SON TARGETING FERTILITY BEHAVIOR: SOME CONSEQUENCES AND DETERMINANTS*

DEEPANKAR BASU AND ROBERT DE JONG

This article draws out some implications of son targeting fertility behavior and studies its determinants. We demonstrate that such behavior has two notable implications at the aggregate level: (a) girls have a larger number of siblings (sibling effect), and (b) girls are born at relatively earlier parities within families (birth-order effect). Empirically testing for these effects, we find that both are present in many countries in South Asia, Southeast Asia, and North Africa but are absent in the countries of sub-Saharan Africa. Using maximum likelihood estimation, we study the effect of covariates on son targeting fertility behavior in India, a country that displays significant sibling and birth-order effects. We find that income and geographic location of families significantly affect son targeting behavior.

Many developing countries in East, South, and Southeast Asia, and North Africa are characterized by a strong son preference—that is, a strong preference for male as opposed to female offspring (Arnold, Choe, and Roy 1998; Clark 2000; Jensen 2002). This strong preference is reflected in son targeting fertility behavior, also referred to in the literature as differential stopping behavior (DSB) or male-preferring stopping rules (see, e.g., Clark 2000). The main idea behind such stopping rules is that the sex composition of current children determines the subsequent fertility behavior of families; for evidence on DSB see Arnold et al. (1998) and Larsen, Chung, and Das Gupta (1998). In our analysis, we concretize DSB 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 determine to be feasible (given their resource constraints). The theoretical results in the article are derived on the basis of the assumption that every couple in the population follows this behavior; in the empirical model, we relax this assumption by allowing each couple to follow this behavior with some probability lying between 0 and 1.

In this article, we highlight two important implications of son targeting fertility behavior. First, we demonstrate that if a population practices son targeting fertility behavior, girls will be born into relatively larger families; we call this the "sibling effect." Second, we show that such behavior will also imply that girls are born relatively earlier within families; we call this the "birth-order effect." Both effects might have important implications for gender inequality even in the absence of intrahousehold allocation biases, an issue we wish to study in the future.

Fertility stopping rules have been previously studied in the demography literature. For instance, Yamaguchi (1989)[1] defined the following stopping rule: couples continue childbearing until they reach their desired number, k, of sons. Thus, in contrast to the stopping rule we use, Yamaguchi's rule set no limit on the maximum number of children that couples could have in their attempt to attain the target number of boys. Our framework is more general; Yamaguchi's model can be considered a special case of our model if we let N go to infinity. The analysis becomes substantially simplified when N can increase without bounds

1. Similar ideas can also be found in Ray (1998), whose proposed male preferring stopping rule is very similar to Yamaguchi's (1989).

because the number of children in a family follows a standard probability distribution (the negative binomial distribution); when N is a finite integer, there is no longer 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 (DHS 1992), for instance, about 94% of the households have five or fewer children.

Jensen (2002) arrived at results similar to what we have called the sibling effect, though he used a different stopping rule. In his model, couples want n children and b boys; but if they reach n children with fewer 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); Jenson's model 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, whereas we discuss both the sibling effect and the birth-order effect, Jensen (2002) limited 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 maximum likelihood estimation.

To focus attention on the issue of birth order, we differentiate between mean *absolute* and mean *relative* birth order. We compute the mean absolute birth order of girls (boys) by averaging the position of the female (male) child within the sequence of births in her (his) family, where averaging is done over all children in the population. To compute mean relative birth order of girls (boys) in the population, we first calculate the average position of female (male) children *within each family* and then average over families.

An example might 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 will have a girl and a boy (in that order), and another fourth will have 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 firstborn children, and one-third are second-born, and so the mean absolute birth order is 4/3 ($= 1 \times 2/3 + 2 \times 1/3$) for both boys and girls.

Let us now compute the mean relative birth orders. To proceed, we define an average within-family birth order (AWFBO) score for boys and girls respectively in each family and then average across families. Notice that boys have an AWFBO score of 1 in families that have only one child (half of the families) and an AWFBO score of 2 in families that have a girl as the first child and a boy as the second child (one-fourth of the families). Girls, on the other hand, have an AWFBO score of 1 for families that have a girl as the first child and a boy as the second child (one-fourth of the families) and an AWFBO score of 3/2 ($= (1 + 2) / 2$) for families with two girls. Hence, the mean relative birth order for boys is 4/3 ($= 1 \times 2/3 + 2 \times 1/3$) because we average across families. Similarly, the mean relative birth order for girls is 5/4 ($= 1 \times 1/2 + 3/2 \times 1/2$). Thus, boys have a higher relative birth order than girls.

Yamaguchi's (1989) discussion of birth order refers to *relative* birth order. He concluded 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 1989:459). This result, which is at odds 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 do in our model, the result no longer holds; the mean relative birth order for boys turns out to be greater than that for girls, as the above example shows and as we demonstrate later (see Observation 2). On the other hand, Jensen's (2002:6) conclusion that "fertility-related characteristics such as birth order...do *not* differ between boys and girls…" is a statement about mean *absolute* birth order. Thus, even though both Yamaguchi (1989) and Jensen (2002) seem to have arrived at the same result, they were really referring to different measures of birth order.

