Working Paper in Economics No. 781

Son Preferences and Education Inequalities in India

Heather Congdon Fors and Annika Lindskog

Department of Economics, October 2019

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Son Preferences and Education Inequalities in India*

Heather Congdon Fors[1] and Annika Lindskog[2]

School of Economics, Business and Law, University of Gothenburg.

Abstract

We investigate the impact of son preferences in India on gender inequalities in educational performance and investment. Son preferences can create gender inequalities through two channels: preferential treatment of boys within the families and gender-biased fertility strategies (gender-specific fertility stopping rules and sex-selective abortions). We distinguish the impact of direct favoritism towards sons in general and towards eldest sons from the impact of fertility strategies. Our empirical strategy relies on dividing families based on the gender of the first-born, and employing family fixed effects. In short, our results suggest strong favoritism towards boys in general. In comparison to the advantage that boys have over girls within the family, the extra advantage of the eldest son is small and only statistically significant for pecuniary investment. The advantage of boys over girls within the family that can be attributed to gender-biased fertility strategies is also small in comparison to the advantage that can be attributed to favoritism of sons. We further document systematic differences between the families that girls and boys live in due to gender-biased fertility strategies, with implications for education investment and performance. Again, the resulting gender inequalities are smaller than the inequalities due to families' favoritism of sons over daughters. Regardless of the reason behind favoritism, the consequence is inequality of opportunities between girls and boys.

JEL codes: D13, I20, J16, O15

Keywords: Son preferences, Gender, Sex-selection, Fertility-stopping rules, Human Capital, Education, Birth order

*We gratefully acknowledge financial support from the Swedish Research Council. We are also thankful for insightful comments from Rohini Somanathan, Sonia Bhalotra, and from participants at presentations at University of Gothenburg, New Delhi School of Economics, and the 2019 ASWEDE conference.

[1]ORCID-id 0000-0002-3917-1099

[2] Corresponding author. E-mail annika.lindskog@economics.gu.se; Telephone +46-31-7864412; ORCID-id 0000-0003-1932-8674.

1. Introduction

Son preferences influence a wide range of behaviors in India. This is not the least apparent in the skewed sex ratios and the large number of missing girls and women (Sen, 1992; Clark, 2000; Klasen & Wink, 2002; Jha et al., 2006; Anderson & Ray, 2010; Milazzo, 2018). Studies have also documented unequal health investments between girls/women and boys/men (Arnold et al., 1998; Mishra et al., 2004; Jayachandran & Kuziemko, 2011; Dercon & Singh, 2013; Barcellos et al., 2014), as well as gender gaps in education investment within families (Azam and Kingdon, 2013; Kaul 2018).

In this paper, we investigate the impact of son preferences in India on gender inequalities in education investment and performance. While it is easy to imagine that son preferences should translate into unequal investments in girls compared to boys, it is less clear exactly through which channels this occurs. Gender gaps could result from within-household favoritism, but also from strategies that parents employ to have sons. One such strategy is to continue having children until the birth of a son, i.e. gender-specific fertility stopping. In families without boys, the birth of an additional girl increases the expected family size, thereby reducing the resources available for her (Jayachandran and Pande, 2017). In particular, parents' decision to continue childbearing and their decision on how soon they try to have another child will create inequalities between girls and boys in early-childhood investments, which has potential consequences for education performance and investment later in life. Jensen (2003) show how gender-specific fertility stopping leads to girls living in larger families than boys, and how this can create inequalities in education investment even if girls and boys were treated equally within families. The combination of gender-specific fertility stopping and favoritism towards sons could also create within-family gender inequalities. Families that would not favor sons over daughters for their preferred family size might do so if the family exceeds the preferred family size, such that credit-constraints become more binding.

The other strategy is sex-selective abortions.[1] It has been suggested that sex-selective abortions should reduce gender inequalities since girls will more often be born into families that actually

[1] Other possible strategies are infanticide and abandonment (Jeffery et al., 1984; Miller, 1987; Sudha and Rajan 1999), but these are likely to be less prevalent.

want them (Goodkind, 1996; Anukriti et al, 2016).[2] However, high SES families tend to have stronger son preferences (Edlund, 1999; Borker et al., 2017)[3] and are likely to desire fewer children, which will increase the use of sex-selective abortions if parents are keen to have at least one son (Chakraborty and Kim, 2010; Bhalotra and Cochrane, 2010). Therefore, sex selection is likely to exacerbate between family inequalities between boys and girls. We will refer to gender-specific fertility stopping rules and sex-selective abortions broadly as gender-biased fertility strategies. Whether gender inequalities are primarily due to favoritism towards sons or fertility strategies matters for the impact of different policy measures or courses of events on education gender inequalities.

In addition to possibly creating gender inequalities, the existence of gender-biased fertility strategies complicate the estimation of within-family favoritism. This is because gender cannot be treated as exogenous anymore. Note that it is not enough to employ family fixed-effects to deal with selection of gender in this case. Gender is endogenous within the family.

In our investigation of the impact of son preferences on gender inequalities, we distinguish the impact of direct favoritism towards sons in general from favoritism towards the eldest sons. We also distinguish the impact of favoritism from the impact of gender-biased fertility strategies. Favoritism by definition occurs within families, while gender-biased fertility strategies could create inequalities both within and between families. We use a wide range of education indicators, including indicators of performance (completed grades and test scores) as well as of time investment (enrollment and hours spent on school) and pecuniary investment (the private-public school choice and education expenditure). In addition, we estimate the impact on height-for-age, which is a potential link between inequalities in early-life environment and investment and later-

[2] To the extent that sex-selection and gender-specific fertility stopping rules are alternative strategies to ensure a son, availability of sex-selection could also decrease within-household inequalities, by crowding out gender-specific fertility stopping.

[3] Sex-selective abortions are more common among high caste families and well-off urban families than among economically disadvantaged and rural families (Chakraborty and Kim, 2010; Bhalotra and Cochrane, 2010). There can be many underlying reasons behind this. One is that better-off families have stronger reasons to want sons. Edlund (1999) shows that this will be the case if families care about marriage of their children. The skewed sex ratios resulting from sex-selective abortions and excessive mortality among girls makes men the abundant sex on marriage markets. High status men can still expect to get married, while low status men might end up single. For high SES families the choice is thus between a married daughter and a married son, while for low SES families the choice is rather between a married daughter and a potentially unmarried son. This gives low SES families less incentives to abort girls. There could also be differences in availability and affordability of sex-selective abortions, even though it appears to be widely available and cheap (Bhalotra and Cochrane, 2010).

life education outcomes. This is interesting since the earlier literature documents implications of gender-biased fertility strategies on early-life investments. Parents who aim for a son after the birth of a girl might for example invest less in her to save resources for the son they hope to have, or mothers might breastfeed shorter to try to get pregnant with a son (Barcellos et al., 2014; Jayachandran and Pande, 2017).

In short, our results suggest strong favoritism towards boys in general within families. In comparison to the within-family advantage that boys have over girls, the extra advantage of the eldest son is mostly small and statistically insignificant. However, eldest sons are favored over and above the advantage they have because of their gender and birth order for pecuniary investment. Gender-biased fertility strategies do not appear to drive within-household gender inequalities to a large extent. Perhaps surprisingly, given the literature on the impact of gender-specific fertility stopping on early life health investments, we do not find impacts on height-for-age among school age children. Gender-biased fertility strategies cause systematic differences in the types of families where girls live compared to the types of families where boys live. While we are not able to exactly quantify the implied education inequalities between boys and girls, we can estimate a lower bound. Though not negligible, these inequalities are smaller than those due to within-family favoritism of sons.

