Son Preference and Gender Inequality

Deepankar Basu* Robert de Jong†

September 19, 2007

Abstract

This paper draws out some implications of son targeting fertility behaviour for gender inequality in developing economies. It is demonstrated that such behaviour has two notable implications at the aggregate level: (a)larger number of siblings for girls (Sibling Effect), and (b)a higher within-family birth order for boys (Birth Order Effect). While the first tends to worsen gender inequality through monetary factors, the second does so in terms of non-monetary factors. Empirically testing for these effects, we find that both are present in many countries in South Asia, South-East Asia and North Africa but are absent in the countries of Sub-Saharan Africa. Using maximum likelihood estimation, we proceed to study the effect of covariates on son targeting and fertility behaviour for India, a country which displays significant sibling and birth-order effects.

JEL Codes: J1, J7.
Keywords: son preference, stopping rules, gender inequality.

1 Introduction

Gender inequality is a pervasive phenomenon in many developing countries. Even a cursory glance at the indicators of well-being like literacy, infant mortality, life expectancy, primary, secondary or college enrollment rates, differentiated along gender lines, makes this amply clear. This has not only been well documented (see, for instance, the various issues of the

*Department of Economics, Ohio State University, 317 Arps Hall, Columbus, OH 43210, email: basu.15@osu.edu. I would like to thank Anindya Bhattacharya, Stephen Cosslett, Sugato Dasgupta, Trevon Logan, Anjan Mukherji, Alita Nandi, G. V. Ravindra and participants in the Micro Lunch at the Department of Economics, Ohio State University for helpful comments.

†Department of Economics, Ohio State University, 429 Arps Hall, 1945 N High Street, Columbus, OH 43202, email: dejong@econ.ohio-state.edu.

World Development Indicators, published by the World Bank) but also widely commented on [Bardhan, 1974, 1982; Sen and Sengupta, 1983; Sen, 1990; Filmer et al., 1998; Clark, 2000; Jensen, 2002].

To explain gender inequality, many scholars have surmised that the intra-household allocation of resources might be skewed in favour of male children [Bardhan, 1974, 1982; Sen and Sengupta, 1983; Clark, 2000]. Many studies have even attempted to derive the gender differentiated expenditure on children's development using inter-temporal optimization models of family expenditure. If this expenditure differential against female children were indeed true, it would go a long way in explaining the continuing disadvantaged position of women in these societies.

Econometric studies to test the hypothesis of biases in intra-household allocation of resources against the girl child have, however, largely produced negative results. Numerous papers following the method pioneered in Deaton (1989) and using data from a wide cross-section of developing countries have almost all come up with the finding that intra-household biases are statistically insignificant (see Jensen, 2002 for a list of references to this econometric literature). Nevertheless, two recent studies have cast doubt on these findings by reporting the presence of systematic resource biases against girls within families (Burgess and Zhuang, 2000; Clark, 2000).

Our concern in this paper will not be an engagement with these conflicting findings on the presence/absence of intra-household biases against the girl child. Instead, we propose to develop and test an alternative framework to explain gender inequality in developing countries. The simple model that we develop in this paper shows that even in the absence of intentional biases against female children at the household level, a disadvantaged position for them can be generated at the aggregate level by fairly plausible assumptions regarding fertility behaviour of families.

To develop this alternative framework, we use a well-recognized aspect of developing countries (in particular, the countries of East, South and South-East Asia and North Africa) namely, son preference. By this we mean, following anthropologists and sociologists (Arnold et al., 1998; Clark, 2000; Jensen, 2002), the existence in society of a strong preference for male as opposed to female offspring. Furthermore, this strong preference is reflected in son targeting fertility behaviour, also referred to in the literature as differential stopping behaviour (DSB) or male-preferring stopping rules [for instance see, Clark, 2000]. The main idea behind such stopping rules is that the sex composition of already-existing children determines the subsequent fertility behaviour of families.[1]In our model, we concretize this idea as follows: couples continue childbearing until they reach their "desired" number, k, of sons or when they hit the ceiling for the maximum number,N, of children that they think to be feasible (given their resource constraints).

We show that there are two important implications of such fertility behaviour for gender inequality. First, girls will be born into relatively larger families sharing resources with a larger sibling cohort; we call this the "sibling effect". Second, boys will be born as relatively younger children within families; we call this the "birth order effect". Both have important

[1]For evidence on this see (Arnold, et al., 1998 and Larsen, et al., 1998).

implications for gender inequality even in the absence of intra-household allocation biases, as we discuss in detail later.

Yamaguchi (1989)[2] works in terms of the following stopping rule: couples continue childbearing until they get their "desired" number,k, of sons. Thus, in contrast to the stopping rule we use, there is no limit on the maximum number of children that couples can have in their attempt to attain the target for boys. It can be easily seen that our framework is more general; Yamaguchi (1989) is a special case of our model if we let N go to infinity. It should be noted that the analysis becomes substantially simplified in this case (when N can increase without bounds) because the number of children in a family follows a standard probability distribution (the negative binomial distribution); when N is a finite integer, we no longer have any standard distribution. Additionally, our framework is more realistic; it seems unreasonable to assume that couples have no limit on the number of children they can produce. In the sample for India, for instance, about 94 percent of the households have five or less children.

Jensen (2002) arrives at results similar to what we have called the sibling effect, though he uses a different stopping rule. In his model, couples want n children and b boys; but if they reach n children with less than b boys, they continue childbearing until they attain b boys or reach some maximum number of children, $n + k$. This stopping rule is a variant of that used by Seidl (1995); this is also a special case of our model with the desired number of sons $k = b$ and the maximum number of children $N = n + k$. However, there are two major differences between our work and Jensen's (2002). First, while we discuss both the sibling effect and the birth order effect, Jensen (2002) limits himself only to the former. Second, unlike Jensen (2002) we use household level data on birth sequences and desired family size to estimate the full model with MLE.

To focus attention on the issue of birth order and to put this paper in perspective, we would like to differentiate between mean *absolute* and mean *relative* birth order. The mean absolute birth order of boys (girls) is computed by averaging the position of the male (female) child within the sequence of births in his (her) family, where averaging is done over all children in the population. On the other hand, to compute mean relative birth order of boys (girls) in the population, we first calculate the average position of male (female) children within each family and then average over all families.

The difference between absolute and relative birth orders is important because the latter affects gender inequality in a sense that the former does not. To understand this crucial point consider two children born into different families which are identical in all respects (income, education of parents, etc.). Suppose the first child under consideration is a first-born in a family with only one child; and the second child is a first-born in a family with four children. Are the prospects for development of both children the same? We think not. The difference is especially important in the context of poorer families in developing countries where a substantial burden of child-rearing falls on the shoulders of elder children. In the second scenario considered above, the position of the first-born will be much less favourable

[2]Similar ideas can also be found on pp. 342 in Ray (1998); his suggestions to solve the problems, in terms of the male preferring stopping rule, is very similar to Yamaguchi (1989).

for self-development because (s)he will have to share in the responsibilities of caring for the younger children in the family. In the first scenario, the first-born will not be burdened with these responsibilities and will therefore be able to devote much more time and energy for her/his own development. It is this difference that will be captured by the mean relative birth order but not by the mean absolute birth order.

An example might further clarify the difference. Consider the following scenario from Jensen (2002): couples want one child but will have a second one if the first is not a boy. In this case, half the families will end up with a boy, one-fourth with a girl and a boy (in that order) and another fourth with two girls. Now, if we compute the mean absolute birth order of boys and girls, we see that it is the same for both at 4/3. Two-thirds of both boys and girls are first-born children while one-third are second-born, and so the mean absolute birth order is 4/3 (= 1x2/3+2x1/3) for both boys and girls.

Let us now compute the mean relative birth orders. To proceed, we will define an average within-family birth order (AWFBO) score for boys and girls respectively in each family and then average across all families. Notice that boys have a AWFBO score of 1 in those families which have only one child (half of the families) and a AWFBO score of 2 in those families which have a girl as the first and a boy as the second child (one-fourth of the families). Girls, on the other hand, have a AWFBO score of 1 for families which have a girl as the first child and a boy as the second child (one-fourth of the families); and a AWFBO score of 3/2=(1+2)/2 for families with two girls. Hence the mean relative birth order for boys is 4/3 (= 1x2/3+2 x1/3) because we average across families; in a similar way, the mean relative birth order for girls is 5/4 (= 1x1/2+3/2x1/2). Not only are they different, boys have a higher relative birth order than boys.

Note that when Yamaguchi (1989) refers to his result on birth order he means the *relative* birth order; he concludes his analysis by stating that "male-preferring stopping rules do not have a differential effect on the mean birth order of boys and girls" (Yamaguchi, 1989, pp.459). This result, which is at variance with ours, is really an artifact of the assumption that N can increase without bounds in his model. If we limit N to a finite integer, as we have done in our model, the result no longer holds; the mean relative birth order for boys turns out to be greater than for girls, as the above example shows and as we demonstrate later[3]. On the other hand, when Jensen (2002) refers to the birth order he means the *absolute* birth order; his conclusion that "fertility-related characteristics such as birth order...do *not* differ between boys and girls..." (Jensen, 2002, pp.6) is a statement about mean absolute birth orders. Thus, even though both Yamaguchi (1989) and Jensen (2002) seem to have arrived at the same result, they are really referring to different measures of birth order.