To summarize, the contribution of this article is twofold. First, the male-preferring stopping rule that we analyze is both more general and more realistic than that used by

Yamaguchi (1989). Second, we highlight not only the sibling effect but also the birth order effect,[2] an issue that seems to have been neglected so far in the literature (e.g., Jensen 2002). Our empirical results show that both sibling and birth-order effects are present in a host of countries in East Asia, South Asia, Southeast Asia, and North Africa. These effects are absent in countries of sub-Saharan Africa. When we compute the same effects for the Indian states, we find their strong presence in the states of North, West, and Central India; the effects are absent in Kerala and in several states in Northeastern India. This is more or less in line with anecdotal evidence on patriarchal tendencies in the different geographic regions of India. Our maximum likelihood estimation for India reveals that geographic location of the family (urban versus rural) has a large and significant impact on both desired fertility and son targeting: family size and the probability of son targeting behavior is higher in rural areas after we control for age, formal education, and other observable characteristics.

The rest of the article is organized as follows. The next section presents the main results; the following section 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 conjectures on possible extensions of the current analysis. Proofs of the main theoretical claims, Observations 1 and 2 below, are available upon request from the authors; the appendix provides details about computations that are relevant for calculating the log-likelihood function.

MAIN RESULTS

Many developing economies in South, East, and Southeast Asia, and North Africa display strong son preference (Clark 2000). This preference is reflected in fertility behavior in the form of DSB. In our analysis, we concretize DSB as follows:

> **Assumption 1.** *Couples continue childbearing until they attain a desired target number of sons, k, or hit a ceiling for the maximum number of children, N, with $k \leq N$.*

Let us denote the probability of a male birth as q, with $0 < q < 1$. Table 1 gives the various possible completed family structures (in terms of the number of children and their gender) that would emerge in a population practicing DSB and the probabilities associated with those family structures.

Note that DSB does not affect the sex ratio at birth (SRB), defined as male births per 100 female births. This fact has an important implication for countries in Asia and Africa that have witnessed increasing population sex ratios (males per 100 females) and SRBs in recent decades. Because the presence of DSB by itself does not affect the SRB, an increasing trend in SRBs can result from only one of the following: biological factors, as Oster (2005) suggested; the increasing prevalence of sex-selective abortion of female fetuses, as has been recently reported in the press (see, e.g., the coverage by BBC at http://news.bbc.co.uk/2/hi/south_asia/4592890.stm); or economic development and reductions in infant mortality rates, as suggested by Bhaskar and Gupta (2007). Ascertaining which of these alternative explanations can better explain the increasing SRB remains an important research question.

The Sibling Effect

The first implication of DSB that we wish to highlight is that, on average, girls are born into larger families. We call this effect of DSB the "sibling effect," and we offer two quantitative measures of it.

Averaging over families. The first measure of the sibling effect is computed as the difference between (1) the expected number of siblings for children in families with at least

2. When we mention birth order without any qualification in this article, we mean the relative birth order.

Table 1. **Family Structures and Number of Siblings**

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

one male child and (2) the expected number of siblings for children in families with at least one female child. Because all children within a family share the same number of siblings, the expected number of siblings for children in families with at least one male child gives an estimate of the average number of siblings that boys have in the population; similarly, the expected number of siblings for children in families with at least one female child gives an estimate of the average number of siblings that girls have in the population. The difference between the two, therefore, gives an estimate of what we have called the sibling effect.

Using Table 1, it can be seen that the expected number of siblings for families with at least one boy, $\overline{M_s^1}$, is

$$\overline{M_s^1} = \frac{C(k,N) - (N-1)(1-q)^N}{1 - (1-q)^N};\tag{1}$$

and the expected number of siblings for families with at least one girl, $\overline{F_s^1}$, is

$$\overline{F_s^1} = \frac{C(k,N) - (k-1)(q)^k}{1 - (q)^k},\tag{2}$$

where

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

This gives us the first measure of the sibling effect:

$$SE_1 = \overline{F_s^1} - \overline{M_s^1}.\tag{4}$$

Averaging over children. The second measure of "sibling effect" is computed as the difference between (1) the expected number of siblings for male children, and (2) the expected number of siblings for female children. Averaging over children takes account of differences in family sizes, that is, the actual number of boys and girls in each family.

Using Table 1 again, it can be seen that the expected number of siblings for a child given that it is a boy, $\overline{M_s^2}$, becomes

$$\overline{M_s^2} = \frac{\sum_{i=1}^{N-k} i(k-1+i)\binom{k-1+i}{i}(1-q)^i q^k + (N-1)\sum_{i=0}^{k-1}(N-i)\binom{N}{i}(1-q)^{N-i}q^i}{\sum_{i=1}^{N-k} i\binom{k-1+i}{i}(1-q)^i q^k + \sum_{i=0}^{k-1}(N-i)\binom{N}{i}(1-q)^{N-i}q^i}, \tag{5}$$

and the expected number of siblings for a child given that it is a girl, $\overline{F_s^2}$, is

$$\overline{F_s^2} = \frac{\sum_{i=1}^{N-k} k(k-1+i)\binom{k-1+i}{i}(1-q)^i q^k + (N-1)\sum_{i=0}^{k-1}i\binom{N}{i}(1-q)^{N-i}q^i}{\sum_{i=1}^{N-k} k\binom{k-1+i}{i}(1-q)^i q^k + \sum_{i=0}^{k-1}\binom{N}{i}(1-q)^{N-i}q^i}. \tag{6}$$