Our estimation strategy relies on both division of families into those with first-born girls and those with first-born boys and on family fixed effects. The division of families into those with first-born girls and those with first-born boys serves to distinguish between families that might use gender-biased fertility strategies from families that should not. Since sex-selection is not common for first births (Pörtner, 2013; Rosenblum, 2015, 2017), gender of the first-born can be considered random. However, the gender of the first-born has important consequences for the use of gender-biased fertility strategies. To ensure the birth of a son, families with a first-born girl might use either sex-selective abortions or gender-specific fertility stopping. Families with a first-born son have less reason to use such strategies. In short, we will treat gender of the first birth as exogenous and we will treat gender of all children as exogenous in the families with first-born boys. We will also use first-born boys' families as the counterfactual without gender-biased fertility strategies when we estimate the impact of these strategies.

Family fixed effects are crucial, but not enough, to estimate within-family favoritism. They will control for eventual family size, which is by necessity correlated with birth order, and in the Indian case also correlated with gender (Jensen, 2003). Since we observe families before many have completed childbearing and in general do not know which families have and have not completed childbearing, family fixed effects is the only way to fully control for eventual family size.

Our analysis proceeds in four steps. First, to estimate favoritism towards sons in general we use the sample of families with first-born boys. The estimated gender effect in this sample should not be affected by gender-biased fertility strategies. Since the first-born boy himself could be special, we compare later born boys in these families to later born girls, controlling for birth order.

Second, we estimate favoritism towards the eldest son, who in the earlier literature has been suggested to be particularly advantaged.[4] Favoritism towards boys may be targeted to eldest sons in particular, rather than towards sons in general. (Jayachandran and Pande, 2017; Kaul, 2018). We do so in a very straightforward way, by adding an eldest son dummy to a within-family estimation that also controls for both gender and birth order. In spite of the simplicity of this approach, we are to the best of our knowledge the first to use it in a model that fully accounts for birth order.[5]

Third, to estimate within-family gender inequalities that are due to gender-biased fertility strategies we compare the first-born girls and the first-born boys' samples. We essentially use a difference-in-difference strategy where first-born boys' families are used as the counterfactual without gender-biased fertility strategies and first-born girls' families are the "treated" group who might resort to gender-biased fertility strategies. The interaction term between first-born girl family and gender reveal if girls fare worse in the families that might resort to gender-biased fertility than in the families that do not.

Fourth, we investigate between-family inequalities due to fertility strategies. If boys are born into high SES families and girls into low SES families that also end up having more children, there should be implications for aggregate human capital investment into girls and into boys. To

[4] The desire of parents to have at least one son might be related to the role of the eldest son in providing for them in old age, taking over family land, and performing important rituals (Mullatti, 1995; Jayachandran and Pande, 2017).
[5] Kaul 2018 uses an eldest son dummy and controls for gender and being first born. However, conditional on not being first born, being the eldest son is still correlated with a lower birth order.

investigate if this is the case, we again compare the first-born girls' and the first-born boys' families. We first test whether there are systematic differences in the first-born girls' families regarding where subsequent girls end up in comparison to those where subsequent boys end up. Next, we estimate the consequence of the inferred difference between boys' families and girls' families for educational investment and outcomes. Since there can be additional differences between girls' and boys' families these will be lower bounds.

We contribute to the literature on son preferences in several ways. This literature has mostly considered early-life outcomes and/or survival of females versus males (Sen, 1992; Clark, 2000; Klasen & Wink, 2002; Jha et al., 2006; Anderson & Ray, 2010; Milazzo, 2018; Arnold et al., 1998; Mishra et al., 2004; Jayachandran & Kuziemko, 2011; Dercon & Singh, 2013; Barecello et al., 2014; Jayachandran & Pande, 2017). Education investment and outcomes has received less attention. While there are papers documenting gender gaps (Azam and Kingdon, 2013; Kaul 2018), these typically do not explain the connection to son preferences and gender-biased fertility strategies. The seminal paper by Jensen (2003) showed that gender-specific fertility stopping could and did create gender inequalities even for a subset of families treating girls and boys equally. We extend the analysis of Jensen (2003) by analyzing more recent data, when sex-selective abortions were widely available and fertility rates were much lower. The impact of gender-biased fertility strategies should be different in this changed context. We also use a different strategy to identify gender inequalities that are due to different sources: within-household discrimination, within-household inequalities due to gender-biased fertility strategies, and between household inequalities due to gender-biased fertility strategies. Moreover, we look at a wide range of education indicators.

As pointed out by Edlund (1999), the more frequent use of sex-selective abortions in high compared to low SES families implies that girls and boys are born into different families, thus potentially creating a female under-class. The combination of sex-selective abortions and gender-specific fertility stopping could create large differences in the types of families that girls and boys live in. We investigate this, and the implications thereof on education investment and outcomes.

We contribute to the literature that estimate gender gaps in education. In this literature, within-family gender gaps are typically interpreted as favoritism (Kingdon & Azam, 2013; Kaul, 2018) However, gender-biased fertility strategies imply that gender cannot be treated as exogenous in the Indian context. Moreover, family fixed effects, which is the strategy typically employed in the literature, is not enough to deal with endogeneity of gender. Within family gender gaps do not only reflect within-family favoritism of the boys, but are also influenced by gender-biased fertility strategies. We contribute to the literature by estimation of a within-household gender effect that can be interpreted causally.

Several authors suggest strong preferences for at least one son and favoritism of the eldest son rather than sons in general in India (Rosenblum, 2013; Jayachandran, 2017; Jayachandran and Pande, 2017; Kaul, 2018). Kaul (2018) uses a within-family model with a control for being the first-born child and finds that eldest sons (who can be of different birth order) face an advantage in enrollment and educational expenditure. We extend that analysis by controlling for the birth order of later-born children in the family in addition to a control for being the first-born. This is important given the systematic difference in the birth order of eldest sons in families where the first-born is a boy (where the eldest son is always the first-born child) versus families where the first-born is a girl (where the eldest son is never the first-born child). It is also important given that the latter family type is more likely to resort to gender-biased fertility strategies.

The remainder of the paper is structured as follows: in section 2 we describe the data and present total gender inequalities and evidence of gender-biased fertility strategies, while in section 3 we present the empirical strategy. In section 4 we present the empirical results, and section 5 concludes the paper.

2. Data and descriptive patterns

2.1 Data and variables

The data comes from two rounds of the India Human Development Survey (IHDS), collected in 2004-05 and 2011-12. This is a nationally representative survey of over 40000 households in India.[6] We use birth histories of women age 15 to 49 to create a sample of full siblings.[7] The sample includes cases where there is more than one sibship per household, as there are cases of extended families living in the same household. Multiple birth children (twins, triplets) are excluded from

[6]The data has been collected jointly by the University of Maryland in the United States and the National Council of Applied Economic Research in India. The surveys were administered via interviews conducted in the local language, and cover a wide variety of socioeconomic topics.

[7] We restrict the sample to families where both the mothers and their husbands have not been previously married. Divorce is very unusual in India, and only 3.9% of children have one or two previously married parents.

the analysis, since their birth order is not well-defined. We use household weights from the first round to account for the fact that there is some oversampling of certain groups in the data.

Our main estimation sample is children aged 6 to 17; however, test scores are only available for children age 8-11, and height-for-age is mostly available for children age 8-11. For a sibship to be included in the estimation sample there needs to be non-missing data from at least two children. Often there is data on more than one child from each of the two surveys. Sibships are also included if there is data from one child in the first round and another child in the second round.[8] For the outcomes where age ranges from 6 – 17, there are on average about 3 children per sibship. For the outcomes where the age range is 8 – 11, there are on average just over 2 children per sibship in the sample.

Our main explanatory variable of interest is gender. We create a dummy variable *female* that takes a value of one if the child is a girl, and zero otherwise. We also control for absolute birth order, using dummy variables for birth orders one, two, three and four plus,[9] for age using age dummies, and for survey year.