To summarize, the contribution of this paper is two fold: one, the male-preferring stopping rule that we analyze in this paper is both more realistic and general than that found in Yamaguchi (1989); and two, we highlight not only the sibling effect but also the birth order effect[4], an issue that seems to have been neglected so far in the literature (for eg. Jensen,

[3]See Theorem 2 below.

[4]Whenever we mention birth order without any qualification in this paper, we will mean the relative birth order.

2002). Additionally, we confront our model with data from a wide range of countries in Asia and Africa in a novel manner, using both descriptive statistics and maximum likelihood estimation. To the best of our knowledge, such empirical analysis does not currently exist in the economics, sociology or demography literature.

Our empirical results show that both the sibling and the birth order effects are present in a host of countries in East Asia, South Asia, South-East Asia and North Africa; these effects are absent in countries of Sub-Saharan Africa. When we compute these same effects for the India states, we find their strong presence in the states of North, West and Central India; the effects are absent in Kerala, and several states in North-Eastern India. This is more or less in line with anecdotal evidence on patriarchal tendencies in the different geographical regions of India. Our maximum likelihood estimation for India reveals that formal education of women has a large and significant impact on both desired fertility and son-targeting: education reduces both family size and the probability of targeting behaviour.

Before moving onto the main text of the paper, we would like to briefly comment on a possible source of criticism of our approach. It might be objected that this paper does not have any family-level optimization model to undergird the analysis. We have deliberately avoided an optimization-based analysis, along the lines of Becker (1960) or Becker and Lewis (1973) because our intention is merely to highlight the *effects* of son targeting behaviour, not to derive it from first principles. Additionally, our results about the sibling effect and birth order effects will be in operation whenever families implement male-preferring stopping rules on their fertility behaviour, no matter what the 'optimal' family size or son-target is. Hence, it seemed unimportant to derive the exact numerical values of the optimal family size or son-target via optimization exercises.

The rest of the paper is organized as follows: the next section presents the main results; section 3 tests the empirical implications of the model with data from a wide range of countries in Asia and Africa; and the next section concludes the discussion with some policy implications. Proofs of all propositions are collected together in Appendix A.

2 Main Results

Many developing economies in South, East and South-East Asia and North Africa display "son preference", a strong preference for male children (Clark, 2000). This preference is reflected in fertility behaviour in the form of male-preferring DSB. We take this as given and try to draw out its implications for gender inequality.

2.1 Son Preference and the Sex Ratio at Birth

In our model, 'son preference' affects fertility behaviour through DSB; more concretely, DSB implies the following stopping rule: couples continue childbearing till they attain a desired 'target' number of sons (k) or hit a ceiling for the maximum number of children (N). Thus, two quite distinct stopping rules influence childbearing decisions: the target number of sons and the maximum number of children. These two stopping rules operate precisely in that order and couples stop childbearing whenever one of them becomes effective. We assume that the probability of a male and female birth are equal.

Table 1 gives the various possible completed family structures (in terms of the number of children and their birth sequence) that would emerge in a population practicing DSB (in the sense defined above) where the target number of sons is k and the ceiling for the maximum number of children is N, with $k \leq N$. The first column gives the total number of children in a family; the second and third columns give the number of boys and girls respectively; and the last column gives the probability of occurrence of the family in the population. We use the following notation: $\binom{n}{k}$ refers to the number of ways of choosing k out of n objects.

Note that each couple would keep continue childbearing till they reach the k^{th} male child or hit the ceiling (N) for the maximum number of children. In such a scenario, it is easy to see that the minimum number of children born in any family would be k; no family would stop at less than k children. Couples would stop when they get $k, k+1, k+2, \ldots, N$ children.

Table 1: Family Structures and Number of Sibling

Total Children	Boys	Girls	Sibling per Child	Probability
k	k	0	$k-1$	$(1/2)^k$
$k+1$	k	1	k	$\binom{k}{1}(1/2)^{k+1}$
$k+2$	k	2	$k+1$	$\binom{k+1}{2}(1/2)^{k+2}$
$k+3$	k	3	$k+2$	$\binom{k+2}{3}(1/2)^{k+3}$
$\vdots$		$\vdots$		$\vdots$
$N-1$	k	$N-1-k$	$N-2$	$\binom{N-2}{N-1-k}(1/2)^{N-1}$
N	k	$N-k$	$N-1$	$\binom{N-1}{N-k}(1/2)^N$
N	$k-1$	$N-k+1$	$N-1$	$\binom{N}{k-1}(1/2)^N$
N	$k-2$	$N-k+2$	$N-1$	$\binom{N}{k-2}(1/2)^N$
$\vdots$		$\vdots$		$\vdots$
N	1	$N-1$	$N-1$	$\binom{N}{1}(1/2)^N$
N	0	N	$N-1$	$(1/2)^N$

Furthermore, any couple that stops at k children would do so only if all the k children are male. Similarly couples would stop at $k+1, k+2, \ldots, N-1$ children only if the last child is born as the k^{th} male child. So, for all families with total number of children between k and $N-1$, the binding constraint on fertility behaviour is the intended target of k male children; such families stop childbearing because they get the "desired number" of male children, i.e., k male children.

For families with N children, it no longer matters whether the target for boys has been attained or not; the binding constraint becomes the total number of children already born.

Couples stop childbearing because they have already hit the ceiling (which is finite). Note that for families with N children, there can be several possibilities in terms of the number of boys that couples finally end up with: couples could get $0, 1, 2, \ldots, k$ male, and the rest female, children.

Since the number of children in a randomly chosen family does not have a standard probability distribution, we need to demonstrate that the sample space is properly specified. The following proposition, by demonstrating that the probabilities of all possible outcomes of the conceptual experiment enumerated in Table 1 (i.e., entries in the last column) add up to unity, ensures that the probability space has been properly specified.

Proposition 1 *Let* $P(N,k) = \sum_{i=k}^{N} \binom{i-1}{k-1} (\frac{1}{2})^i + \sum_{i=0}^{k-1} \binom{N}{i} (\frac{1}{2})^N$ *; then* $P(N,k) = 1$, $\forall 1 \leq k < N$.

Proof See Appendix A.

The first result that we would like to draw attention to is the following: DSB does not have any effect on the sex ratio at birth (SRB). We note this as:

Observation 1 Let $\bar{S} = 1$ denote the ratio of newborn male to female children; then $\bar{S} = 1, \forall N, 1 \leq k < N$.

Proof: This is obvious by the use of any standard law of large numbers. Childbearing in different families can be considered to be identical and indepenedent experiments. Thus the averge number of male and female children will converge to their expected values (probability of a male and female birth). We have assumed that male and female births have equal probability (the probability is bounded away both from zero and unity), and so the result follows. □

This result has an important implication for countries in Asia and Africa which have witnessed declining sex ratios and SRB's in recent decades. Since the presence of DSB by itself does not affect the SRB, a declining trend in SRB's probably implies that sex-selective abortion of female foetuses or other forms of female infanticide are gaining ground as has been recently reported in the press[5].

2.2 The Sibling Effect

The next implication of DSB that we wish to highlight is that even when there are no intra-household biases operating against female children, there might be grounds to expect a markedly disadvantageous situation for her in the aggregate. This is simply because the number of siblings per girl child is, on average, more than the number of siblings per male child. So, even when there is no bias in intra-household allocation of resources, female children, on average, will get less resource for their development.

[5]For instance, see the coverage in BBC at `http://news.bbc.co.uk/2/hi/south_asia/4592890.stm`

To see this, let us compute the expected number of siblings for male and female children. We will do so by calculating the expected number of siblings for a child conditional on the sex of the child. Using Table 1, it can be seen that the expected number of siblings for a child given that it is a boy, $\bar{M}_s$, becomes:

$$\bar{M}_s = \frac{C(k,N) - (N-1)(\frac{1}{2})^N}{1 - (\frac{1}{2})^N} \tag{1}$$

and the expected number of siblings for a child given that it is a girl, $\bar{F}_s$, is:

$$\bar{F}_s = \frac{C(k,N) - (k-1)(\frac{1}{2})^k}{1 - (\frac{1}{2})^k} \tag{2}$$

where,

$$C(k,N) = \sum_{i=k}^{N} (i-1)\binom{i-1}{k-1}(\frac{1}{2})^i + (N-1)\sum_{i=0}^{k-1}\binom{N}{i}(\frac{1}{2})^N \tag{3}$$

In our model, we define the "sibling effect" as the difference in the expected number of siblings for female and male children:

$$SE = \bar{F}_s - \bar{M}_s. \tag{4}$$

Now, we are in a position to state one of the main results of this paper.

Theorem 1 $SE > 0, \forall N, 1 \leq k < N.$

Proof See Appendix A.