This gives us the second measure of the sibling effect:

$$SE_2 = \overline{F_s^2} - \overline{M_s^2}. \tag{7}$$

Observation 1. *If Assumption 1 is satisfied, then $SE_1 \geq 0, \forall N, 1 \leq k < N$.*

An algebraic proof of Observation 1 is available from the authors upon request. We avoided a similar algebraic result for SE_2 because it seemed much more cumbersome. Numerical computation, however, demonstrates that SE_2 is always positive. We do not include these computations in the article for the sake of brevity, but they can be obtained from the authors upon request.

The Birth-Order Effect

As we noted earlier, DSB implies that couples are more likely to stop childbearing if they have boys, rather than girls, at early parities. One implication of this is that boys will be born relatively later within families. We try to capture this quantitatively using the notion of an average within-family birth order (AWFBO) score: 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, G refers to a girl, and time moves from left to right). Here, the firstborn child was a boy, the second-born was a girl, the third- and fourth-born were boys, and 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 that families with no boys will not have an AWFBO score for boys, and those with no girls will not have an AWFBP score for girls.

To compute the mean AWFBO scores for boys and girls in the population, we 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. The first row does not have an AWFBO score for girls because these families have no female children; similarly, the last row does not have an AWFBO score for boys because these families do not have male children.

Using Table 2, the mean relative birth order for boys is

$$\overline{M}_{BO} = \frac{1}{1-(1-q)^N}\left\{\frac{k+1}{2}q^k + \frac{k+1}{k}\sum_{i=k}^{N-1}\binom{i}{k-1}\frac{i+1}{2}q^k(1-q)^{i+1-K}\right.$$

$$\left. + \frac{N+1}{2}\sum_{i=1}^{k-1}\binom{N}{i}q^i(1-q)^{N-i}\right\}. \tag{8}$$

Similarly, the mean relative birth order for girls is

Table 2. **Family Structures and AWFBO[a] Scores**

Total Children	AWFBO Score, Boys	AWFBO Score, Girls	Sibling per Child	Probability
k	$\dfrac{k+1}{2}$	—	$k-1$	q^k
$k+1$	$\dfrac{k+1}{2}\left(1+\dfrac{1}{k}\right)$	$\dfrac{k+1}{2}$	k	$\binom{k}{1}q^k(1-q)$
$k+2$	$\dfrac{k+2}{2}\left(1+\dfrac{1}{k}\right)$	$\dfrac{k+2}{2}$	$k+1$	$\binom{k+1}{2}q^k(1-q)^2$
...	...	...	...	...
N	$\dfrac{N}{2}\left(1+\dfrac{1}{k}\right)$	$\dfrac{N}{2}$	$N-1$	$\binom{N-1}{N-k}q^k(1-q)^{N-k}$
N	$\dfrac{N+1}{2}$	$\dfrac{N+1}{2}$	$N-1$	$\binom{N}{k-1}q^{k-1}(1-q)^{N-k+1}$
...	...	...	...	...
N	—	$\dfrac{N+1}{2}$	$N-1$	$(1-q)^N$

[a]The AWFBO score is the average within-family birth order score; for details see the text.

$$\overline{F}_{BO} = \frac{1}{1-q^k}\left\{\frac{N+1}{2}(1-q)^N + \frac{k+1}{k}\sum_{i=k}^{N-1}\left(k\frac{i}{-}1\right)\frac{i+1}{2}q^k(1-q)^{i+1-K}\right.$$

$$\left.+\frac{N+1}{2}\sum_{i=1}^{k-1}\binom{N}{i}q^i(1-q)^{N-i}\right\}. \tag{9}$$

Observation 2. *If Assumption 1 is satisfied, then* $BOE = \overline{M}_{BO} - \overline{F}_{BO} > 0, \forall N, 1 \le k \le N.$

Details of how to derive the expressions for $\overline{M}_{BO}$ and $\overline{F}_{BO}$, and an algebraic proof of Observation 2 for the case $q = 0.5$ are available from the authors upon request. We avoided a similar algebraic proof for a general value of q because it is more cumbersome and provides no additional insight. However, numerical computation demonstrated that $BOE = \overline{M}_{BO} - \overline{F}_{BO} > 0$ for a range of plausible values of q.

Some limitations of our definition of the birth-order effect deserve attention; future work will need to refine the AWFBO score to account for such issues. For instance, consider two families with the following birth sequences: BGGB and GBBG. For both families, the AWFBO scores are same for girls. On average, however, girls might be better off in the second family because one of the girls in the second family is the youngest child and might be able to escape all household responsibilities, whereas both girls in the first family might have to take care of younger sibling. Another problem of our formulation is that we have not explicitly taken account of birth spacing.

EMPIRICAL ANALYSIS

We carry out the empirical analysis 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 behavior, total fertility rates, and the interaction between the two (using household-level data on birth histories) using maximum likelihood estimation.