A particular strength of the data set is that it allows us to analyze an unusually rich set of educational variables. We can broadly categorize our dependent variables into educational performance of the child and educational investment. The indicators of child performance are the test scores on reading, writing and mathematics and the number of completed grades.[10] We use standardized test scores such that they measure age-specific standard deviations from the mean, using the sample population as the age-specific reference. [11,12]

The indicators of educational investment are *Enrollment*, *Total hours*, *Private school* and *School expenses*. The first two are indicators of time invested in schooling. *Enrollment* is a dummy

[8] This substantially increases the test scores estimation sample.
[9] Birth order 4 plus takes a value of one if the child's birth order is 4 or higher and zero otherwise.
[10] The tests are administered by the interviewers.
[11] The reading score runs from 0 (cannot read) to 4 (read a story), with the intermediate values 1 (letter), 2 (word) and 3 (paragraph). The writing score is equal to 0 if the child cannot write and 1 if the child can write with 2 or less mistakes. The math score runs between 0 (cannot count) and 3 (division), with the intermediate values 1 (number) and 2 (subtraction).
[12] The test scores variables are the same as Makino (2012) uses in her analysis. We have an additional round of data from 2010-11 and thus have a much larger sample of families with at least two children in the data. This allows us to rely on a within-sibship analysis.

variables taking a value of 1 if the child is enrolled in school and zero otherwise. [13] *Total hours* combines all hours related to schooling, including the hours in school, hours of homework and hours of private tuition per week used by the child.[14] In our regressions, we set *Total hours* to zero for all children who are not enrolled and estimate on the full sample.

We include two measures of investment into school quality: *Private school* and *School expenses*.[15] These outcomes are only collected for children who are enrolled in school [16] *Private school* is a dummy variable taking a value of 1 if the child attends a private school, and 0 if the child attends a public school. *School expenses* measures the cost of school fees, books, uniforms, bus fare and private tuition fees in rupees.

In addition to education indicators, we use the height-for-age z-score (HAZ). HAZ is relevant since it is a measure that will capture differences in early life investment and environment (Silventoinen, 2003; Li et al., 2003), and since gender-biased fertility strategies have been found to matter for early life investment. HAZ has been shown to be correlated with both health human capital and cognitive and non-cognitive skills (Glewwe et al., 2001; Alderman et al., 2001). The *HAZ* was constructed using the WHO reference tables from 2007. The height data is available for all ages, but was collected with a primary focus on children under 5 and between the ages 8 to 11, with other ages included based on availability at the time of the survey.

Table 1: Descriptive statistics

Variable	N	Mean	Std dev	Min	Max
Dependent variables					
Enrollment	69,523	0.848	0.359	0	1
Total hours spent on school in a week	63,064	35.400	19.661	0	216
Private school	55,620	0.266	0.442	0	1
School expenses in rupees	49,962	2587.703	5281.526	0	201000
Completed grades	69,514	4.500	3.203	0	16
Reading test score	8,578	2.371	1.408	0	4
Writing test score	8,478	0.663	0.473	0	1
Math test score	8,545	1.387	1.007	0	3
Height for age Z score	31,002	-1.842	1.452	-5.999	5.855

[13] While it is possible that children are enrolled without actually attending school, less than 1% of our enrolled sample report spending zero hours on schooling, and approximately 90% report spending at least 29 hours a week on schooling.

[14] Private tuition depends on pecuniary investment, but we still choose to include it to count all hours equally. If children did not study with private tutors they could instead study alone or with someone else.

[15] Though the effect of these investments on human capital accumulation remain unclear, parents are likely to make them with the intent to improve the child's human capital.

[16] Note that this, for example, implies that a family is dropped if only one child has attended school, even if there are more children of the right ages in the data.

Birth order 1	69,523	0.273	0.445	0	1
Birth order 2	69,523	0.327	0.469	0	1
Birth order 3	69,523	0.207	0.405	0	1
Birth order 4 or higher	69,523	0.194	0.395	0	1
Age 6	69,523	0.078	0.269	0	1
Age 7	69,523	0.085	0.279	0	1
Age 8	69,523	0.079	0.270	0	1
Age 9	69,523	0.069	0.254	0	1
Age 10	69,523	0.107	0.309	0	1
Age 11	69,523	0.071	0.256	0	1
Age 12	69,523	0.117	0.322	0	1
Age 13	69,523	0.091	0.288	0	1
Age 14	69,523	0.092	0.289	0	1
Age 15	69,523	0.081	0.272	0	1
Age 16	69,523	0.069	0.254	0	1
Age 17	69,523	0.061	0.239	0	1
Girl	69,523	0.484	0.500	0	1
In first-born girl family	69,523	0.523	0.499	0	1
Year 2011	69,523	0.473	0.499	0	1
Household variables (Part 4.D)					
Poor	24,692	0.254	0.435	0	1
Urban	24,692	0.282	0.450	0	1
Mother’s ed	24,692	4.074	4.570	0	16
Father’s ed	24,692	6.380	4.827	0	16
SC/ST	24,692	0.302	0.459	0	1
OBC	24,692	0.424	0.494	0	1
Brahmin	24,692	0.045	0.207	0	1
Muslim	24,692	0.125	0.331	0	1
Sibship size	24,692	3.101	1.238	2	13
Ideal no. kids	24,692	2.568	0.930	0	18

Descriptive statistics on all variables used in our analysis are presented in Table 1. Enrollment is rather high, at about 85%. The average height-for-age Z score is approximately -1.84. Though this is very low, and quite close to the limit for stunting, it is in line with earlier findings from India (Tarozzi, 2008).

2.2 The total education gender gap

To know whether within-family favoritism and gender-biased fertility strategies are important explanations of education inequalities we need a benchmark for comparison. Table 2 shows the correlations between gender and education outcomes from regressions that do not use family fixed effects and that do not distinguish between families with first-born boys and families with first-born girls. Note that these total inequalities are used merely as a benchmark. In particular, we will not decompose the total inequalities into different components. Our aim is to test the importance of various possible sources of gender inequalities, focusing on credible identification of the mechanisms. This implies that we will not consider all sources behind the total gender inequalities. For example, we do not estimate the difference in outcomes between first-born girls and first-born boys. While gender of the first-born can be treated as exogenous, we still cannot tell if differences in outcomes are because first-born boys are treated different from first-born girls, or because first-born girl families use gender-biased fertility strategies while first-born boy families do not. Another source of between-family variation that we do not consider is how total human capital investment in the family respond to child gender.

Table 2: The total gender inequalities in education - descriptive regressions

	Completed grades	Reading	Writing	Math	HAZ
Female	-0.054**	-0.129***	-0.133***	-0.208***	-0.123***
N	69,514	8,578	8,478	8,545	31,002
	Enrolled	Total hours	Private	Expenses	
Female	-0.034***	-1.886***	-0.048***	-447.343***	
N	69,523	63,064	55,620	49,962	

Note: The estimations also include a constant, a full set of child age dummies, and a year dummy.
* p<0.1; ** p<0.05; *** p<0.01. Standard errors, clustered at the sibship level, within parenthesis.

Girls exhibit a disadvantage compared to boys for all outcomes. Both the reading and the writing test scores are about 0.13 age specific standard deviations lower for girls than for boys, while the math test score is about 0.21 lower. Girls' HAZ is about 0.12 lower than boys' are. On average girls are 3.4 percentage points less likely than boys to be enrolled (the mean is 0.848), and they spend 1.9 hours less every week on their schooling (the mean is 35.4). If enrolled, girls are 4.8 percentage points less likely to be in a private school (the mean is 26.6 percent) and families spent 447.34 rupees less on their education (the mean is 2587.7 rupees).