A positive sibling effect implies that, on average, female children will have more siblings than male children. Intuitively, this comes about because of DSB. If families practice DSB then they are more likely to stop childbearing if they have boys, rather than girls, in the lower parities (i.e. early in the birth history). Put another way, couples are more likely to continue childbearing if they have girls in the early parities. Thus, female children will, on average, have more siblings than male children. If we assume that endowment (or income) is divided equally across all families in society (which, in fact, it is not) and if we assume further that there is no intra-family bias against female children, even then female children, on average, will get lower resources for their health and educational development.

Of course, income is not divided equally in society; there are significant differences in the average incomes of rich and poor households. It is also a well-known fact that poorer families are also the larger families. Hence, the structural disadvantages of female children will, in reality, get heightened by these twin factors of larger family size and lower average family income for those larger families. This disadvantage can partially explain the worse performance of female children in comparison to male children in terms of most indicators of social well-being.

2.3 The Birth Order Effect

As we have noted earlier, DSB implies that couples are more likely to stop childbearing if they have boys, rather than girls, in the early parities. One implication of this is that boys will be born as relatively younger children with families. We try to capture this quantitatively using the notion of an average within family birth order (AWFBO) score and then draw out its implications for gender inequality. As a simple example will immediately demonstrate, the AWFBO score for boys (girls) measures the relative position of boys (girls) *within* the birth history of the family.

For instance, consider a family with the following birth sequence: BGBBG (where B refers to a boy and G to a girl and time moves from left to right). Here, the first-born child was a boy, the second-born was a girl, ..., the last-born (i.e., the youngest) was a girl. For this family, the AWFBO score for boys would be 8/3=(1+3+4)/3 and the AWFBO score for girls would be 7/2=(2+5)/2. Note, in passing, that families with no boys (girls) will not have a AWFBO score for boys (girls).

To compute the mean AWFBO scores for boys and girls in the population, we will use Table 2 (a variant of Table 1). The first and last columns of Table 2 are identical to the corresponding columns in Table 1; the second and third columns give the AWFBO scores for boys and girls respectively. Note that the first row does not have a AWFBO score for girls because these families have no female children; similarly, the last row does not have a AWFBO score for boys because these families do not have male children.

To facilitate the computation, we have divided the families in the population (equivalently, the rows of Table 2) into four groups. Group one has families with k children where all k are boys (first row of Table 2); group two has families with k boys and more than k children (first $N-k$ rows other than the first row); group three has families with N children and at least one boy (the last k rows apart from the last row); and group four has families with N female children (last row).

In all the families belonging to the second group, the last child will always be a male child; that is precisely why these families stop childbearing at that stage. In general, a family of size n in this group (where n runs from $k+1$ to N) gets the desired number of k boys, only when the last child is a boy and the other $k-1$ boys are uniformly distributed among the first $n-1$ births. So, the AWFBO score for boys, in such a case, is:

$$\frac{1}{k}\{\sum_{i=1}^{n-1} i(\frac{k-1}{n-1}) + n\} = \frac{n}{2}(1+\frac{1}{k}) \tag{5}$$

Proceeding in a similar manner, we can find the AWFBO score for girls in the second group of families. In the general case of n total children, there will be $(n-k)$ girls, and since the last child is a boy, these $(n-k)$ girls will be uniformly distributed over the first $(n-1)$ births. So, the AWFBO score for girls in such a scenario is:

$$\frac{1}{n-k}\{\sum_{i=1}^{n-1} i(\frac{n-k}{n-1})\} = \frac{n}{2} \tag{6}$$

Table 2: Family Structure and AWFBO Scores

Total Children	AWFBO score (boys)	AWFBO score (girls)	Probability
k	$\frac{k+1}{2}$		$(1/2)^k$
$k+1$	$\frac{k+1}{2}(1+\frac{1}{k})$	$\frac{k+1}{2}$	$\binom{k}{1}(1/2)^{k+1}$
$k+2$	$\frac{k+2}{2}(1+\frac{1}{k})$	$\frac{k+2}{2}$	$\binom{k+1}{2}(1/2)^{k+2}$
$k+3$	$\frac{k+2}{2}(1+\frac{1}{k})$	$\frac{k+2}{2}$	$\binom{k+2}{3}(1/2)^{k+3}$
$\vdots$		$\vdots$	$\vdots$
$N-1$	$\frac{N-1}{2}(1+\frac{1}{k})$	$\frac{N-1}{2}$	$\binom{N-2}{N-1-k}(1/2)^{N-1}$
N	$\frac{N}{2}(1+\frac{1}{k})$	$\frac{N}{2}$	$\binom{N-1}{N-k}(1/2)^{N}$
N	$\frac{N+1}{2}$	$\frac{N+1}{2}$	$\binom{N}{k-1}(1/2)^{N}$
N	$\frac{N+1}{2}$	$\frac{N+1}{2}$	$\binom{N}{k-2}(1/2)^{N}$
$\vdots$		$\vdots$	$\vdots$
N	$\frac{N+1}{2}$	$\frac{N+1}{2}$	$\binom{N}{1}(1/2)^{N}$
N		$\frac{N+1}{2}$	$(1/2)^{N}$

This gives us the entries in the first part of columns two and three in Table 2.

Families in group three are characterized by the following: there is no *a priori* rule to suggest what the gender of the last child will be. In fact, in all the families in this group, boys and girls are uniformly distributed among the N birth sequences. In such a situation the AWFBO score for boys and girls will be the same and will be given by $(N+1)/2$. For instance, a family with j boys and N children, will have the following AWFBO score for boys:

$$\frac{1}{j}\sum_{i=1}^{N} i\frac{j}{N} = \frac{N+1}{2}$$

which is independent of j. Similarly, the AWFBO score for girls in such families is given by

$$\frac{1}{N-j}\sum_{i=1}^{N} i\frac{N-j}{N} = \frac{N+1}{2}$$

which is also independent of j. Moreover, the two are equal.

For families in group one (where there are no gilrs), the AWFBO score for boys will be $(k+1)/2$; and for families in group four (where there are no boys), the AWFBO score for girls will be $(N+1)/2$.

Using Table 2, we can now compute the mean relative birth order for boys and girls by averaging across families. The mean relative birth order for boys is

$$\bar{M}_{BO} = \frac{1}{1-(1/2)^N}\{\frac{k+1}{2^{k+1}} + \frac{k+1}{k}\sum_{i=k}^{N}\binom{i-1}{k-1}(i/2)(\frac{1}{2})^i + (\frac{1}{2})^N(N+1)/2\sum_{i=1}^{k-1}\binom{N}{i}\}. \quad (7)$$

Similarly, the mean relative birth order for female children is

$$\bar{F}_{BO} = \frac{1}{1-(1/2)^k}\{\frac{N+1}{2^{N+1}} + \sum_{i=k}^{N}\binom{i-1}{k-1}(i/2)(\frac{1}{2})^i + (\frac{1}{2})^N(N+1)/2\sum_{i=1}^{k-1}\binom{N}{i}\}. \quad (8)$$

Now we are in a position to state our next result.

Theorem 2 $\bar{M}_{BO} > \bar{F}_{BO}, \forall N, 1 \leq k < N.$

Proof See Appendix A.

There are two important implications of theorem 2. One: boys, on average, will benefit more from the real income growth of society than girls because they enter the world at a stage when the per capita resources of society are at a higher level. We can think of the value of the resources that go into the upbringing of any child as the present value of the whole future income stream that is devoted for the child's development calculated at the time of the birth of the child. Even assuming that there is no systematic bias in favour of any child because of its gender, boys, on average, will get a higher value of resources because each term in their present value sum will be higher than for girls. This is simply a reflection of the fact that boys are born later relative to girls or are, on average, younger children in any family. In the context of a grwoing economy, this will mean a better starting point. But this effect will be swamped by another effect: by being born later into families, boys will have to share resources with a larger number of siblings over the years. Since the addition of siblings into a family divides resources at a rate far higher that the growth rate of any developing economy, the birth-order effect will generally tend to improve gender equity.

Note that the precise magnitude of this effect will depend on the average time span between the first and last births within families in the population. If the average time span is large, children born in the higher parities (boys in the main) will tend to benefit because parental income will be supplemented by elder children entering the work force; if, on the other hand, the time span is short, children born in the lower parities (mainly girls) will not have a large advantage because new-born children will arrive pretty soon. These effects seem to suggest an optimal time span which will improve gender equality the most. [6]

Of course, if we look at non-monetary aspects of gender inequality, the story becomes a little more complicated: the fact that girls, on average, are born as elder children in families will act against them in the context of a developing economy. In poorer and larger families, where both parents work to make ends meet, part of the parental responsibility

[6] Computation of the optimal rate is beyond the scope of this paper.

towards younger children will be passed on to elder children in the family. Since most of the elder children are girls, this responsibility will be disproportionately borne by them. Being burdened with sundry housework and the responsibilities associated with caring for younger children, these girls will not be able to devote their full time and energy towards their own education or other recreational activities. This, it must be stressed, is not the result of any intentional bias in the household against girls; all elder children in larger and poorer families will have to shoulder the responsibility of bringing up the younger children. Since girls are, on average, more likely to be born as elder children, the lion's share of this responsibility fall on their shoulder, despite the best intentions of parents. Additionally, it must be noticed that this kind of disadvantageous position of girls will not get reflected in household expenditure data. So, even when there are no expenditure biases against female children within families (which most econometric studies seem to suggest), a substantially disadvantageous position for the girl child can be expected on the basis of the non-monetary effects of the birth order effect alone. This largely counteracts the positive impact of the birth-order effect for girls, if any.