Table 3. Sibling and Birth-Order Effects for Asia and North Africa

Country	Survey Date[a]	Sample Size	Sibling Effect[b]		Birth-Order Effect[c]	
			Effect	t Statistic	Effect	t Statistic
Asia						
Bangladesh	1999–2000	10,544	0.07	2.62	0.04	2.36
India	1998–1999	90,303	0.13	14.93	0.03	4.92
Indonesia	2002–2003	29,483	0.00	0.83	0.03	2.81
Nepal	2001	8,726	0.18	6.57	0.03	1.61
Pakistan	1990–1991	6,611	0.06	1.33	−0.05	−1.46
Philippines	2003	13,633	0.03	1.15	−0.01	−0.57
Sri Lanka	1987	5,865	0.04	1.33	0.02	0.65
Thailand	1987	6,775	0.03	0.82	0.00	0.08
North Africa						
Egypt	2000	15,573	0.06	2.92	0.04	2.86
Morocco	2003	16,798	0.04	1.43	−0.02	−1.07

[a]Data for this analysis come from the Demographic and Health Surveys.

[b]Sibling effect is measured according to Eq. (4) in the text.

[c]For details of how the birth-order effect is measured, refer to Observation 2 in the text.

Sibling and Birth-Order Effects

For empirical analysis, we use data from the Demographic and Health Surveys (DHS; 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 the latest available survey or the one closest to the year 2000. (See Tables 3 and 4 for details of the data set for each country.) Apart from being a comprehensive survey covering almost all relevant aspects of health and educational indicators, the DHS 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 permits us to test our hypotheses regarding the sibling effect and the birth-order effect.[3]

Because male-preferring stopping rules entail decision-making (about having more children) conditional on the sex composition of existing children, we limit our calculations to living children within each family. Since couples try to reach a desired number of *living* sons and stop when they hit a ceiling for the maximum number of *living* children, looking at living children is more relevant than looking at all the children ever born. Additionally, we limit our computations to families that have completed their birth histories because the two effects will emerge fully only after fertility is complete. We present results for several countries in South Asia, Southeast Asia, North Africa, and sub-Saharan Africa in Tables 3 and 4. In these tables, we report SE_1 as the measure of the sibling effect. We also present results for India disaggregated at the state level in Table 5.

The results are along expected lines. Most of the countries in South Asia, Southeast Asia, and North Africa display significant sibling and birth-order effects. Countries in sub-Saharan Africa, on the other hand, do not display any statistically significant sibling or birth-order effects. There might be several reasons for the relative absence of both the

---

3. DHS data can be downloaded, with prior permission, from http://www.measuredhs.com.

Table 4. **Sibling and Birth-Order Effects for Sub-Saharan Africa**

Country	Survey Date[a]	Sample Size	Sibling Effect[b]		Birth-Order Effect[c]	
			Effect	*t* Statistic	Effect	*t* Statistic
Benin	2001	6,219	0.03	0.44	0.00	0.04
Burkina Faso	2003	12,477	0.03	0.75	−0.02	−0.53
Burundi	1987	3,970	0.06	0.74	−0.05	−0.72
Cameroon	2004	10,656	0.01	0.09	−0.04	−0.92
CAR	1994–1995	5,884	−0.01	−0.13	−0.01	−0.23
Chad	1996–1997	7,454	−0.05	−0.70	0.06	1.18
Comoros	1996	3,050	−0.13	−1.23	0.02	0.19
Cote d'Ivoire	1998	3,040	0.08	0.76	−0.09	−1.15
Ethiopia	2000	15,367	0.01	0.28	−0.02	−0.73
Gabon	2000–2001	6,183	−0.03	−0.36	0.00	0.06
Ghana	1998–1999	4,843	0.00	−0.03	−0.08	−1.79
Guinea	1999	6,753	−0.01	−0.26	−0.01	−0.34
Kenya	2003	8,195	0.02	0.50	−0.02	−0.56
Madagascar	1997	7,060	−0.02	−0.41	0.01	0.22
Malawi	2000	13,220	−0.04	−1.13	0.02	0.88
Mali	2000	12,849	0.02	0.37	−0.03	−0.97
Mozambique	2003	12,418	−0.02	−0.56	0.02	0.73
Namibia	2000	6,755	−0.04	−0.82	0.02	0.71
Niger	1998	7,577	0.03	0.41	0.00	−0.06
Nigeria	1999	9,810	0.11	1.78	−0.08	−1.77
Rwanda	2000	10,421	0.01	0.24	0.01	0.38
South Africa	1998	11,735	0.03	0.88	−0.03	−1.45
Sudan	1989–1990	5,860	−0.02	−0.31	−0.02	−0.43
Tanzania	1999	4,029	0.03	0.38	0.01	0.13
Uganda	2000–2001	7,246	0.01	0.12	0.01	0.40
Zambia	2001	7,658	−0.01	−0.09	0.00	−0.09
Zimbabwe	1999	5,907	0.00	0.04	0.02	0.55

[a]Data for this analysis come from the Demographic and Health Surveys.

[b]Sibling effect is measured according to Eq. (4) in the text.