2.2 Evidence on gender-biased fertility strategies in the data

There is clear evidence of gender-biased fertility strategies when looking at the number of females relative to males by birth order. When the sample is split between children who are the last-born in the family and those who are not, a striking pattern emerges. As can be seen in Figure 1, the number of girls per 1000 boys who are not the last-born is dramatically larger than the number of girls per 1000 boys who are the last-born, particularly at birth orders 2 and 3. This could be either because parents use sex selective abortions before the birth of their last child or because parents continue childbearing if they have a girl. Since we do not observe completed fertility of the mothers these numbers are likely to under-estimate true differences. Some last-born girls might not end up being last born.

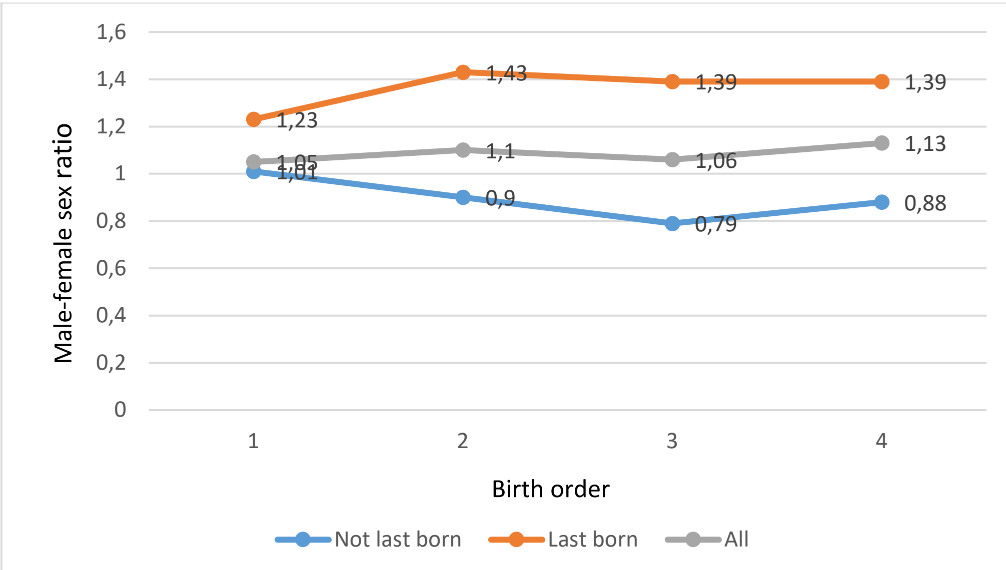

Figure 1: Ratio of boys to girls, by birth order

Our empirical strategy assumes that gender of the first-born is random. As can be seen in Figure 1 above, the sex ratio for all first-born children is 1.05, which is well within the range that is considered biologically normal (Anderson and Ray, 2010)[17]. Rosenblum (2017) tests for systematic differences in family characteristics that should be exogenous to the gender of the first-born child using the first round of the India Human Development Survey (IHDS), and finds no significant evidence of sex-selection. As we are using both the first and second round of the IHDS, we do similar tests for evidence of sex-selection among first-born children also in the data from the second round. We test for systematic differences in the following family characteristics that should be exogenous to the gender of the first-born child: parental age and education, caste, religion, and whether they live in an urban or rural location. The results are presented in table A1 in the appendix, and show essentially no significant differences between families with first-born girls versus families with first-born boys[18]. Hence, the gender of the first-born can be considered

[17] The biologically normal rate is 106.
[18] There is a weakly significant difference in father's age between first-born girls' and first-born boys' families, with fathers in first-born girls' families being on average 0.3 years older.

random and there should be no a priori systematic selection into families that have a first-born girl versus a first-born boy.

Even if first-born girls do not appear to end up in systematically different families than first-born boys, it is clear from the data that this does not apply to girls and boys in general. Gender-biased fertility strategies imply that girls and boys live in systematically different families: girls have on average significantly more siblings, are significantly more likely to be in a household that is below the poverty line, and are more likely to belong to a lower caste. The pattern is the same for children in families where the first-born is a girl.

Table 3: Differences in means, children up to age 17.

	Female		Male		
	N	Mean	N	Mean	diff
Sibship size	54,861	3.43	59,980	3.22	0.21***
Income per capita	54,065	11455	59,080	12062.96	-607.96***
Poor	54,813	0.302	59,901	0.276	0.026***
Ideal number of sons	48,708	1.51	52,720	1.54	-0.03***
SC/ST	54,402	0.32	59,481	0.31	0.007*
	First born girl family		First born boy family		
	N	Mean	N	Mean	diff
Sibship size	59,673	3.46	55,168	3.16	0.30***
Income pc	58,742	11574.85	54,403	11981.51	-406.66***
Poor	59,612	0.29	55,102	0.28	0.01***
Ideal number of sons	52,969	1.50	48,459	1.55	-0.05***
SC/ST	59,165	0.32	54,718	0.31	0.005
	Not last-born		Last-born		
	N	Mean	N	Mean	diff
Birth order 1 - %female	29,800	0.50	8,070	0.45	0.05***
Birth order 2 - %female	19,099	0.53	16,764	0.41	0.12***
Birth order 3 - %female	9,667	0.56	11,502	0.42	0.14***
Birth order 4+ - %female	9,036	0.53	10,903	0.42	0.11***

3. Empirical Strategy

To investigate the impact of son preferences on human capital inequalities we combine two main strategies. First, we identify families with first-born boys and those with first-born girls. The gender of the first-born should be largely exogenous in India despite sex-selective abortions since these are not common before the birth of the first child (Bhalotra and Cochrane 2010; Jha et al. 2011; Pörtner, 2015; Rosenblum, 2015, 2017). In the previous section, we confirmed that this holds in our data. However, as also indicated in the previous section investigating patterns in our data, the gender of the first-born leads to important differences between the families. Families with first-born boys have less reason to use either gender-specific fertility stopping rules or sex-selective abortions. This can be exploited in two important ways to learn about the mechanisms though which son preferences create gender inequalities. First, gender should be random also for later birth orders in the first-born boy sample, implying that the gender coefficient can be interpreted causally.[19] Second, we can compare families with first-born boys to families with first-born girls to find the impact of gender-biased fertility strategies on gender inequalities.

Our second main strategy is to employ family fixed effects models to identify within family inequalities. Families in our sample are nuclear ones, consisting of full siblings whose parents do not have any children with other partners. Given the extreme low levels of both out-of-wedlock births and divorces in India, this restriction excludes almost no one. A within-family specification ensures that we do not confuse differences in human capital accumulation across families, which depend for example on family size, with within-family inequalities.

In our first model, we aim to identify gender gaps that are due to within-household favoritism. Note that it is not enough to employ family fixed effects to achieve this. Fixed effects capture differences in unobserved preferences that matter for use of gender-biased fertility strategies, e.g. strength of son preferences and preferred family size: However, gender is endogenous within the family since the same family will use gender-biased fertility strategies differently depending on the gender of all previously born children. To estimate favoritism of boys we use the gender coefficients from within-family estimations in the sample of first-born boys' families. Since families that already have a son have no reason to employ gender-biased fertility strategies, gender is exogenous in this sample.[20]

(1)

$$y_{ist} = \alpha + \beta_1 * female_{is} + \beta_2 * bo2_{is} + \beta_3 * bo3_{is} + \beta_4 * bo4plus_{is} + \boldsymbol{age_{ist}}\pi + \varphi_t + \gamma_s + \varepsilon_{ist}$$

[19] Note, however, that the consequence of being born female in a family that has an eldest son might be different than it is in a family without sons.

[20] If gender is exogenous in first-born boy families it should not be systematically related to birth order, such that the inclusion of birth order controls should have no significant impact on the estimated *female* coefficient. Back of the envelope calculations of the difference in *female* coefficient with and without birth order controls, assuming no covariance between the regression coefficients, indeed indicate no significant difference.

The coefficient β_1 will give us the impact of within-family favoritism of sons over daughters. We use linear sibship fixed effects regressions for all outcomes in our main within-family estimations. For the binary outcomes *enrollment* and *private school* we therefore estimate the linear probability model. Standard errors are always clustered at the sibship level.