3 Empirical Analysis

Empirical analysis in this paper is carried out in two steps. In the first step, we test for the presence of sibling and birth order effects in the sample as evidence of DSB; in the second step, we estimate the effect of covariates on son targeting behaviour, total fertility rates and the interaction between the two (using household level data on birth histories) with the method of maximum likelihood.

3.1 Sibling and Birth Order Effects

We use data from the Demographic and Health Surveys (formerly known as the World Fertility Survey and the Contraceptive Prevalence Survey), which is part of a standardized survey conducted in over 70 developing countries by USAID; we use data from the latest available survey or the one closest to the year 2000[7]. Apart from being a comprehensive survey covering almost all relevant aspects of health and educational indicators, it also provides detailed information on the birth history of the interviewed women (between the ages of 15 and 49 years). The detailed birth history allows us to know the exact family structure and birth sequence for each of the interviewed women, and thus test our hypotheses regarding the sibling effect and the birth order effect[8]. We present results for several countries in South Asia, South-East Asia, North Africa and Sub-Saharan Africa in Tables 5 and 6. We also present results for India disaggregated at the state level in Table 7.

The results are along expected lines. Most of the countries in South Asia, South-East Asia and North Africa display significant sibling and birth order effects. Countries in Sub-

[7]Details of the data-set for each country is given in Table 3 and 4

[8]DHS data can be downloaded, with prior permission, from the following website: http://www.measuredhs.com/

Saharan Africa, on the other hand, do not throw up any statistically significant sibling or birth order effects. A population which has both high fertility (high value of N) and a high numerical value of the son-target (high value of k) can have small or zero sibling and birth order effects. This, rather than the absence of son preference, seems to be the case for countries in Sub-Saharan Africa. This conclusion emerges from two facts. One, when we numerically compute the sibling and birth order effects, we see that whenever N and k are close together, sibling and birth order effects are small or zero. Two, commonly used measures of son preference (like the ratio of desired sons to daughters, the proportion of families using contraceptive after two sons compared to those using contraceptives after two girls, etc) show that many countries of Sub-Saharan Africa display strong son preference[9].

For India, the two effects are strong for the states in the Northern, Central and Western regions; Kerala in the South and the states of the North East generally do not show these effects. This is in accord with much evidence (anecdotal and otherwise) on the prevalence of patriarchal practices in different geographical regions of the country.

3.2 The Empirical Model

In the second step of the empirical analysis, we analyze the effect of important covariates on son targeting behaviour, fertility behaviour and the interaction between the two for countries which display the presence of DSB (as seen in the previous sub-section). To do so, we estimate the parameters of our simple model using the method of maximum likelihood.

We start by introducing some notation. Let S_i denote the completed birth sequence for the i^{th} family; for instance, S_i could be BBG (where B stands for boy and G for girl). Let N_i denote the maximum number of children that family i would like to have; let k_i denote the target number of boys for family i. Let X_i denote a vector of covariates which determine the probability of son targeting behaviour for family i; let Z_i denote a vector of covariates which determines fertility behaviour (i.e., the desired maximum number of children, N_i) of family i.

Note that this analysis concerns the population of families with completed birth histories. To estimate the effect of covariates on targeting and fertility behaviour, we will calculate the joint likelihood of observing a given birth sequence (S_i) *and* maximum number of children (N_i); in other words, we will compute $P(S_i, N_i)$ where $P(.)$ denotes probability. To do so, we proceed as follows.

We introduce T_i, a dichotomous *unobservable* variable which indicates whether family i targets sons or not. $T_i = 0$ means that the family does not target; and $T_i = 1$ implies that family i is a son targeter. Finally, we let T_i be determined by a vector of covariates, X_i, in the following manner:

$$T_i = \begin{cases} 0 & \text{if} \quad X_i'\beta + \varepsilon_i \leq 0 \\ 1 & \text{if} \quad X_i'\beta + \varepsilon_i > 0 \end{cases} \tag{9}$$

[9] We do not report these results in any depth here because they are not the main issues of this paper; details are, however, available from the authors upon request.

where X_i is a $(k \times 1)$ vector of co-variates which determine whether a particular family 'targets' sons or not, β is a $(k \times 1)$ vector of parameters to be estimated and $\varepsilon_i \sim N(0, 1)$. To allow for the fact that targeting is unobservable, we make it stochastic.

To get the likelihood for the observed birth sequence and maximum number of children for the i^{th} family, note that

$$
\begin{aligned}
&P(S_i, N_i) = P(S_i, N_i|T_i = 0)P(T_i = 0) + P(S_i, N_i|T_i = 1)P(T_i = 1) \\
&= P(S_i|N_i, T_i = 0)P(N_i|T_i = 0)P(T_i = 0) \\
&\quad + P(S_i|N_i, T_i = 1)P(N_i|T_i = 1)P(T_i = 1) \\
&= I(n(S_i) = N_i)(\frac{1}{2})^{N_i} P(N_i|T_i = 0)P(T_i = 0) + P(S_i|N_i, T_i = 1)P(N_i|T_i = 1)P(T_i = 1) \\
&= I(n(S_i) = N_i)(\frac{1}{2})^{N_i} P(N_i|T_i = 0)\Phi(-X_i'\beta) \\
&\quad + P(S_i|N_i, T_i = 1)P(N_i|T_i = 1)[1 - \Phi(-X_i'\beta)] \qquad (10)
\end{aligned}
$$

where $\Phi(.)$ denotes the standard normal cdf, $I(.)$ denotes the indicator function and $n(S_i)$ denotes the number of children in birth sequence S_i.

Note that, because of DSB, when a family does not target sons, the effective stopping rule for childbirth becomes the maximum number of children that the family desires to have, N_i; hence the probability of observing S_i given N_i in such a case (i.e., $P(S_i|N_i, T_i = 0)$) is $I(n(S_i) = N_i)(\frac{1}{2})^{N_i}$. The indicator function is meant to rule out the possibility that a family which stops childbearing before hitting the desired maximum number of children could be a non-targeter. Any family which stops childbearing before hitting the ceiling, N_i, has to be a son targeter in our model. This gives us the first term in (10). The second term in (10) comes from son targeters; so we need to compute $P(S_i|N_i, T_i = 1)$.

Note that when a family targets sons, it's target, k_i, can range anywhere from 1 to $N_i - 1$; targeting $k_i = N_i$ sons with a ceiling for the maximum number of children at N_i is equivalent to not targeting. Since we cannot observe k_i (the target number of sons for a family), we condition on k_i and then integrate it out as follows:

$$
P(S_i|N_i, T_i = 1) = \sum_{k_i=1}^{N_i-1} P(S_i|N_i, k_i, T_i = 1)P(k_i|N_i, T_i = 1) \qquad (11)
$$

where $P(S_i|N_i, k_i, T_i = 1)$ is the probability of observing S_i given N_i, k_i and $T_i = 1$ (son targeting) [10]; $P(k_i|N_i, T_i = 1)$ is the probability of targeting k_i sons given that the maximum number of children is N_i. The summation follows from an application of the law of total probability.