[c]For details of how the birth-order effect is measured, refer to Observation 2 in the text.

effects in sub-Saharan Africa. Populations in sub-Saharan Africa might not have strong preference for sons. Alternatively, the absence of effects might be due to high fertility. A population that has both high fertility (high value of N) and a high numerical value of the son target (high value of k) can have small sibling and birth-order effects. This, rather than the absence of son preference, seems to be the case for countries in sub-Saharan Africa for two reasons. First, 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. Second, commonly used measures of son preference (e.g., the ratio of desired sons to daughters, or the proportion of families using contraceptives after two sons compared with

Table 5.   Sibling and Birth-Order Effects for Indian States

State	Sibling Effect[a]		Birth-Order Effect[b]	
	Effect	t Statistic	Effect	t Statistic
Andhra Pradesh	0.04	1.16	0.03	4.92
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
Maharashtra	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	4.70	0.02	0.78
Punjab	0.18	5.10	0.21	7.48
Rajashthan	0.19	6.19	0.09	3.90
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

[a]Sibling effect is measured according to Eq. (4) in the text.

[b]For details of how the birth-order effect is measured, refer to Observation 2 in the text.

those using contraceptives after two girls) show that many countries of sub-Saharan Africa display son preference.[4]

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 northeast generally do not show these effects. This is in accord with much evidence (anecdotal and otherwise) on the prevalence of patriarchal practices in different regions of the country.

_____

4. We do not report these results in any depth here because they are not the focus of this article; details are available from the authors upon request.

The Empirical Model

In the second step of the empirical analysis, we analyze the effect of covariates on son targeting behavior, fertility behavior, and the interaction between the two for India, a country that displays DSB (as seen in the previous subsection). To do so, we estimate the parameters of our simple model using the method of maximum likelihood with data from the 1992 DHS for India.

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 indicates girl). Let N_i denote the maximum number of children that family *i* would *like* to have, and let k_i denote the target number of boys for family *i*. Let $\mathbf{X}_i$ denote a vector of covariates that determines the probability of son targeting behavior for family *i*, and let $\mathbf{Z}_i$ denote a vector of covariates that determines the desired maximum number of children, N_i, for family *i*; note that $\mathbf{X}$ and $\mathbf{Z}$ can contain common variables.

This analysis concerns the population of families with completed birth histories. To estimate the effect of covariates on targeting and fertility behavior, we will calculate the joint likelihood of observing a given birth sequence (S_i) *and* desired 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 that indicates whether family *i* targets sons. $T_i = 0$ means that the family does not target sons, and $T_i = 1$ implies that family *i* does target sons. Finally, we let T_i be determined by a vector of observable covariates, $\mathbf{X}_i$, in the following manner:

$$T_i = \begin{cases} 0 \text{ if } \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i \le 0 \\ 1 \text{ if } \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i > 0 \end{cases}, \tag{10}$$

where $\mathbf{X}_i$ is a $(1 \times k)$ vector of covariates that determine whether a particular family targets sons; $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of parameters to be estimated; and $\varepsilon_i \sim N(0,1)$ captures the unobservable, stochastic factors that affect son targeting behavior.

To obtain the likelihood for the observed birth sequence and maximum number of children for the *i*th family, note that

$$P(S_i,N_i) = P(S_i,N_i \mid T_i = 0)P(T_i = 0) + P(S_i,N_i \mid T_i = 1)P(T_i = 1)$$

$$= P(S_i,N_i \mid T_i = 0)P(T_i = 0) + P(S_i,N_i \mid T_i = 1)P(T_i = 1)$$

$$= P(S_i,N_i \mid T_i = 0)P(N_i \mid T_i = 0)P(T_i = 0) + P(S_i \mid N_i,T_i = 1)P(N_i \mid T_i = 1)P(T_i = 1)$$

$$= I(n(S_i) = N_i)q^{b_i}(1 - q)^{g_i}P(N_i \mid T_i = 0)P(T_i = 0) + P(S_i \mid N_i,T_i = 1)P(N_i \mid T_i = 1)$$
$$\quad P(T_i = 1)$$

$$= I(n(S_i) = N_i)q^{b_i}(1 - q)^{g_i}P(N_i \mid T_i = 0)\,\Phi(-\mathbf{X}_i\boldsymbol{\beta}) + P(S_i \mid N_i,T_i = 1)P(N_i \mid T_i = 1)$$
$$\quad [1 - \Phi(-\mathbf{X}_i\boldsymbol{\beta})],$$

where $\Phi(.)$ denotes the standard normal cumulative distribution function, $I(.)$ denotes the indicator function, b_i is the number of boys and g_i is the number of girls in the *i*th family, and $n(S_i)$ denotes the number of children in birth sequence S_i; q is the probability of a male birth, and for the estimation, we use a value of $q = 0.514$, which is a widely accepted figure (see, e.g., Hesketh and Wei Xing 2006).

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 when the family does not target sons (i.e., $P(S_i \mid N_i,T_i = 0)$) is $I(n(S_i) = N_i)q^{b_i}(1 - q)^{g_i}$. The indicator function is meant to rule out the

possibility that a family that stops childbearing before reaching its desired maximum number of children might not target sons: in our model, any family that stops childbearing before hitting the ceiling, N_i, targets sons. This gives us the first term in the joint probability expression above. The second term comes from families that target sons, so we need to compute $P(S_i \mid N_i, T_i = 1)$.