Our second model aims to identify a possible eldest son advantage. It uses a very simple and straightforward strategy: to add an eldest son dummy to a within-household estimation that controls for gender (*female*) and birth order (*bo*). Note that the eldest son dummy is not an interaction term since the eldest son could be of any birth order. We also control for a full set of age dummies and survey round.

The model is

(2)

$$y_{ist} = \alpha + \beta_1 * eldestson_{is} + \beta_2 * female_{is} + \beta_3 * bo2_{is} + \beta_4 * bo3_{is} + \beta_5 * bo4plus_{is} + \boldsymbol{age_{ist}}\pi + \varphi_t + \gamma_s + \varepsilon_{ist}$$

where y_{ist} is the outcome of child *i* in sibship *s* at time *t*, $\boldsymbol{age_{ist}}$ is a full set of child age dummies, and φ_t is a survey round dummy. γ_s are sibship fixed effects, which captures differences in family size, and all other time constant differences across families. Our coefficient of interest is β_1, which informs us about whether eldest sons are favored on top and above any potential benefits they have from being male and of their specific birth order.

Third, we aim to investigate the impact of gender-biased fertility strategies on within-household gender inequalities. For this purpose, a within-household model with an *fbg* interaction terms is used. The interest is on the coefficient of the interaction term between *female* and *fbg*. In essence, we use a difference-in-difference strategy where first-born boy families are used as the counterfactual without gender-manipulation strategies. We include a control for being first born, but no additional birth order controls. This is because later birth orders are endogenous and related to gender in families that apply gender-manipulation strategies. If birth order matters for education outcomes we still need to control for first births since these are systematically of different genders between the two types of families (even if gender of the first born is ex ante exogenous).

(3)

$$y_{ist} = \alpha + \beta_1 * female_{is} + \beta_2 * fbg_s + \beta_3 * (fbg_s * female_{is}) + \beta_4 * firstborn_{is} + \boldsymbol{age_{ist}}\pi + fbg_s * \boldsymbol{age_{ist}}\pi + \varphi_t + fbg_s * \varphi_t + \gamma_s + \varepsilon_{ist}$$

Fourth, we aim to find the effects of gender-biased fertility on between-family gender inequalities. Do gender-biased fertility strategies imply that girls end up in families that invest less in children's human capital? And what does this imply for aggregate human capital inequalities between girls and boys? Since human capital investment is not likely to be fixed, but to respond to child gender we cannot directly test if girls end up in families that invest less. If girls live in families that invest less in education, this could be either because girls ended up in types of families that invest less or because the families invest less when they have more girls. For the same reason, we cannot simply compare models with and without family fixed effects. The total between-family inequality will not only capture the fact that girls and boys end up in different types of families, but also responses in these families to child gender.

We employ a two-step strategy. In the first step we test whether and how much gender-biased fertility strategies affect the types of families that girls and boys end up in. We do this by regressing a female dummy, an *fbg* family dummy, and an interaction term between these two dummies on family characteristics that are likely to matter for human capital accumulation. Again, first-born boy families can be seen as providing the counterfactual, not affected by gender-biased fertility strategies.

(4) $$x_{is} = \alpha + \beta_1 * female_{is} + \beta_2 * fbg_s + \beta_3 * (fbg_s * female_{is}) + \boldsymbol{age_{ist}}\pi + fbg_s * \boldsymbol{age_{ist}}\pi + \varepsilon_{is},$$

where x_{is} are family characteristics. We do not include birth order controls since these are endogenous to gender-biased fertility strategies. If the interaction term is statistically significant, gender-biased fertility strategies lead to systematic differences between the families that girls and boys end up in. If families with first-born boys do not employ gender-biased fertility strategies and if the dependent variable is predetermined, such that it cannot respond to child gender, the coefficient on the female dummy should not be statistically significant.

In the second step, we estimate the correlation between the family characteristics investigated in the first step and the education indicators (Eq. 5).

(5) $$y_{is} = \sum x_{is}\gamma + \varepsilon_{is}$$

Note that it is not important whether the family characteristic's impact on the education outcome is causal or not for our purpose. If girls, for example, more often end up in families where parents have less education, they will on average fare worse than boys, whether the impact of parents education on the education outcome is causal or not. Since we are not likely to include all family characteristics that matter for children's education investment and outcomes, we will estimate lower bounds.

4. Results

A. Favoritism towards sons in general

To find exogenous gender effects, that are not affected by gender-biased fertility strategies, we consider the female dummy among later born children in first-born boy families. Table 4 and 5 present the female and birth order effects in first-born boy families.

Table 4: The effect of gender and birth order in families with first-born boys on indicators of education performance – coefficients from linear sibship fixed effects estimations

	Completed grades	Reading	Writing	Math	HAZ
Female	-0.009	-0.113**	-0.083	-0.152***	-0.081*
	(0.043)	(0.055)	(0.063)	(0.051)	(0.043)
Second born	-0.583***	-0.162**	-0.106	-0.233***	-0.265***
	(0.048)	(0.067)	(0.074)	(0.062)	(0.049)
Third born	-0.847***	-0.328***	-0.113	-0.348***	-0.475***
	(0.084)	(0.098)	(0.117)	(0.096)	(0.079)
Fourth+ born	-0.964***	-0.549***	-0.234	-0.566***	-0.935***
	(0.137)	(0.145)	(0.176)	(0.141)	(0.127)
R^2	0.68	0.03	0.02	0.05	0.05
N	33,148	3,914	3,864	3,899	14,258
Sibships	10,678	1,838	1,816	1,831	5,242

Note: The estimations also include a constant, a full set of child age dummies, a year dummy, and sibship fixed effects. * p<0.1; ** p<0.05; *** p<0.01. Standard errors, clustered at the sibship level, within parenthesis.

Table 5: The effect of gender and birth order in families with first-born boys on educational investment – coefficients from linear sibship fixed effects estimations

	Enrolled	Total hours	Private	Expenses
Female	-0.029***	-1.592***	-0.021***	-267.991***
	(0.008)	(0.372)	(0.007)	(56.392)
Second born	-0.030***	-1.903***	-0.026**	-439.443***
	(0.008)	(0.391)	(0.011)	(83.020)
Third born	-0.021	-2.186***	-0.020	-564.831***
	(0.013)	(0.645)	(0.023)	(138.506)
Fourth+ born	0.004	-1.727*	0.004	-709.967***
	(0.022)	(1.019)	(0.046)	(206.370)
R^2	0.14	0.12	0.01	0.16
N	33,153	30,029	26,328	23,589
Sibships	10,679	9,848	9,176	8,443

Note: The estimations also include a constant, a full set of child age dummies, a year dummy, and sibship fixed effects.

* p<0.1; ** p<0.05; *** p<0.01. Standard errors, clustered at the sibship level, within parenthesis.

Controlling for unobserved family characteristics and restricting the analysis to families that should not resort to gender-biased fertility strategies does make a difference to estimated female coefficients. The estimated female coefficients are smaller than in the descriptive estimations in Table 2. However, within-family favoritism matters - girls do face a disadvantage compared to boys of the same birth order even within families that are not expected to use gender-biased fertility strategies. This disadvantage is significant for all outcomes with the exception of *Completed grades* and *Writing*. The greatest reduction in the female effect is on completed grades and pecuniary investments.

B. Directly testing favoritism towards the eldest son

In this sub-section we investigate within-household favoritism of the eldest son. In the case of school performance and HAZ, the eldest son coefficient is statistically insignificant and generally small in magnitude. This indicates that the potential advantage that eldest sons have in these measures is explained by their birth order and gender. Similarly, eldest sons are not favored within households in terms of time investments. Eldest sons do, however, appear to be favored in terms of pecuniary measures of investment, over and above what their birth order and gender can explain. Eldest sons are 2.5 percentage points more likely to attend a private school than what their gender and birth order can explain, and families spend about 152 extra rupees on their education. In summary, eldest sons are favored in terms of pecuniary investment, but this does not translate into better school performance as measured here, and it does not remove the negative birth order and female effects within families. Hence, favoritism towards the eldest son does not appear to be a main reason behind inequalities between the genders.