Two things should be immediately noted about (11). First, the summation runs till $(N_i - 1)$ because $k_i = N_i$ is equivalent to not targeting. Second, we only consider cases

[10] In Appendix B, we have sketched a simple method for computing the probabilities $P(S_i|N_i, k_i, T_i = 1)$.

where $N_i \geq 2$; this follows from the intuition that families with a planned maximum number of children below 2 cannot 'target' sons in any meaningful sense. Using (11), therefore, (10) becomes:

$$P(S_i, N_i) = I(n(S_i) = N_i)(\frac{1}{2})^{N_i} P(N_i|T_i = 0)\Phi(-X_i'\beta) + [1 - \Phi(-X_i'\beta)] \sum_{k_i=1}^{N_i-1} P(S_i|N_i, k_i, T_i = 1)P(k_i|N_i, T_i = 1)P(N_i|T_i = 1).$$

Next, we assume that N_i conditional on T_i is distributed as a Poisson random variable with conditional mean given by λ_i. We try to capture two crucial facts with this formulation: one, that N_i conditional on T_i is a count variable; and two that there is interaction between the decision of son targeting and the total number of children planned (the interaction term appears in the expression for the conditional mean, λ_i).

$$P(N_i|T_i) = \frac{\exp(-\lambda_i)[\lambda_i]^{N_i}}{N_i!},$$

where

$$\lambda_i = \exp(Z_i'\gamma + \alpha T_i). \tag{12}$$

Note that in the above expression α captures the effect of 'son targeting' on the total fertility rate and Z_i is a set of covariates which affects family size. Moreover, since T_i is a dichotomous variable, we have

$$P(N_i|T_i = 0) = \frac{\exp(-\exp(Z_i'\gamma))[\exp(Z_i'\gamma)]^{N_i}}{N_i!} \tag{13}$$

and

$$P(N_i|T_i = 1) = \frac{\exp(-\exp(Z_i'\gamma + \alpha))[\exp(Z_i'\gamma + \alpha)]^{N_i}}{N_i!} \tag{14}$$

Using (13), (14) and (11), we can write (10) as:

$$P(S_i, N_i) = I(n(S_i) = N_i)(\frac{1}{2})^{N_i} \frac{\exp(-\exp(Z_i'\gamma))[\exp(Z_i'\gamma)]^{N_i}}{N_i!}\Phi(-X_i'\beta) + \sum_{k_i=1}^{N_i-1} P(S_i|N_i, k_i, T_i = 1)P(k_i|N_i, T_i = 1) \times \frac{\exp(-\exp(Z_i'\gamma + \alpha))[\exp(Z_i'\gamma + \alpha)]^{N_i}}{N_i!}[1 - \Phi(-X_i'\beta)] \tag{15}$$

The log-likelihood for the observed sample, then, becomes:

$$l = \ln(L) = \sum_{i=1}^{n} \ln(P(S_i, N_i)) \tag{16}$$

where n is the number of families in the sample and $P(S_i, N_i)$ is substituted from (15). Maximizing l will give the estimates of the parameters of interest in the model: α (interaction term), β (targeting behaviour), γ (determinants of family size) and the probabilities $P(k_i|N_i, T_i = 1)$. The following interpretations naturally emerge for the parameters in our model: α captures the effect of son targeting on the total fertility rate; β captures the effect of covariates on the probability of targeting sons; γ provides the effect of covariates on the maximum number of children (the total fertility rate) desired by families and $P(k_i|N_i, T_i = 1)$ is the probability of targeting k_i sons given that the family desires a maximum of N_i children.

3.3 Results for India

Results for India are presented in Tables 8, 9 and 10. The following covariates have been used in the analysis: *age* is the age of the respondent in years; *edu* measures the years of formal education; *rur* is a dummy variable for location (it is 0 for urban areas and 1 for rural areas); *work* is a dummy variable to capture whether the respondent participates in the labour market or not; *rich, middle* and *poor* are income/class [11] dummies (we have left out *rich*); *other, scaste, stribe* and *obc* are caste dummies (we have left out *other*); *hindu* and *muslim* are religion dummies capturing membership in the two largest religious communities in India; *jfam* is a dummy for whether the respondent lives in a joint family or not; and, *contra* is a dummy for use of contraceptives (it takes the value of 1 if the respondent uses contraceptives and 0 otherwise).

The first thing to note is that the signs on most variables in Table 8 are along expected lines. Let us first look at the targeting equation (9). The covariates that do not seem to affect the probability of targeting are the following: *edu, middle, obc.* In particular, the estimates suggest that the probability of targeting is not affected by the years of formal education of the respondent (mother). The income dummies show that compared to rich families, the middle-class families target neither more or nor less. Along the same lines, the caste dummies indicate that other backward caste (OBC) families target neither more nor less than upper caste families.

The covariates that affect the probability of targeting are the following: *rur, work, poor, jfam, contra, scaste, stribe, hindu, muslim.* The estimates indicate that families in rural areas are more likely to target sons through male-preferring stopping behaviour than urban families; thus geographical location seems to matter for gender inequality. Next, we see that participation in the labour force by the respondent does affect the probability of targeting

[11] Since there are no income variables in DHS, we had to construct income/class dummies in the following manner: a respondent is designated *rich* if she owns a car/truck and has electricity connection in her house; a respondent is designated *middle* if she has electricity in her house but does not own a car/truck; and a respondent is designated poor if she neither owns a car/truck nor has an electricity connection to her house.

and that too positively; this is rather surprising. Income dummies show that compared to rich families, poorer families have a higher probability of son targeting; this is more or less in agreement with anecdotal evidence. Residence in a joint-family increases the probability of targeting; this is along expected lines because the respondent's (i.e., daughter-in-law's) authority is severely curtailed in a joint family set-up. Next, we see that the use of contraceptives significantly increases the probability of targeting; this is to be expected because without contraceptives, it would be difficult to actually enforce differential stopping (or any other kind of stopping) behaviour. The caste dummies suggests that compared to 'other' (the upper caste), both scheduled caste and schedule tribe families are more probable to practice targeting; the difference, on the other hand, between upper caste and OBC families seem to be insignificant, as already noted. Another counter-intuitive result is the one for religion dummies. Our results suggest that Hindu families are less likely to target sons than non-Hindu families; the same result obtains for Muslim families too. Anecdotal evidence on Hinduism suggests that there is a religious sanction behind son preference; hence, it is surprising that this does not show up in the data.

When we turn to the equation for the determination of family size (12), we notice that all the variables in Table 8 significantly affect the dependent variable; the weakest effect is from the work dummy, which has a negative impact on desired family size. As would be expected, years of formal education reduces the average family size; rural families have larger families (again, along expected lines). The income dummies show some interesting results: compared to rich households, both middle-class and poor households have larger average family sizes, and the effect is stronger for poorer families. The caste dummies show that compared to the upper caste families, SC, ST and OBC families have larger average family sizes. Turning to the religion dummies, we see that both hindu and muslim families have larger family size than non-hindu and non-muslim families, though the effect is much stronger for the latter group.

Along expected lines, the interaction term comes out as positive and highly significant; families which are more likely to be targeters will tend to have larger families. This is to be expected because son preference will induce couples to continue childbearing in their attempt to attain the target for boys. If we look at the targeting probabilities in Table 9, we notice a simple pattern: conditional on targeting, couples are more likely to target higher than lower number of sons. So, for instance, families which have a ceiling at five children are more likely to target four boys than three than two than one. The same is true for families having a ceiling at three and six. But families with a desired family size of four display a slightly different behaviour; for these families, the probability of targeting two sons is marginally higher than the probability of targeting three sons.

To test for the joint significance of all the variables in the targeting equation (9), we performed an LR test. The results for the restricted model is given in Table 10. The LR test statistic, a χ^2 random variable with 12 degrees of freedom, has a value of 877.94; thus, we can easily reject the null hypothesis that all the variables in the targeting equation are insignificant.

4 Conclusion

In this paper we have developed a simple model that can account for the generation and perpetuation of gender inequality in societies marked by son preference. Our simple analysis shows that even in the absence of intra-household biases in resource allocation against the girl child there is reason to expect a substantially disadvantageous position for her. There are two distinct mechanisms - one static and the other dynamic - by which the existence of son preference impacts on gender inequality. If we look at a developing economy at any point in time, the distribution of children into families is such that the average number of siblings for female children are greater than that for their male counterparts. If we look at the same society over time, we find that boys are born as relatively younger children in families. We have argued that both these facts have important implications for gender inequality. Our empirical analysis suggests that both these effects are strong in many countries of South Asia, South-East Asia and North Africa. We find an absence of these effects in the countries of Sub-Saharan Africa.

One implication of the simple model of this paper is that DSB does not affect the population sex ratio at birth (SRB). Since the SRB has been declining in India over the past few decades (Agnihotri, 2001), this seems to suggest that sex-selective abortion (foeticide) and female infanticide are gaining ground in developing countries like India. One recent paper (Oster, 2005) has tried to suggest a biological explanation for this phenomenon: the author suggests that the prevalence of hepatitis-B might be able to explain the declining SRB. We would like to conduct further research on this area as an extension of the analysis developed in this paper.

One policy implication of this analysis is that in societies where son preference is widespread, one way to counter the negative effects of DSB on female children is for the State to actively participate in the provision of education and health. Not only should the government provide free primary and secondary education, it should also make it compulsory. Additionally, the government should provide free health-care for all children up to the age of 25 years. These policy initiatives might be able to counter the effects of the "sibling effect" and the "birth order effect" which rests on expenditure for children's development being channeled through the family.

5 Appendix A

We will need the following two lemmas for proving Proposition 1.