When a family targets sons, its target, k_i, can range anywhere from 1 to $N_i - 1$; targeting $k_i = N_i$ sons with a ceiling for the desired maximum number of children set at N_i is equivalent to not targeting sons. Because 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:

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

where $P(S_i \mid N_i, k_i, T_i = 1)$ is the probability of observing S_i given N_i, k_i, and $T_i = 1$ (son targeting), and $P(k_i \mid N_i, T_i = 1)$ is the probability of targeting k_i sons given that the desired maximum number of children is N_i. We do not observe these probabilities; in our model, we treat them as parameters and estimate them jointly with other parameters. The summation in Eq. (11) follows from an application of the law of total probability.

Three things should be noted about Eq. (11). First, the summation runs until $(N_i - 1)$ because $k_i = N_i$ is equivalent to not targeting. Second, we consider only cases in which $N_i \geq 2$; this follows from the intuition that families with a desired maximum number of children below 2 cannot target sons in any meaningful sense. Third, in computing the conditional probabilities $P(S_i \mid N_i, k_i, T_i = 1)$, we use information not only about the number of sons and the number of total children but also about the *birth order* of the children; details are available in the appendix.

Using Eq. (11), therefore, the expression for the joint probability becomes

$$P(S_i, N_i) = I(n(S_i) = N_i) q^{b_i} (1-q)^{g_i} = P(N_i \mid T_i = 1) \Phi(-\mathbf{X}_i \boldsymbol{\beta})$$
$$+ [1 - \Phi(-\mathbf{X}_i \boldsymbol{\beta})] \sum_{k_i=1}^{N_i-1} P(S_i \mid N_i, k_i, T_i = 1) P(k_i \mid N_i, T_i = 1) P(N_i \mid T_i = 1).$$

Next, we assume that N_i conditional on T_i is distributed as a Poisson random variable with a conditional mean given by λ_i. We try to capture two crucial facts with this formulation: (1) that N_i conditional on T_i is a count variable, and (2) that there is an interaction between the decision of son targeting and the desired maximum number of children (the interaction term appears in the expression for the conditional mean, λ_i).

$$P(N_i \mid T_i) = \frac{\exp(-\lambda_i) \lambda_i^{N_i}}{N_t!},$$

where

$$\lambda_i = \exp(\mathbf{Z}_i \boldsymbol{\gamma} + \alpha T_i). \tag{12}$$

In Eq. (12), α captures the effect of son targeting on the total fertility rate, and $\mathbf{Z}_i$ is a set of covariates that affects desired family size. Moreover, because T_i is a dichotomous variable, we have

$$P(N_i \mid T_i = 0) = \frac{\exp(-\exp(\mathbf{Z}_i \boldsymbol{\gamma}))[\exp(\mathbf{Z}_i \boldsymbol{\gamma})]^{N_i}}{N_i!}, \tag{13}$$

and

$$P(N_i \mid T_i = 1) = \frac{\exp(-\exp(\mathbf{Z}_i \boldsymbol{\gamma} + \alpha))[\exp(\mathbf{Z}_i \boldsymbol{\gamma} + \alpha)]^{N_i}}{N_i!}. \tag{14}$$

Using Eqs. (13), (14), and (11), we can write the expression for the joint probability as follows:

$$P(S_i, N_i) = I(n(S_i) = N_i) q^{b_i} (1-q)^{g_i} \frac{\exp(-\exp(Z_i\gamma))\left[\exp(Z_i\gamma)\right]^{N_i}}{N_i!} \Phi(-X_i\beta)$$

$$+ \left[1 - \Phi(-\mathbf{X}_i\beta)\right] \sum_{k_i=1}^{N_i-1} P(S_i \mid N_i, k_i, T_i = 1) P(k_i \mid N_i, T_i = 1)$$

$$\times \frac{\exp(-\exp(\mathbf{Z}_i\gamma + \alpha))\left[\exp(\mathbf{Z}_i\gamma + \alpha)\right]^{N_i}}{N_i!}. \tag{15}$$

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

$$l = \log(L) = \sum_{i=1}^{n} \log(P(S_I, N_i)),$$

where n is the number of families in the sample, and $P(S_i, N_i)$ is substituted from Eq. (15). Maximizing l will give the estimates of the parameters of interest in the model: α (interaction term), β (targeting behavior), γ (determinants of family size), and the probabilities $P(k_i \mid N_i, T_i = 1)$. 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 $P(k_i \mid N_i, T_i = 1)$ is the probability of targeting k_i sons given that the family desires a maximum of N_i children.

An alternative method of empirical analysis might proceed by estimating a model of parity-progression rates, as in Gray and Evans (2005), for instance. We choose to follow our method for two reasons. First, our empirical method fits the theoretical part of our argument much more closely and follows from it naturally. Second, our theoretical and empirical analysis allows us to draw out some implications of son targeting behavior that might have important effects on gender inequality that we wish to explore in the future. A model of parity-progression rates would allow us to detect the presence of son targeting behavior without highlighting the sibling and birth-order effects.