Table 6: The effect of gender and birth order on indicators of education performance in models with an eldest son dummy – coefficients from linear sibship fixed effects estimations

	Completed grades	Reading	Writing	Math	HAZ
Eldest son	0.034	-0.012	0.031	0.045	0.001
	(0.034)	(0.044)	(0.050)	(0.043)	(0.035)
Female	-0.031	-0.141***	-0.121**	-0.176***	-0.096***
	(0.035)	(0.044)	(0.050)	(0.041)	(0.036)
Second born	-0.504***	-0.164***	-0.092*	-0.179***	-0.235***
	(0.033)	(0.044)	(0.051)	(0.044)	(0.035)
Third born	-0.805***	-0.294***	-0.121	-0.321***	-0.448***
	(0.056)	(0.071)	(0.086)	(0.070)	(0.060)
Fourth+ born	-0.931***	-0.402***	-0.221*	-0.521***	-0.788***
	(0.093)	(0.107)	(0.119)	(0.101)	(0.095)
R^2	0.68	0.03	0.02	0.04	0.05
N	69,514	8,578	8,478	8,545	31,002
Sibships	21,453	3,993	3,947	3,977	11,046

Note: The estimations also include a constant, a full set of child age dummies, a year dummy, and sibship fixed effects.
* p<0.1; ** p<0.05; *** p<0.01. Standard errors, clustered at the sibship level, within parenthesis.

Table 7: The effect of gender and birth order on educational investment in models with an eldest son dummy - coefficients from linear sibship fixed effects estimations

	Enrollment	Hours	Private	Expenses
Eldest son	0.006	0.224	0.025***	151.779***
	(0.006)	(0.310)	(0.006)	(56.270)
Female	-0.031***	-1.604***	-0.032***	-380.594***
	(0.006)	(0.309)	(0.006)	(47.901)
Second born	-0.027***	-1.768***	-0.008	-279.134***
	(0.006)	(0.273)	(0.008)	(60.291)
Third born	-0.029***	-2.413***	-0.006	-356.590***
	(0.009)	(0.444)	(0.014)	(93.112)
Fourth+ born	-0.018	-2.492***	-0.011	-504.566***
	(0.014)	(0.700)	(0.025)	(140.058)
R^2	0.15	0.12	0.02	0.15
N	69,523	63,064	55,620	49,962
Sibships	21,454	19,852	18,630	17,190

Note: The estimations also include a constant, a full set of child age dummies, a year dummy, and sibship fixed effects.
* p<0.1; ** p<0.05; *** p<0.01. Standard errors, clustered at the sibship level, within parenthesis.

C. The impact of gender-biased fertility strategies on gender inequalities within families

We here investigate the impact of gender-specific fertility stopping on inequalities between girls and boys using an interaction between the indicator that a child is in a first-born girl family (*fbg*) and the female indicator.

In terms of education performance, the effect of being female in a family where the first-born is a girl is always negative, and in some cases relatively large, but is not statistically significant. In the case of human capital investment, girls in families with first-born girls once again exhibit more negative female effects, and these effects are statistically significant in the case of pecuniary investments. Therefore, fertility strategies only have a significant negative impact on girls' education in the case of pecuniary investments. The significantly more negative female effect in pecuniary investments may indicate that girls in first-born girl families more often face binding financial constraint.

Table 8: The effect of gender in families with first-born girls versus first-born boys on indicators of education performance - coefficients from linear sibship fixed effects estimations

	Completed grades	Reading	Writing	Math	HAZ
Female	-0.023	-0.091*	-0.073	-0.144***	-0.072*
	(0.041)	(0.055)	(0.063)	(0.051)	(0.043)
Female*fbg	-0.081	-0.058	-0.124	-0.019	0.020
	(0.054)	(0.072)	(0.094)	(0.078)	(0.062)
R^2	0.68	0.02	0.02	0.03	0.04
N	69,514	8,578	8,478	8,545	31,002
Sibships	21,453	3,993	3,947	3,977	11,046

Note: The estimations also include a constant, a full set of child age dummies, a year dummy, and sibship fixed effects.
* p<0.1; ** p<0.05; *** p<0.01. Standard errors, clustered at the sibship level, within parenthesis.

Table 9: The effect of gender in families with first-born girls versus first-born boys on educational investment - coefficients from linear sibship fixed effects estimations

	Enrolled	Total hours	Private	Expenses
Female	-0.030***	-1.593***	-0.020***	-255.713***
	(0.008)	(0.368)	(0.007)	(55.581)
Female*fbg	-0.002	0.110	-0.032***	-288.008***
	(0.011)	(0.567)	(0.011)	(101.255)

R^2	0.15	0.12	0.02	0.15
N	69,523	63,064	55,620	49,962
Sibships	21,454	19,852	18,630	17,190

Note: The estimations also include a constant, a full set of child age dummies, a year dummy, and sibship fixed effects.
* p<0.1; ** p<0.05; *** p<0.01. Standard errors, clustered at the sibship level, within parenthesis.

D. The impact of gender-biased fertility on between household gender inequalities

Do gender-biased fertility strategies create gender inequalities due to girls and boys ending up in different types of families? In table 10 below, we investigate if because of fertility strategies girls are born in to families that differ from those that boys are born in to with regard to family characteristics that matter for education investment and performance. We consider mostly predetermined characteristics such as parents' education, religion and caste. Urban residence and total household income is also likely to be largely predetermined. Sibship size is, however, likely to respond to child gender, through gender-specific fertility stopping behavior. Below are difference-in-difference estimates where the first difference is gender of the first-born and the second difference is gender of the child in question. The first-borns themselves are excluded from the analysis.

Table 10: Differences in gender and family-type composition by family characteristics

	Mother's education	Father's education	Poor	Sibship size	Ideal no children
Female	0.042	-0.010	0.014	0.089***	-0.001
	(0.074)	(0.095)	(0.009)	(0.029)	(0.021)
Fbg family	0.169*	0.241*	-0.006	0.101**	-0.042
	(0.098)	(0.126)	(0.013)	(0.042)	(0.036)
Female*fgb	-0.183*	-0.161	0.026*	0.241***	0.117***
	(0.106)	(0.135)	(0.014)	(0.042)	(0.032)
R^2	0.01	0.00	0.01	0.04	0.01
N	40,283	37,886	40,502	40,506	37,728
	Urban	SC/ST	OBC	Brahmin	Muslim
Female	-0.005	0.008	-0.007	0.002	-0.001
	(0.007)	(0.009)	(0.010)	(0.004)	(0.007)
Fbg family	-0.002	0.013	-0.016	0.005	-0.034***
	(0.009)	(0.015)	(0.014)	(0.005)	(0.009)
Female*fgb	0.005	0.006	0.008	-0.009*	0.012
	(0.010)	(0.014)	(0.014)	(0.005)	(0.009)

R^2	0.01	0.00	0.00	0.00	0.00
N	40,449	40,449	40,449	40,449	40,506

Note: The estimations also include a constant and a full set of child age dummies.
* p<0.1; ** p<0.05; *** p<0.01. Standard errors, clustered at the sibship level, within parenthesis.

Results suggest that gender-biased fertility strategies result in important differences in the families that boys and girls end up in. In first-born girl families, boys on average end up in families with better-educated mothers than girls, and more often in higher caste families. Boys less often end up in poor families and in families where the mother express a stronger preference for many children. They also end up in smaller families. Girls end up in larger families than boys even in the first-born boy families, and first-born girl families are in general larger than first-born boy families.

