring What Works to Achieve Balance: The Situation in Selected Districts of Nepal*. CREHPA. Lalitpur, Nepal.

Pörtner, C. (2015). *Sex-Selective Abortions, Fertility, and Birth Spacing*. February-2015. World Bank Policy Research WP No. 7189.

Sawada, Y. and Lokshin, M. (2009). Obstacles to school progression in rural Pakistan: An analysis of gender and sibling rivalry using field survey data. *Journal of Development Economics*. *88*(2), 335-347.

Valente, C. (2014). Access to abortion, investments in neonatal health, and sex-selection: Evidence from Nepal. *Journal of Development Economics*. 107(March): 225-243.

Veron, J. (2008). The demography of South Asia from the 1950s to the 2000s. A summary of changes and a statistical assessment. *Population*. *63*(1), 9-92.

Table 1. Number of ever born children

Number of children	Male-first-born	Female-first-born	All
1	12	8	20
2	44	28	72
3	64	55	119
4	58	56	114
5	23	31	54
6	25	30	55
7	7	5	12
8	3	9	12
9	2	2	4
10	0	2	2
N	238	226	464

Table 2. Number of live children

Number of children	Male-first-born	Female-first-born	All
1	19	13	32
2	54	32	86
3	78	67	145
4	48	49	97
5	23	39	62
6	12	16	28
7	1	4	5
8	1	4	5
9	0	1	1
10	0	1	1
N	236	226	462

Table 3. Descriptive statistics for people of 19+ years (N=787)

Variables	short-name	mean	std.dev.	min	max
Completed years of education	edunfb	7.036	4.486	0	18
Completed 10-years of education	slcnfb	0.311	0.463	0	1
Male	male	0.525	0.500	0	1
Female first-born sibling	female-first-born	0.525	0.500	0	1
Interaction: male* female first-born sibling	inter-male-femfb	0.262	0.440	0	1
Two female first-born (instrument for N-alive)	female-two-first	0.287	0.453		
Number of alive children of age 19+	N-alive	4.342	1.550	1	10
Indigenous group	tharu	0.344	0.475	0	1
Dalit group	bantar	0.121	0.326	0	1
High-caste group	BC	0.105	0.307	0	1

Table 4. Years of completed education among people of 19+ years

	Male		Female		Male	Female
Years of education	First-born sister	First-born brother	First-born sister	First-born brother	First-born	First-born
Illiterate	11	21	41	26	11	37
1	8	13	22	17	9	24
2	3	9	5	3	6	8
3	7	9	5	5	3	0
4	12	8	9	14	8	7
5	15	8	14	9	19	13
6	12	15	2	7	7	5
7	9	12	13	10	11	5
8	17	16	11	10	14	9
9	14	21	9	12	13	2
10	13	20	13	12	15	18
SLC	46	34	42	33	28	29
Studying	11	2	7	2	3	7
Intermediate	20	12	13	5	11	8
Bachelor	5	5	1	2	5	1
MA+	3	2	0	0	1	0
Median years	9	8	7	7	8	5
Mean years	8.3***	7.2	6.2	6.2	7.6***	5.7
95%-conf-int.	(7.3-9.4)	(6.0-8.4)	(4.9-7.6)	(5.1-7.4)	(6.7-8.5)	(4.6-6.9)
Illiterate-%	5.3*	10.1	19.8	15.6	6.7***	21.4
95%-conf-int.	(0.0-10.7)	(3.9-16.4)	(10.2-29.4)	(6.1-25.1)	(1.0-12.5)	(12.7-30.0)
SLC-%	41.3***	26.6	30.4	25.1	29.3	26.0
95%-conf-int.	(31.4-51.1)	(17.3-35.9)	(18.9-42.0)	(16.7-33.6)	(19.8-38.8)	(16.7-35.3)
N	206	207	207	167	164	173

Literate but no schooling is coded as completed one year of education
Confidence intervals corrected for intra-PSU correlations
*** Significant different from next column at 1%-level
** Significant different from next column at 5%-level
* Significant different from next column at 10%-level

Table 5. Occupation of people of 19+ years

	Male		Female		Male	Female
Occupation	First-born sister	First-born brother	First-born sister	First-born brother	First-born	First-born
Farmer	26	27	57	45	25	50
Business	17	14	17	11	14	19
Rickshaw	4	3	0	0	4	0
Farm labor	7	6	43	37	4	37
Factory labor	43	45	7	2	40	2
Constr labor	27	26	1	4	29	1
Other labor	5	20	1	0	10	1
Empl rest	7	7	0	0	6	0
Empl shop	6	3	1	0	5	0
Empl priv off	5	2	1	1	4	0
Empl priv oth	17	21	3	1	6	5
Empl gov	4	8	2	5	3	3
Other work	2	2	4	2	2	1
Housework	0	1	40	41	0	40
Student	35	19	27	17	10	13
No work	1	3	3	1	2	1
%-farmer	13%	13%	28%	27%	15%	29%
%-farm labor	3%	3%	21%	22%	2%	21%
%-fact labor	21%	22%	3%	1%	24%	1%
N	206	207	207	167	164	173