Results for India

Covariates and the estimates for India are presented in Tables 6 and 7. Signs on most variables in Table 6 are along expected lines. Let us first look at the targeting equation (Eq. (10)). The covariates that do not seem to affect the probability of targeting are years of formal education, whether the respondent lives in an extended family, and middle income.

The estimates also indicate that age of the mother has a negative impact on the probability of son targeting. In addition, families in rural areas are more likely than those in urban areas to target sons through male-preferring stopping behavior. Thus, geographic location seems to matter for gender inequality. Respondents' participation in the labor force positively affects the probability of targeting; this is rather surprising and runs counter to earlier suggestions (see, e.g., Sen 1990). Income dummy variables show that compared with wealthy families, low- and middle-income families have a higher probability of son targeting, which is more or less in agreement with anecdotal evidence. The use of contraceptives significantly increases the probability of targeting; this is to be expected because without contraceptives, it would be difficult to enforce differential stopping behavior (or any other kind of stopping behavior). Our results suggest that Hindu families are more likely to target sons than are non-Hindu families. Based on anecdotal evidence on Hinduism suggesting that a religious sanction is behind son preference, this finding is not surprising.

Table 6.   Estimation Results[a] for India, 1992

Variable	Coefficient	SE
Average Family Size		
Age of the mother	0.019	0.0003
Education	−0.007	0.0005
Work	−0.007	−0.005
Rural	0.065	−0.006
Low income	0.546	−0.013
Middle income	0.042	−0.012
Hindu	0.159	−0.009
Extended family	0.002	−0.008
Interaction Term		
Intercept	0.042	−0.008
Son Targeting		
Intercept	−1.18	−0.094
Age of the mother	−0.005	−0.001
Education	0.0008	−0.002
Work	0.047	−0.022
Contraception use	0.562	−0.023
Rural	0.08	−0.026
Extended family	0.017	−0.031
Hindu	0.081	−0.039
Low income	0.235	−0.079
Middle income	0.066	−0.077

[a]For a definition of the model, refer to Eq. (15).

Turning to the equation for the determination of family size (Eq. (12)), all the variables in Table 6 significantly affect the dependent variable except the dummy variables for work and extended family. Along expected lines, age of the mother is associated with increases in the average family size, and years of formal education is associated with reductions in the average family size; rural families have larger families. The income dummy variables show expected results: compared with wealthy households, both middle-income and low-income households have larger average family sizes, and the effect is stronger for lower-income families. Turning to the religion dummy variable, Hindu families have larger family sizes than non-Hindu families.

Along expected lines, the interaction term is positive and highly significant; families that are more likely to target sons 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 targeted number of boys. The targeting probabilities in Table 7 reveal a simple pattern: conditional on targeting, couples are more likely to target a higher than lower number of sons. For instance, families that set a ceiling at five children are more likely to target four boys than three, three boys more than two, and two boys more than one.

One possible criticism of the results of this empirical analysis might be the following: we treated the probability of male birth as exogenous in our model, but in reality, it might be affected by human intervention in the form of sex-selective abortion and thus may be endogenous. Although this concern is valid in general, we think that our empirical analysis

Table 7.   **Estimated Targeting Probabilities[a]**

Targeting Probability	Estimate
$P(k = 1 \mid N = 3, T = 1)$	.273
$P(k = 2 \mid N = 3, T = 1)$	.726
$P(k = 1 \mid N = 4, T = 1)$	.204
$P(k = 2 \mid N = 4, T = 1)$	.399
$P(k = 3 \mid N = 4, T = 1)$	.398
$P(k = 1 \mid N = 5, T = 1)$	.054
$P(k = 2 \mid N = 5, T = 1)$	.205
$P(k = 3 \mid N = 5, T = 1)$	.315
$P(k = 4 \mid N = 5, T = 1)$	.426
$P(k = 1 \mid N = 6, T = 1)$	.058
$P(k = 2 \mid N = 6, T = 1)$	.177
$P(k = 3 \mid N = 6, T = 1)$	.204
$P(k = 4 \mid N = 6, T = 1)$	.261
$P(k = 5 \mid N = 6, T = 1)$	.300

[a]For a definition of these probabilities, refer to the appendix. All probabilities reported here are statistically significant at the 5% level.

is not affected by this problem. The reason is that ultrasound technology for the sex determination of the fetus became available only from the late-1980s onward in India and could not have affected most of the births recorded in the DHS 1992 (the data set that we used for our ML estimation); the effects of sex-selective abortion can be detected only in surveys from a later period. Hence, the problem of endogeneity of the probability of male birth should not be a serious problem for the empirical results in this article.

CONCLUSION

In developing economies, son preference has often been expressed through son targeting fertility behavior. In this article, we drew out two demographic implications of son targeting fertility behavior: (1) on average, girls will be born into larger families, and (2) on average, girls will be born at relatively earlier parities within families. These two implications of son targeting fertility behavior might be useful for explaining the generation and perpetuation of gender inequality even in the absence of intrahousehold biases in resource allocation against daughters. The fact that girls are born into larger families means that they have to share resources with many siblings; this might put them at a disadvantage even when parents do not discriminate against girls. The fact that girls are born at relatively earlier parities might also work to their disadvantage. In poorer and larger families in which both parents work to make ends meet, part of the parental responsibility for younger children is passed on to older children in the family. Because most of the older 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 might not be able to devote their full time and energy to their own education or other recreational activities. We wish to explore these possibilities in our future research.