What are the implied gender inequalities of the fact that girls and boys live in different types of families? In tables 11 and 12 below the family characteristics are regressed on education indicators. Coefficients from table 10 and tables 11-12 can then be used to predict resulting gender inequalities in education outcomes. These are displayed in table 13.

Table 11: Correlations between family characteristics and indicators of education performance – coefficients from a linear regression.

	Completed grades	Reading	Writing	Math	HAZ
Poor	-0.151***	-0.178***	-0.151***	-0.175***	-0.264***
	(0.036)	(0.028)	(0.032)	(0.027)	(0.037)
Urban	0.128***	0.113***	0.092***	0.166***	0.086***
	(0.026)	(0.021)	(0.022)	(0.022)	(0.030)
Mother's edu	0.048***	0.032***	0.035***	0.035***	0.020***
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
Father's edu	0.060***	0.031***	0.017***	0.029***	0.008**
	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
SC/ST	-0.197***	-0.072**	-0.113***	-0.116***	-0.137***
	(0.044)	(0.032)	(0.033)	(0.031)	(0.045)
OBC	-0.080**	0.027	-0.088***	-0.035	-0.074*
	(0.033)	(0.027)	(0.027)	(0.029)	(0.042)
Brahmin	-0.115**	0.097*	-0.041	0.051	-0.063
	(0.050)	(0.052)	(0.051)	(0.060)	(0.065)
Muslim	-0.367***	-0.049	-0.025	-0.052	-0.008
	(0.043)	(0.034)	(0.036)	(0.037)	(0.051)
Sibship size	-0.155***	-0.063***	-0.068***	-0.057***	0.015
	(0.017)	(0.011)	(0.011)	(0.010)	(0.013)
Ideal no. kids	-0.120***	-0.071***	-0.019	-0.053***	-0.045**

	(0.023)	(0.013)	(0.015)	(0.013)	(0.018)
R^2	0.72	0.17	0.12	0.18	0.03
N	24,686	12,469	12,377	12,419	16,962

Table 12: Correlations between family characteristics and measures of current education investment – coefficients from a linear regression.

	Enrolled	Total hours	Private	Expenses
Poor	-0.033***	-3.062***	-0.109***	-1,212.642***
	(0.007)	(0.352)	(0.008)	(60.502)
Urban	0.008*	1.483***	0.211***	1,672.406***
	(0.004)	(0.291)	(0.008)	(124.337)
Mother's edu	0.005***	0.387***	0.007***	221.107***
	(0.001)	(0.038)	(0.001)	(19.901)
Father's edu	0.012***	0.595***	0.014***	127.422***
	(0.001)	(0.037)	(0.001)	(12.663)
SC/ST	-0.013*	-1.288***	-0.061***	-885.157***
	(0.007)	(0.443)	(0.010)	(136.701)
OBC	-0.002	0.334	0.018**	-632.142***
	(0.006)	(0.364)	(0.009)	(149.363)
Brahmin	0.011	-0.546	0.071***	76.008
	(0.007)	(0.626)	(0.019)	(282.633)
Muslim	-0.054***	-3.738***	-0.009	-398.210***
	(0.009)	(0.452)	(0.010)	(144.952)
Sibship size	-0.011***	-0.970***	0.002	-164.953***
	(0.002)	(0.137)	(0.003)	(36.179)
Ideal no. kids	-0.019***	-1.513***	0.000	-100.038**
	(0.004)	(0.214)	(0.004)	(45.662)
R^2	0.20	0.19	0.20	0.19
N	24,686	23,395	22,964	18,817

Table 13: Lower bounds on gender inequalities due to girls and boys ending up in systematically different types of families

	Completed grades	Reading	Writing	Math	HAZ
Later-born female in fbg family	-0.079	-0.040	-0.033	-0.037	-0.014
	Enrolled	Total hours	Private	Expenses	
Later-born female in fbg family	-0.009	-0.699	-0.006	-152.709	

Note that the gender inequalities in table 13 are lower bounds, since we might miss important family characteristics. Later-born girls in first-born girl families do face a disadvantage on all outcomes as compared to boys in first-born girl families. For most outcomes the inequalities are between 1/6 and 1/3 of the descriptive total inequalities between girls and boys in Table 2 (about 1/6 for HAZ and private school enrollment, 1/5 for math scores, 1/4 for writing scores and enrollment, 1/3 or more for reading writing scores, hours, and rupees spent on education). The disadvantage is, however, smaller than the within-household disadvantage that girls face (in tables 5 and 6) for all outcomes except completed grades.

5. Discussion and Conclusion

We show that son preferences create inequalities in education investment and outcomes between girls and boys in several ways. The most straightforward way is favoritism of boys within the family. In a context where some families use gender-specific fertility stopping rules or sex selective abortions, gender is not exogenous and within-family discrimination cannot be estimated by simply adding family fixed effects. To estimate a gender effect that is due to within-family favoritism only we use a within-family model on a sub-sample that is unlikely to use gender-biased fertility strategies: families with first-born boys. Favoritism towards sons within the family turns out to be the quantitatively most important way in which son preferences create gender inequalities in education. Boys are clearly favored over girls.

Eldest sons have been suggested to be particularly favored in Indian families. We estimate the advantage that eldest sons enjoy on top and above the advantage of their gender and birth order. Eldest sons in particular are favored with regard to private school enrollment and education expenses, but not for other outcomes. For the probability of private school enrollment, their extra advantage is larger than that of boys in general (which they also enjoy), while it is smaller for education expenses.

Gender-specific fertility stopping rules can lead to within-family inequalities, especially for early-life investments, while sex-selective abortions might instead work to reduce within-family inequalities. To identify the impact of gender-biased fertility strategies on education indicators we compare the impact of being a girl in a sub-sample that is likely to use gender-biased fertility strategies (first-born girl families) with a sub-sample that is not likely to do so (first-born boy families). We find that gender-biased fertility strategies create additional gender inequalities with regard to the probability of private school enrollment and education expenses, but not for other indicators. A possible reason is that families that end up larger due to gender-specific fertility stopping rules face more binding resource constraints and prioritize boys. Perhaps surprisingly given the literature that has found impacts of gender-specific fertility stopping on early-life health investments, we do not find that gender-biased fertility strategies create gender inequalities in height for age among school-aged children. We also find no impacts on test scores, grade completion and enrolment.

Both gender-specific fertility stopping rules and sex-selective abortions, and in particular their combination, could result in girls ending up in systematically different types of families than boys. Gender-specific fertility stopping rules imply that girl will on average live in families that ended up larger, which might results in less human capital investment for all children in the family. Sex-selective abortions tend to be used more often by better-off families, increasing the share of boys in these-better off families. The total between-family inequality will not only capture the fact that girls and boys end up in different types of families, but also the human capital investment responses in these families to child gender. To get at least a lower bound on between-family education inequalities between boys and girls that can be attributed to gender-manipulation strategies, we use a two-step procedure. In the families that use gender-manipulation strategies, i.e. first-born girl families, we first estimate if there are systematic differences in the types of families that girls and boys live in with regard to family characteristics that matter for education investment and performance. We thereafter estimate the association between these family characteristics and the education indicators. While girls do more often end up in families with worse outcomes on the education indicators, the implied inequalities are generally smaller than those due to within-family favoritism of boys.

Gender-biased fertility strategies have been shown to have significant effects on a range of early-life outcomes (Barcello et al., 2014; Jayachandran and Pande, 2017), and the large number of missing and unwanted girls in India is a tragedy. We show, however, that gender-biased fertility strategies have limited consequences on education inequalities between boys and girls in comparison to unequal treatment within the family (that is not due to gender-biased fertility strategies). Families might have good reasons to invest more in boys than in girls (Kumar, 2013;

Rosenblum, 2017).[21] Independent of the underlying reasons, favoritism towards boys implies that Indian girls and Indian boys do not have equal education opportunities. Favoritism towards boys within the family thus provide a rationale for corrective policies favoring girls of all backgrounds, not only focusing on household characteristics.