Table 6. Wage and salary among people of 19+ years

	Male		Female		Male	Female
Wage	First-born sister	First-born brother	First-born sister	First-born brother	First-born	First-born
Median wage	300	300	150	150	300	150
Mean wage 95%-conf-int.	297 (258-336)	324 (285-362)	147 (130-165)	148 (130-166)	338*** (314-362)	148 (132-165)
N-wage	41	57	48	40	43	39
Salary						
Median salary	10000	10000	5000	11500	12000	7000
Mean salary 95%-conf-int.	12281 (9885-14678)	11381 (9459-13304)	5761** (3102-8420)	10230 (7034-13426)	14432** (9647-19217)	7580 (4303-10875)
N-salary	77	80	13	10	60	10

Wage is paid per day, salary per month

*** Significant different from next column at 1%-level

** Significant different from next column at 5%-level

Table 7. Occupation and wages for mothers

Occupation	First-born daughter	First-born son
Farmer	115	117
Business	20	24
Farm labor	41	45
Factory labor	5	8
Constr labor	3	2
Other labor	1	1
Empl gov	1	2
Other work	2	2
Housework	33	31
Pensioner	0	1
No work	5	3
%-farmer	51%	50%
%-farm labor	18%	19%
%-fact labor	2%	3%
N	226	236
Median wage	100	128
Mean wage	136	135
N-wage	50	54

Table 8. Education among men of 19+ years

	Years of education		SLC		Illiteracy	
Number of children	First-born sister	First-born brother	First-born sister	First-born brother	First-born sister	First-born brother
1-2 (N=39)	7.1	**9.1**	35.7	**48.0**	14.3	12.0
3 (N=115)	***9.1***	6.9*	*50.0*	25.4*	*0.0*	7.5
4 (N=105)	8.5	7.6	*43.4*	23.1*	1.9	**1.9**
5 (N=84)	*8.2*	6.7	37.3	24.2*	3.9	9.1
6-10 (N=70)	7.8	6.0**	35.0	20.0**	15.0	30.0*

** Significantly different from the bold reference category in same column at 5%-level
* Significantly different from the bold reference category in same column at 10%-level
Italic means significant horizontal difference for same variable at 5%-level

Table 9. Determinants of years of schooling among adult children

	1	2	3	4	5
Dep: edunfb	OLS	IV	Red-form	OLS	OLS
male	0.801*	0.947*	0.957**	0.957**	1.226***
	(0.452)	(0.454)	(0.442)	(0.441)	(0.359)
female-first-born	0.314	-0.010	-0.008	-0.033	-0.132
	(0.639)	(0.796)	(0.819)	(0.646)	(0.558)
inter-male-femfb	**1.145**	**1.164***	**1.152**	**1.165***	**1.023***
	(0.680)	(0.663)	(0.749)	(0.657)	(0.574)
female-two-first			-0.033		
			(0.732)		
N-alive	-0.542***	-0.036			
	(0.165)	(0.791)			
Tharu					1.509**
					(0.697)
Bantar					-2.502***
					(0.816)
BC					4.043***
					(0.837)
Constant	8.506***	6.396*	6.246***	6.246***	5.550***
	(0.845)	(3.137)	(0.541)	(0.540)	(0.660)
Observations	787	787	787	787	787
R-squared	0.071	0.042	0.038	0.038	0.179
First-stage					
dep: N-alive					
male		-0.288			
		(0.204)			
female-first-born		-0.044			
		(0.265)			
inter-male-femfb		**0.324**			
		(0.249)			
female-two-first		0.918**			
		(0.381)			
Constant		4.168			
		(0.167)			

Kleiberger-Paap rk Wald F statistics for the clustered IV is 5.799
Without cluster-corrections the Cragg-Donald Wald F statistic would be 33.093
If only second-stage is clustered (as in ivreg, cluster()), then the first stage F-statistics is 19.44.
Bold means that femalefb+intmalfemfb is significant at the 5%-level (10%-level in the first stage)
Clustered at the PSU level standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10. Determinants of SLC among adult children

	1	2	3	4	5
Dep: slcnfb	OLS	IV	Red-form	OLS	OLS
male	0.004	0.008	0.014	0.014	0.029
	(0.045)	(0.043)	(0.048)	(0.048)	(0.043)
female-first-born	0.076	0.066	0.066	0.053	0.048
	(0.069)	(0.080)	(0.082)	(0.067)	(0.062)
inter-male-femfb	**0.093**	**0.093**	**0.087**	**0.094**	**0.082**
	(0.079)	(0.079)	(0.088)	(0.078)	(0.072)
female-two-first			-0.018		
			(0.059)		
N-alive	-0.036***	-0.020			
	(0.013)	(0.062)			
Tharu					0.081
					(0.061)
Bantar					-0.161***
					(0.050)
BC					0.353***
					(0.088)
Constant	0.403***	0.335	0.251***	0.251***	0.204***
	(0.067)	(0.251)	(0.041)	(0.041)	(0.050)
Observations	787	787	787	787	787
R-squared	0.033	0.030	0.019	0.019	0.095
First-stage					
dep: N-alive					
male		-0.288			
		(0.204)			
female-first-born		-0.044			
		(0.265)			
inter-male-femfb		**0.324**			
		(0.249)			
female-two-first		0.918**			
		(0.381)			
Constant		4.168			
		(0.167)			

Kleiberger-Paap rk Wald F statistics for the clustered IV is 5.799
Without cluster-corrections the Cragg-Donald Wald F statistic would be 33.093
If only second-stage is clustered (as in ivreg, cluster()), then the first stage F-statistics is 19.44.
Bold means that femalefb+intmalfemfb is significant at the 5%-level (10%-level in the first stage)
Clustered at the PSU level standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1