Our empirical analysis shows that both the implications of son targeting fertility behavior are present in several countries in Asia and North Africa, though they are absent in sub-Saharan countries. This might be evidence for the existence and strengthening of son preference in the former group of countries. When we turn our attention to the

determinants of targeting behavior, we find that geographic location and income strongly affect son targeting. We also find that families that target sons are also more likely to have larger families.

APPENDIX

In this appendix, we sketch the method that we have used to compute the conditional probabilities, $P(S_i \mid N_i, k_i, T_i = 1)$, that have been used in the maximum likelihood estimation;[5] for illustration, we work with a probability of male birth of (1/2) here. The logic of our method is straightforward. For every family, we are given a completed birth sequence (S_i) and the desired maximum number of children (N_i). For such a family, we must compute the following ($N_i - 1$) conditional probabilities: $P(S_i \mid N_i = 4, k_i = 1, T_i = 1)$, $P(S_i \mid N_i = 4, k_i = 2, T_i = 1)$, and $P(S_i \mid N_i = 4, k_i = 3, T_i = 1)$. We need to compute all these probabilities because we do not observe the desired target for sons.

Because, *a priori*, we do not know the desired target (for sons) for family i, we need to allow for all feasible possibilities. Thus, when a family states that the maximum number of children that it desires 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 these possibilities, but we cannot rule any of these out *a priori*.

To compute something like $P(S_i \mid N_i, k_i, T_i = 1)$, we merely need to observe whether the family has *any* child after the k_ith son. If there is a child after the k_ith son, we assign zero probability to $P(S_i \mid N_i, k_i, T_i = 1)$; otherwise 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 desires to have is 4 and the birth sequence (starting with the first-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: $P(GGBG \mid N_i = 4, k_i = 1, T_i = 1)$, $P(GGBG \mid N_i = 4, k_i = 2, T_i = 1)$, and $P(GGBG \mid N_i = 4, k_i = 3, T_i = 1)$. Because there is a child after the first boy, this family could not possibly be targeting one son; hence, $P(GGBG \mid N_i = 4, k_i = 1, T_i = 1) = 0$. But the family could conceivably be targeting two or even three sons; these possibilities are not ruled out by the observed birth sequence. Hence, $P(GGBG \mid N_i = 4, k_i = 2, T_i = 1) = (1 / 16)$; similarly, $P(GGBG \mid N_i = 4, k_i = 3, T_i = 1) = (1 / 16)$.

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 completed birth sequence for the same family to be BGB. Because there is a child after the first boy, this family could not possibly be targeting one son; hence, $P(BGB \mid N_i = 4, k_i = 1, T_i = 1) = 0$. Because there is no child after the second boy, the family could possibly be targeting 2 sons; hence, $P(BGB \mid N_i = 4, k_i = 2, T_i = 1) = (1 / 8)$. And because the family stops at three children (with two sons), it cannot be targeting three sons. Hence, $P(BGB \mid N_i = 4, k_i = 3, T_i = 1) = 0$.

REFERENCES

Arnold F., M.K. Choe, and T.K. Roy. 1998. "Son Preference, the Family-Building Process and Child Mortality in India." *Population Studies* 52:301−15.

Bhaskar, V. and B. Gupta 2007. "India's Missing Girls: Biology, Customs, and Economic Development." *Oxford Review of Economic Policy* 23:221−38.

Clark, S. 2000. "Son Preference and Sex Composition of Children: Evidence From India." *Demography* 37:95–108.

Gray, E. and A. Evans. 2005. "Parity Progression in Australia: What Role Does Sex of Existing Children Play?" *Australian Journal of Social Sciences* 40:505–20.

---

5. R code for these and other computations, including the maximum likelihood estimation, are available from the authors upon request.

Hesketh, T. and Z. Wei Xing. 2006. "Abnormal Sex Ratios in Human Populations: Causes and Consequences." *Proceedings of the National Academy of Sciences of the United States of America* 103:13271–75.

Jensen, R.T. 2002. "Equal Treatment, Unequal Outcomes? Generating Sex Inequality Through Fertility Behaviour." Unpublished document. JFK School of Government, Harvard University.

Larsen, U., W. Chung, and M. Das Gupta. 1998. "Fertility and Son Preference in Korea." *Population Studies* 52:317–25.

Oster, E. 2005. "Hepatitis B and the Case of Missing Women." *Journal of Political Economy* 113:1163–216.

Ray, D. 1998. *Development Economics*. Princeton, NJ: Princeton University Press.

Seidl, C. 1995. "The Desire for a Son is the Father of Many Daughters: A Sex Ratio Paradox." *Journal of Population Economics* 8:185–203.

Sen, A. 1990. "More Than 100 million Women are Missing." *New York Review of Books,* December 20.

Yamaguchi, K. 1989. "A Formal Theory for Male-Preferring Stopping Rules of Childbearing: Sex Differences in Birth Order and in the Number of Siblings." *Demography* 26:451–65.