Compliance with ethical standards

This study was funded by Vetenskapsrådet (The Swedish Research Council) grant number 348-2014-5963. The authors declare that they have no conflict of interest.

[21] A thorough investigation of the underlying reasons behind within-household favouritism of sons is beyond the scope of this paper. However, the geographical variation in within-household favouritism is not correlated with geographical differences in returns to education between women and men (available from the authors on request).

References

Alderman H, Behrman J, Lavy V, Menon R (2001) Child health and school enrollment: a longitudinal analysis. *Journal of Human Resources* 36:185–205.

Anderson, S., & Ray, D. (2010). Missing women: age and disease. *The Review of Economic Studies*, *77*(4), 1262-1300.

Anukriti, S. (2018) Financial Incentives and the Fertility-Sex Ratio Trade-Off. *American Economic Journal: Applied Economics*, 10(2), 27-57.

Anukriti, S., Bhalotra, S. and H. Tam (2016). “On the Quantity and Quality of Girls: New Evidence on Abortion, Fertility, and Parental Investments”, IZA Discussion Paper No. 10271.

Arnold, F., Choe, M. K., & Roy, T. K. (1998). Son preference, the family-building process and child mortality in India. *Population studies*, *52*(3), 301-315.

Arnold, F., Kishor, S., & Roy, T. K. (2002). Sex-selective abortions in India. *Population and development review*, 28(4), 759-785.

Azam, M., & Kingdon, G. G. (2013). Are girls the fairer sex in India? Revisiting intra-household allocation of education expenditure. *World Development*, *42*, 143-164.

Barcellos, S. H., Carvalho, L. S., & Lleras-Muney, A. (2014). Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently?. *American Economic Journal: Applied Economics*, *6*(1), 157-189.

Bhalotra, S. R., & Venkataramani, A. (2015). Shadows of the captain of the men of death: Early life health interventions, human capital investments, and institutions.

Bhalotra, S. and T. Cochrane (2010). “Where Have All the Young Girls Gone? Identification of Sex Selection in India”, IZA Discussion Paper No. 5381.

Borker, G., Eeckhout, J., Luke, N., Minz, S., Munshi, K., & Swaminathan, S. (2017). *Wealth, marriage, and sex selection*. Working Paper, Cambridge University.

Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. Journal of health economics, 24(2), 365-389.

Chakraborty, T. and S. Kim (2010). "Kinship institutions and sex ratios in India." *Demography* 47(4), 989-1012.

Chamarbagwala, R. (2011). Sibling composition and selective gender-based survival bias. *Journal of Population Economics*, 24(3), 935-955.

Clark, S. (2000). “Son preference and sex composition of children: Evidence from India”. *Demography*, *37*(1), 95-108.

Congdon Fors, H. and A. Lindskog (2017). “Within-Family Inequalities in Human Capital Accumulation in India: Birth Order and Gender Effects”. Department of Economics, School of Business, Economics and Law, University of Gothenburg.

Currie, J., & Almond, D. (2011). Human capital development before age five. In *Handbook of labor economics* (Vol. 4, pp. 1315-1486). Elsevier.

De Haan, M., Plug, E., & Rosero, J. (2014). ”Birth Order and Human Capital Development Evidence from Ecuador”. *Journal of Human Resources*, 49(2), 359-392.

Dercon, S., & A. Singh (2013). “From nutrition to aspirations and self-efficacy: gender bias over time among children in four countries”, *World Development*, 45, 31-50.

Doyle, O., Harmon, C. P., Heckman, J. J., & Tremblay, R. E. (2009). Investing in early human development: timing and economic efficiency. *Economics & Human Biology*, *7*(1), 1-6.

Edlund, L. (1999). Son preference, sex ratios, and marriage patterns. *Journal of political Economy*, *107*(6), 1275-1304.

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

Glewwe P, Jacoby H, King E (2001) Early childhood nutrition and academic achievement: a longitudinal analysis. *Journal of Public Economics* 81:345–68.

Heckman, J. J., & Masterov, D. V. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, *29*(3), 446-493.

Jayachandran, S and R Pande (2017) “Why are Indian children so short? The Role of Birth Order and Son Preference”, *American Economic Review*, 107(9), 2600-2629.

Jayachandran, S. and I. Kuziemko (2011). “Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India”, *The Quarterly Journal of Economics*, 126(3), 1485–1538

Jensen, R. T. (2003). *Equal Treatment, Unequal Outcomes?: Generating Sex Inequality Through Fertility Behaviour*. Bazaar Chintan.

Jha, P., Kumar, R., Vasa, P., Dhingra, N., Thiruchelvam, D., & R. Moineddin (2006). ”Low male-to-female sex ratio of children born in India: national survey of 1· 1 million households”, *The Lancet*, 367(9506), 211-218.

Kaul, T. (2018). Intra-household allocation of educational expenses: Gender discrimination and investing in the future. *World Development*, *104*, 336-343.

Klasen, S., & Wink, C. (2002). A turning point in gender bias in mortality? An update on the number of missing women. *Population and Development Review*, *28*(2), 285-312.

Kumar, A. (2013). “Preference based vs. market based discrimination: Implications for gender differentials in child labor and schooling”. *Journal of Development Economics*, 105, 64-68.

Li H, Stein AD, Barnhart HX, Ramakrishnan U, Martorell R (2003). “Associations between prenatal and postnatal growth and adult body size and composition”. *Am J Clin Nutr*.,77:1498–505.

Makino, M. (2018). “Birth Order and Sibling Sex Composition Effects Among Surviving Children in India: Enrollment Status and Test Scores”, *The Developing Economies*,56(3), 157-196.

Milazzo, A. (2018). “Why are adult women missing? Son preference and maternal survival in India”, Journal of Development Economics, 134, 467-484.

Mishra, V., Roy, T. K., & R.D. Retherford (2004). “Sex differentials in childhood feeding, health care, and nutritional status in India”, *Population and development review*, 30(2), 269-295.

Mullatti, L. (1995). Families in India: Beliefs and realities. *Journal of Comparative family studies*, 11-25.

Pörtner, C. (2015) “Sex-Selective Abortions, Fertility, and Birth Spacing”, World Bank Policy Research Working Paper No. 7189

Rosenblum, D. (2017) “Estimating the Private Economic Benefits of Sons Versus Daughters in India”, *Feminist Economics*. 23(1), 77-107.

Rosenblum, D. (2013) “The Effect of Fertility Decisions on Excess Female Mortality in India”, *Journal of Population Economics*. 26(1):147-180.

Sen, A. (1992). Missing women. *BMJ: British Medical Journal*, *304*(6827), 587.

Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of biosocial science*, *35*(02), 263-285.

Tarozzi, A. (2008). Growth reference charts and the nutritional status of Indian children. *Economics & Human Biology*, *6*(3), 455-468.

Tenikue, M., & Verheyden, B. (2010). Birth order and schooling: Theory and evidence from twelve sub-saharan countries. *Journal of African Economies*, 19(4), 459-495.

Table A1: Differences in means – children up to age 17

	First born girl family		First born boy family		
	N	Mean	N	Mean	diff
Mother's age	8,691	30.99	8,900	30.76	0.23
Father's age	7,705	35.70	7,919	35.40	0.30*
Mother's education	8,751	5.44	8,954	5.46	-0.02
Father's education	7,751	7.16	7,959	7.24	-0.08
Brahmin	8,681	0.04	8,880	0.04	-0.001
Other Backward Class	8,681	0.43	8,880	0.43	0.000
Scheduled Caste/Tribe	8,681	0.31	8,880	0.30	0.01
Muslim	8,774	0.13	8,977	0.14	-0.004
Urban	8,693	0.27	8,901	0.27	-0.007