Lemma 1 *Let* $P(N,k) = \sum_{i=k}^{N} \binom{i-1}{k-1} (\frac{1}{2})^i + \sum_{i=0}^{k-1} \binom{N}{i} (\frac{1}{2})^N$*; then* $P(n,k+1) = P(n,k), \forall k \leq N$

Proof: Rewrite $P(N,k) = (\frac{1}{2})^k \sum_{i=0}^{N-k} \binom{k-1}{k+i-1} (\frac{1}{2})^i + (\frac{1}{2})^N \sum_{i=0}^{k-1} \binom{N}{i}$. The expression for N and $(k+1)$ is

$$P(N,k+1) = (\frac{1}{2})^{k+1} \sum_{i=0}^{N-k-1} \binom{k+i}{k} (\frac{1}{2})^i + (\frac{1}{2})^N \sum_{i=0}^{k} \binom{N}{i}$$

This can be re-written using the following equality:

$$\binom{k+i}{k} = \binom{k+i-1}{k} + \binom{k+i-1}{k-1}. \tag{17}$$

Thus

$$\begin{aligned}
P(n,k+1) &= (\frac{1}{2})^{k+1}[\sum_{i=0}^{N-k-1} \binom{k+i-1}{k}(\frac{1}{2})^i + \sum_{i=0}^{N-k-1} \binom{k+i-1}{k-1}(\frac{1}{2})^i] + (\frac{1}{2})^N \sum_{i=0}^{k}\binom{N}{i} \\
&= (\frac{1}{2})^{k+2} \sum_{i=0}^{N-k-1} \binom{k+i-1}{k}(\frac{1}{2})^{i-1} + (\frac{1}{2})^N \sum_{i=0}^{k}\binom{N}{i} + (\frac{1}{2})^{k+1} \sum_{i=0}^{N-k-1} \binom{k+i-1}{k-1}(\frac{1}{2})^i \\
&= A + B
\end{aligned}$$

where

$$\begin{aligned}
A &= (\frac{1}{2})^{k+2} \sum_{i=0}^{N-k-1} \binom{k+i-1}{k}(\frac{1}{2})^{i-1} + (\frac{1}{2})^N \sum_{i=0}^{k}\binom{N}{i} \\
&= (\frac{1}{2})^{k+2}[\sum_{i=0}^{N-k-1} \binom{k+i-1}{k}(\frac{1}{2})^{i-1} + (\frac{1}{2})^{N-k-1}\binom{N-1}{k}] - (\frac{1}{2})^{N+1}\binom{N-1}{k} + (\frac{1}{2})^N \sum_{i=0}^{k}\binom{N}{i} \\
&= (\frac{1}{2})^{k+2} \sum_{j=0}^{N-k-1} \binom{k+j}{k}(\frac{1}{2})^j + (\frac{1}{2})^N \sum_{i=0}^{k}\binom{N}{i} - (\frac{1}{2})^{N+1}\binom{N-1}{k} \\
&= (\frac{1}{2})\{(\frac{1}{2})^{k+1} \sum_{i=0}^{N-k-1} \binom{k+i}{k}(\frac{1}{2})^i + (\frac{1}{2})^N \sum_{i=0}^{k}\binom{N}{i}\} + (\frac{1}{2})^{N+1} \sum_{i=0}^{k}\binom{N}{i} - (\frac{1}{2})^{N+1}\binom{N-1}{k} \\
&= (\frac{1}{2})P(n,k+1) + (\frac{1}{2})^{N+1} \sum_{i=0}^{k}\binom{N}{i} - (\frac{1}{2})^{N+1}\binom{N-1}{k}
\end{aligned}$$

and

$$\begin{aligned}
B &= (\frac{1}{2})^{k+1} \sum_{i=0}^{N-k-1} \binom{k+i-1}{k-1}(\frac{1}{2})^i \\
&= (\frac{1}{2})^{k+1}[\sum_{i=0}^{N-k-1} \binom{k+i-1}{k-1}(\frac{1}{2})^i + (\frac{1}{2})^{N-k}\binom{N-1}{k-1}] - (\frac{1}{2})^{N-k}(\frac{1}{2})^{k+1}\binom{N-1}{k-1} \\
&= (\frac{1}{2})^{k+1} \sum_{i=0}^{N-k}\binom{k+i-1}{k-1}(\frac{1}{2})^i - (\frac{1}{2})^{N+1}\binom{N-1}{k-1}
\end{aligned}$$

Adding A and B, we get:

$$\begin{aligned}
P(n,k+1) &= (\frac{1}{2})P(n,k+1) - (\frac{1}{2})^{N+1}\binom{N-1}{k} \\
&\quad + (\frac{1}{2})^{k+1}\sum_{i=0}^{N-k}\binom{k+i-1}{k-1}(\frac{1}{2})^i + (\frac{1}{2})^{N+1}\sum_{i=0}^{k}\binom{N}{i} \\
&= (\frac{1}{2})P(n,k+1) - (\frac{1}{2})^{N+1}[\binom{N-1}{k} + \binom{N-1}{k-1}] \\
&\quad + (\frac{1}{2})^{k+1}\sum_{i=0}^{N-k}\binom{k+i-1}{k-1}(\frac{1}{2})^i + (\frac{1}{2})^{N+1}\sum_{i=0}^{k}\binom{N}{i} \\
&= (\frac{1}{2})P(n,k+1) - (\frac{1}{2})^{N+1}[\binom{N-1}{k} + \binom{N-1}{k-1}] \\
&\quad + (\frac{1}{2})\{(\frac{1}{2})^{k}\sum_{i=0}^{N-k}\binom{k+i-1}{k-1}(\frac{1}{2})^i + (\frac{1}{2})^{N}\sum_{i=0}^{k}\binom{N}{i}\}
\end{aligned}$$

Note that in the second line, we have used (17); rearranging the above, we get:

$$\begin{aligned}
(\frac{1}{2})P(n,k+1) &= (\frac{1}{2})\{(\frac{1}{2})^{k}\sum_{i=0}^{N-k}\binom{k+i-1}{k-1}(\frac{1}{2})^i \\
&\quad + (\frac{1}{2})^{N}\sum_{i=0}^{k-1}\binom{N}{i}\} + (\frac{1}{2})^{N+1}\binom{N}{k} - (\frac{1}{2})^{N+1}[\binom{N-1}{k} + \binom{N-1}{k-1}] \\
&= (\frac{1}{2})P(n,k) + (\frac{1}{2})^{N+1}[\binom{N}{k} - \binom{N-1}{k} - \binom{N-1}{k-1}]
\end{aligned}$$

Hence, $P(n,k+1) = P(n,k)$, where the second term is zero by (17). $\square$

Lemma 2 $P(n,1) = 1, \forall n$

Proof: Since $P(N,k) = (\frac{1}{2})^k \sum_{i=0}^{N-k}\binom{k-1}{k+i-1}(\frac{1}{2})^i + (\frac{1}{2})^N \sum_{i=0}^{k-1}\binom{N}{i}$, we have,

$$\begin{aligned}
P(n,1) &= (\frac{1}{2})\sum_{i=0}^{N-1}\binom{i}{0}(\frac{1}{2})^i + (\frac{1}{2})^N \\
&= (\frac{1}{2})[1 + (\frac{1}{2}) + \ldots + (\frac{1}{2})^{N-1}] + (\frac{1}{2})^N \\
&= 1
\end{aligned}$$

This completes the proof. $\square$

Proof of Proposition 1 The proof now follows because using Lemma 1 and Lemma 2, we have: $P(n,k) = P(n,k-1) = \ldots = P(n,1) = 1$. $\square$

Proof of Theorem 1 Recall from (3) that

$$C(k,N) = \sum_{i=k}^{N} \binom{i-1}{k-1} (\frac{1}{2})^i + (N-1) \sum_{i=0}^{k-1} \binom{N}{i} (\frac{1}{2})^N. \tag{18}$$

Notice that $C(k,N)$ is a weighted average of the following terms:$(k-1), k, (k+1), \ldots, (N-2), (N-1)$; where the weights sum to unity by Proposition 1. Hence

$$k - 1 \leq C(k) \leq N - 1 \tag{19}$$

Now, re-writing (1), we get:

$$\begin{aligned} M_s &= \frac{C(k) - (\frac{1}{2})^N (N-1)}{1 - (\frac{1}{2})^N} \\ &= \frac{(2^N - 1)C(k) + C(k) - N + 1}{2^N - 1} \\ &= C(k) + \frac{C(k) - N + 1}{2^N - 1} \end{aligned}$$

Similarly, rewriting (2), we get:

$$\begin{aligned} F_s &= \frac{C(k) - (\frac{1}{2})^k (k-1)}{1 - (\frac{1}{2})^k} \\ &= C(k) + \frac{C(k) - k + 1}{2^k - 1} \end{aligned}$$

We will have proved Theorem 1 once we prove the following

Lemma 3 $M_s \leq F_s, \forall k < N$

Proof: Taking the difference between M_s and F_s, we have

$$\begin{aligned} M_s - F_s &= \frac{C(k) - N + 1}{2^N - 1} - \frac{C(k) - k + 1}{2^k - 1} \\ &= \frac{C(k)(2^k - 1) - N(2^k - 1) + (2^k - 1) - C(k)(2^N - 1) + k(2^N - 1) - (2^N - 1)}{(2^N - 1)(2^k - 1)} \end{aligned} \tag{20}$$

Since the denominator in the above expression is always positive (because in any meaningful situation both k and N are greater than 0), we only need to verify that the numerator is non-positive. Denoting the numerator by NUM, we have:

$$\begin{aligned} NUM &= C(k)(2^k - 2^N) - N2^k + k2^N + N - k + 2^k - 2^N \\ &\leq (N-1)(2^k - 2^N) - N2^k + k2^N + N - k + 2^k - 2^N, (\text{since} \quad C(k) \leq N - 1) \\ &\leq (-N + k)2^N + (N - k) \\ &\leq (N - k)(1 - 2^N) \\ &\leq 0, \forall N > 0. \end{aligned}$$

This completes the proof. $\square$

We require two simple lemmas for the proof of Theorem 2.

Lemma 4 $(1+\frac{1}{k}) > (1-\frac{1}{2^N})(1-\frac{1}{2^k}), \quad \forall 1 < k < N.$

Proof This is obvious since

$$\frac{1}{k} > \frac{1}{2^N}(\frac{1}{2^k}-1) - \frac{1}{2^k}, \quad \forall 1 < k \leq N.$$

Lemma 5 $\frac{k+1}{2^{k+1}}(1-2^{-k}) > \frac{N+1}{2^{N+1}}(1-2^{-N}), \quad \forall 1 < k < N.$

Proof It suffices to prove the following:

$$\frac{k+1}{2^{k+1}}(1-2^{-k}) > \frac{k+2}{2^{k+2}}(1-2^{-k-1})$$

for all $1 < k < N$; which is equivalent to proving

$$2(k+1)(1-\frac{1}{2^k}) > (k+2)(1-\frac{1}{2^{k+1}})$$

$$\text{i.e.,} \quad 2(k+1) - \frac{k+1}{2^{k-1}} > (k+2) - \frac{k+2}{2^{k+1}}$$

$$\text{i.e.,} \quad k2^{k-1} > (1/4)[4(k+1)-(k+2)]$$

$$\text{i.e.,} \quad 4k.2^{k-1} > 3k-2$$

which is true because $4k.2^{k-1} > 4k > 3k > 3k-2$, for $k > 1$. $\square$

Proof of Theorem 2: We prove the theorem for two cases: $k=1$ and $k>1$.
Case 1: In this case, when $k=1$, we have

$$\bar{M}_{BO} = \frac{1}{1-(1/2)^N}\{\frac{1}{2} + 2\sum_{i=1}^{N}(i/2)(\frac{1}{2})^i)\}$$

and

$$\bar{F}_{BO} = \frac{1}{1-(1/2)}\{\frac{N+1}{2^{N+1}} + \sum_{i=1}^{N}(i/2)(\frac{1}{2})^i\}.$$

So, in this case, it suffices to show that

$$\frac{1}{1-(1/2)^N} > \frac{N+1}{2^{N+1}}$$

which is true since

$$2^N 2^{N+1} > (N+1)2^N > (N+1)2^N - (N+1)$$

Case 2: For this case, when $k > 1$, comparing the expressions in (7) and (8) and using the results in Lemma 4 and Lemma 5, we get the desired result. □

6 Appendix B

In this appendix, we sketch the method that we have used to compute the conditional probabilities, $P(S_i|N_i, k_i, T_i = 1)$, that have been subsequently used in the maximum likelihood estimation [12]. The logic of our method is straightforward. For every family, we are given a completed birth sequence (S_i) and the planned maximum number of children(N_i). For such a family, we must compute the following $N_i - 1$ conditional probabilities: $P(S_i|N_i, k_i = 1, T_i = 1), P(S_i|N_i, k_i = 2, T_i = 1), \ldots, P(S_i|N_i, k_i = N_i - 1, T_i = 1)$. We need to compute all these probabilities because we do not observe the desired target for sons.

Since, *a priori*, we do not know the desired target (for sons) for family i, we need to allow for all the feasible possibilities. So, when a family states that the maximum number of children that it can produce is N_i, we need to allow for the possibilities that the family targets 1 son, 2 sons, ..., $N_i - 1$ sons. Of course the actual birth sequence might assign zero probability to some of this possibilities; but we cannot rule any of thes out *a priori*.

Now, to compute something like $P(S_i|N_i, k_i, T_i = 1)$, we merely need to observe whether the family has *any* child after the k_i^{th} son. If there is a child after the k_i^{th} son, then we assign zero probability to $P(S_i|N_i, k_i, T_i = 1)$; else we assign it a probability of $(1/2)^n$, where n is the number of children in the sequence S_i.

An example might clarify matters. Suppose a family reports that the maximum number of children it can have is 4 and the birth sequence (starting with the fisrt born child) for the family is observed to be GGBG (where G stands for a girl and B stands for a boy). For such a family, we need to compute the following probabilities: $P(GGBG|N_i = 4, k_i = 1, T_i = 1)$, $P(GGBG|N_i = 4, k_i = 2, T_i = 1)$, and $P(GGBG|N_i = 4, k_i = 3, T_i = 1)$. Since there is a child after the first boy, this family could not possibly be targeting one son; hence $P(GGBG|N_i = 4, k_i = 1, T_i = 1) = 0$. But the family could conceivably be targeting two or even three sons; these possibilities are not ruled out by the observed birth sequence. Hence $P(GGBG|N_i = 4, k_i = 2, T_i = 1) = (1/16)$, and similarly $P(GGBG|N_i = 4, k_i = 3, T_i = 1) = (1/16)$.

To clarify matters further, take another example. Suppose the family in question reports a maximum desired family size (number of children) of 4 and we observe the following completed birth sequence for the same family: BGB. Since there is a child after the first boy, this

[12] STATA code for these and other computations, including the maximum likelihood estimation, are available from the authors upon request.

family could not possibly be targeting one son; hence $P(BGB|N_i = 4, k_i = 1, T_i = 1) = 0$. But the family could be targeting two sons. Hence $P(BGB|N_i = 4, k_i = 2, T_i = 1) = (1/8)$. And since the family stops at three children (with two sons), it cannot be targeting three sons; to target three sons, the family should have gone for another child and not stopped at the third child (and second son). Hence, $P(BGB|N_i = 4, k_i = 3, T_i = 1) = 0$.

References

Arnold F, Choe MK, Roy TK. Son Preference, the Family-building Process and Child Mortality in India. Population Studies 1998; 52; 301-315.

Agnihotri SB. 2001. Rising Sons and Setting Daughters: Provisional Results of the 2001 Census. In: Mazumdar V, Krishnaji N (Eds), Enduring Conundrum: India's Sex Ratio: Essays in Honour of Asok Mitra. Centre for Womens Development Studies and Rainbow Publishers: New Delhi; 2001.

Bardhan, P. On Life and Death Questions. Economic and Political Weekly (Special Number) 1974: 1293-1304.

Bardhan P. Little Girls and Deaths in India. Economic and Political Weekly 1982: 1448-50.

Becker G. 1960. Demographic and Economic Change in Developing Countries. Princeton University Press: Princeton, NJ; 1960.

Becker G, Lewis HG. On the Interaction between the Quantity and Quality of Children. The Journal of Political Economy 1973; 81(2), part 2: S279-S288.

Clark S. Son Preference and Sex Composition of Children: Evidence from India. Office of Population Studies (Working Paper) 2001; Princeton University.

Deaton A. 1989. Looking for Boy-Girl Discrimination in Household Expenditure Data. World Bank Economic Review 1989; 3; 1-15.

Filmer D, King EM, Pritchett L. Gender Disparity in South Asia: Comparisons between and within countries. Working paper, World Bank 1998.

Jensen R T. Equal Treatment, Unequal Outcomes? Generating Sex Inequality through Fertility Behaviour. Mimeo, JFK School of Government, Harvard University 2002; Cambridge.

Larsen U, Chung W, Das Gupta M. Fertility and Son Preference in Korea. Population

Studies 1998; 52: 317-325.

Oster E. 2005. Hepatitis B and the Case of Missing Women. Mimeo, Centre for International Development, Harvard University 2005; Cambridge.

Ray D. Development Economics. Princeton University Press: NJ; 1998.

Seidl C. The Desire for a Son is the Father of Many Daughters: A Sex Ratio Paradox. Journal of Population Economics 1995; 8(2): 185-203.

Sen A. More than 100 million Women are Missing. New York Review of Books 1990; December 20.

Sen A, Sengupta S. Malnutrition of Rural Indian Children and the Sex Bias. Economic and Political Weekly 1983; 855-64.

Yamaguchi K. A Formal Theory for Male-Preferring Stopping Rules of Childbearing: Sex Differences in Birth Order and in the Number of Siblings. Demography 1989; 26(3); 451-465.

Table 3: Datasets for Asia and North Africa

Country	Survey	Year	Size (women)
Asia			
Bangladesh	Demographic and Health Survey	1999-2000	10544
India	National Family Health Survey	1998-1999	90303
Indonesia	Demographic and Health Survey	2002-2003	29483
Nepal	Demographic and Health Survey	2001	8726
Pakistan	Demographic and Health Survey	1990-1991	6611
Phillipines	Demographic and Health Survey	2003	13633
Sri Lanka	Demographic and Health Survey	1987	5865
Thailand	Demographic and Health Survey	1987	6775
North Africa			
Armenia	Demographic and Health Survey	2000	6430
Egypt	Demographic and Health Survey	2000	15573
Morocco	Demographic and Health Survey	2003	16798

Table 4: Datasets for Sub-Saharan Africa

Country	Survey	Year	Size
Benin	Enquete Demographique et de Sante	2001	6219
Burkina Faso	Demographic and Health Survey	2003	12477
Burundi	Enquete Demographique et de Sante	1987	3970
Cameroon	Enquete Demographique et de Sante	2004	10656
Central African Republic	Demographic and Health Survey	1994-1995	5884
Chad	Demographic and Health Survey	1996-1997	7454
Comoros	Enquete Demographique et de Sante	1996	3050
Cote d'Ivoire	Enquete Demographique et de Sante	1998	3040
Ethiopia	Demographic and Health Survey	2000	15367
Gabon	Enquete Demographique et de Sante	2000-2001	6183
Ghana	Demographic and Health Survey	1998-1999	4843
Guinea	Demographic and Health Survey	1999	6753
Kenya	Demographic and Health Survey	2003	8195
Madagascar	Enquete Demographique et de Sante	1997	7060
Malawi	Demographic and Health Survey	2000	13220
Mali	Demographic and Health Survey	2000	12849
Mozambique	Demographic and Health Survey	2003	12418
Namibia	Demographic and Health Survey	2000	6755
Niger	Enquete Demographique et de Sante	1998	7577
Nigeria	Demographic and Health Survey	1999	9810
Rwanda	Demographic and Health Survey	2000	10421
South Africa	Demographic and Health Survey	1998	11735
Sudan	Demographic and Health Survey	1989-1990	5860
Tanzania	Demographic and Health Survey	1999	4029
Uganda	Demographic and Health Survey	2000-2001	7246
Zambia	Demographic and Health Survey	2001	7658
Zimbabwe	Demographic and Health Survey	1999	5907

Table 5: Sibling and Birth Order Effects for Asia and North Africa

	Sibling Effect*		Birth Order Effect**	
	Effect	t-stat	Effect	t-stat
Asia				
Bangladesh	0.07	2.62	0.04	2.36
India	0.14	17.47	0.06	9.19
Indonesia	0.0	0.08	0.03	2.81
Nepal	0.18	6.57	0.03	1.61
Pakistan	0.06	1.33	-0.05	-1.46
Phillipines	0.03	1.15	-0.01	-0.57
Sri Lanka	0.04	1.33	0.02	0.65
Thailand	0.03	0.82	0.0	0.08
North Africa				
Armenia	0.09	4.14	0.1	5.18
Egypt	0.06	2.92	0.04	2.86
Morocco	0.04	1.43	-0.02	-1.07

*Sibling effect is the difference in the average number of siblings per child between a girl and a boy.

**Birth order effect is the difference in the average within-family birth order of boys and girls.

Table 6: Sibling and Birth Order Effects for Sub-Saharan Africa

	Sibling Effect*		**Birth Order Effect****	
	Effect	t-stat	Effect	t-stat
Benin	0.03	0.44	0.0	0.04
Burkina Faso	0.03	0.75	-0.02	-0.53
Burundi	0.06	0.74	-0.05	-0.72
Cameroon	0.01	0.09	-0.04	-0.92
Central African Republic	-0.01	-0.13	-0.01	-0.23
Chad	-0.05	-0.7	0.06	1.18
Comoros	-0.13	-1.23	0.02	0.19
Cote d'Ivoire	0.08	0.76	-0.09	-1.15
Ethiopia	0.01	0.28	-0.02	-0.73
Gabon	-0.03	-0.36	0.0	0.06
Ghana	0.0	-0.03	-0.08	-1.79
Guinea	-0.01	-0.26	-0.01	-0.34
Kenya	0.02	0.5	-0.02	-0.56
Madagascar	-0.02	-0.41	0.01	0.22
Malawi	-0.04	-1.13	-0.03	-0.97
Mali	0.02	0.37	0.02	0.88
Mozambique	-0.02	-0.56	0.02	0.73
Namibia	-0.04	-0.82	0.02	0.71
Niger	0.03	0.41	0.0	-0.06
Nigeria	0.11	1.78	-0.08	-1.77
Rwanda	0.01	0.24	0.01	0.38
South Africa	0.03	0.88	-0.03	-1.45
Sudan	-0.02	-0.31	-0.02	-0.43
Tanzania	0.03	0.38	0.01	0.13
Uganda	0.01	0.12	0.01	0.40
Zambia	-0.01	-0.09	0.0	-0.09
Zimbabwe	0.0	0.04	0.02	0.55

*Sibling effect is the difference in the average number of siblings per child between a girl and a boy.

**Birth order effect is the difference in the average within-family birth order of boys and girls.

Table 7: Sibling and Birth Order Effects for Indian States

	Sibling Effect*		**Birth Order Effect****	
	Effect	t-stat	Effect	t-stat
Andhra Pradesh	0.04	1.16	0.07	2.79
Arunachal Pradesh	0.10	1.17	-0.08	-1.29
Assam	0.11	2.35	-0.04	-1.24
Bihar	0.17	5.29	-0.03	-1.39
Delhi	0.14	3.31	0.06	1.99
Goa	0.12	1.77	0.04	0.87
Gujrat	0.20	5.53	0.12	4.42
Haryana	0.28	6.74	0.07	2.37
Himachal Pradesh	0.19	5.35	0.11	4.05
Jammu and Kashmir	0.12	2.75	0.11	3.29
Karnataka	0.08	2.42	0.07	2.54
Kerala	0.02	0.54	0.01	0.19
Madhya Pradesh	0.19	6.36	0.03	1.30
Mahrashtra	0.15	5.41	0.11	5.25
Manipur	0.15	2.05	0.12	2.23
Meghalaya	0.07	0.65	-0.09	-1.04
Mizoram	0.05	0.62	0.10	1.64
Nagaland	0.09	0.83	-0.06	-0.81
Orissa	0.16	4.7	0.02	0.78
Punjab	0.18	5.1	0.21	7.48
Rajashthan	0.19	6.19	0.09	3.9
Sikkim	0.18	2.35	0.00	0.09
Tamil Nadu	0.06	2.11	0.06	2.63
Tripura	0.15	2.13	0.04	0.72
Uttar Pradesh	0.18	6.59	0.01	0.53
West Bengal	0.12	3.12	0.01	0.57

*Sibling effect is the difference in the average number of siblings per child between a girl and a boy.

**Birth order effect is the difference in the average within-family birth order of boys and girls.

Table 8: Estimation* results for India (unrestricted model)

Variable	**Coefficient**	**(Std. Err.)**
Average family size: (γ)		
age	0.02	(0.000)
edu	-0.016	(0.001)
rur	0.053	(0.006)
work	-0.011	(0.005)
middle	0.359	(0.013)
poor	0.411	(0.014)
scaste	0.084	(0.007)
stribe	0.153	(0.009)
obc	0.034	(0.006)
hindu	0.064	(0.009)
muslim	0.246	(0.011)
Interaction term: (α)		
Intercept	0.055	(0.008)
Targeting behaviour: (β)		
edu	0.000	(0.002)
rur	0.059	(0.025)
work	0.074	(0.022)
middle	0.088	(0.07)
poor	0.214	(0.074)
jfam	0.093	(0.028)
contra	0.528	(0.023)
scaste	0.075	(0.031)
stribe	0.302	(0.037)
obc	0.016	(0.026)
hindu	-0.201	(0.035)
muslim	-0.129	(0.049)
Intercept	-1.298	(0.080)

*For a definition of the empirical model, refer to (15) in the main text.

Table 9: Estimated targeting probabilities* for India (unrestricted model)

	targeting probability
P(k=1|N=3)	0.28
P(k=2|N=3)	0.72
P(k=1|N=4)	0.17
P(k=2|N=4)	0.43
P(k=3|N=4)	0.40
P(k=1|N=5)	0.07
P(k=2|N=5)	0.20
P(k=3|N=5)	0.32
P(k=4|N=5)	0.41
P(k=1|N=6)	0.05
P(k=2|N=6)	0.13
P(k=3|N=6)	0.23
P(k=4|N=6)	0.25
P(k=5|N=6)	0.33

*For a definition of these probabilities see Appendix B. All the probabilities reported in this table are statistically siginificant at the 5 % level.

Table 10: Estimation** results for India (restricted model)

Variable	**Coefficient**	**(Std. Err.)**
Average family size: (γ)		
age	0.020	(0.000)
edu	-0.017	(0.001)
rur	0.054	(0.006)
work	-0.011	(0.005)
middle	0.361	(0.013)
poor	0.414	(0.014)
scaste	0.085	(0.007)
stribe	0.156	(0.009)
obc	0.034	(0.006)
hindu	0.063	(0.009)
muslim	0.246	(0.011)
Interaction term: (α)		
Intercept	0.044	(0.007)
Targeting behaviour: (β)		
Intercept	-1.069	(0.011)

**For a definition of the empirical model, refer to (15) in the main text.