WHO ARE SEX-SELECTING, AND WHEN?

Minu Philip*

February 2022

ABSTRACT

Son-preferring parents are known to manipulate the sex-composition of their children using differential fertility stopping or pre-natal sex-selection. In this paper, we examine birth histories of Indian women to identify the heuristics that they largely use in deciding *when* to sex-select, if at all. We exploit the non-relation between the sex of a newborn and the length of the preceding birth-interval to identify instances of pre-natal sex-selection that would artificially increase the interval preceding male births. We combine this with mothers' reported parity preferences to infer heuristics of parental manipulations. Using a novel approach of redefining birth orders relative to the mother's ideal parity, we find that son-preferring parents typically engage in sex-selection when they are at their ideal parity so as to avoid exceeding it. This is also true at the first birth order where some mothers with an ideal parity of one resort to sex-selection. After select birth histories, we also find evidence of reversals in the decision to sex-select following unsuccessful attempts.

Keywords: son preference, sex-selection, sex-ratio, fertility, parity

JEL Codes: I12, I18, I21, J13, J16

*Department of Economics, New York University.
I am very grateful to my advisors — Professor Debraj Ray and Professor Martin Rotemberg — for their continuous guidance and counsel. Their helpful feedback and comments have refined this project to its current form. I am also thankful to Alberto Bisin, Katarína Borovičková and Prashant Bharadwaj for their useful comments and feedback.

1 Introduction

Son-preference, or a parental bias for male children, is widespread, particularly in South and East Asia, and has far-reaching implications for important aspects such as population demographics, marriage market equilibria, and maternal and child health. The use of differential fertility stopping rules (Clark 2000; Jayachandran and Kuziemko 2011; Baland, Cassan, and Woitrin 2020) or pre-natal sex-determining technologies (Jha et al. 2006; Bhalotra and Cochrane 2010; Pörtner 2016, 2020), has been evidenced extensively in the literature, with the consequences also discussed therewith.[1] What is not known, however, are the precise heuristics (or rules of thumb) of these sex-composition manipulating strategies that parents execute at various birth orders.[2]

In this paper, we analyse birth histories of Indian women as recorded in the National Family Health Survey (NFHS) 2015-16 to identify the sex-composition manipulating practices that parents resort to and the precise heuristics that they largely follow. India presents as a suitable context for this study given the widespread prevalence of son-preference as suggested by its very high sex-ratios. Moreover, India's ongoing fertility transition generates sufficient variation in fertility rates to facilitate our analyses, unlike other low-fertility or high-fertility countries.

Identifying heuristics of parental manipulations is, however, quite challenging. This is so because of two reasons: (1) parental manipulations are neither directly observed by the researcher nor truthfully reported (if reported at all), and (2) preferences that motivate these choices are also typically not observed.[3] Any attempt at identifying such heuristics must necessarily focus on *inferring* these unobserved preferences and fertility choices from the birth histories and birth outcomes. A complete birth history records the order, sex and inter-birth interval for all children born to a mother.

1. Differential fertility stopping refers to when parents' decision to stop bearing children depends on their already-born children's sex-composition. It is common for parents with no or few sons to continue childbearing unlike others with the same number of children.

2. By sex-composition manipulating strategies, we mean practices that potentially modify the sex-composition of all children *born live* to a parent. This includes differential fertility stopping and sex-selection, but excludes post-natal discrimination. Sex-selection, here, refers to *all* methods that allow parents to directly influence the sex of their successive child. These include: post-implantation manipulation (such as pre-natal sex determination and subsequent abortion of fetuses of unwanted sex), as well as pre-implantation interventions such as sperm sorting or embryo selection.

3. This is because pre-natal sex determination of fetuses is illegal in India under the Pre-conception and Prenatal Diagnostic Techniques Act (PCPNDT, 1994); and parents may misrepresent their preferences to avoid acknowledging that they are son-preferring.

Even with detailed birth histories, it is not trivial to separate instances of sex-selection from differential fertility stopping. This is why the literature has often studied the two separately. Driven by a common primitive of son-biased preferences, differential fertility stopping and sex-selection often either have congruent or ambiguous implications for *observed* fertility outcomes. For example, a key feature of sex-selection which is supposed to differentiate it from differential stopping behavior is that it alters the sex-ratio at birth by artificially disallowing girls from being born. However, differential stopping behavior could also increase the ratio of males to females at the highest birth order, since parents' decision to stop childbearing becomes endogenous to the sex of the last child. Therefore, comparing sex-ratios at various birth orders for parents who have different parity (i.e. total number of children) is not a sound approach. It is also obvious that any of these methods are likely resorted to following similar birth histories — those with few or no sons.[4] Quite apart from this, there is also the issue of a lack of consensus on what the appropriate value of natural birth sex ratio should be for us to benchmark against.

To get around the problem of correctly identifying instances of sex-selection, we exploit the natural sex-determination system in humans that renders the sex of a newborn unrelated to the length of the preceding inter-birth spacing, and independent of the sex-composition of any children already born to the mother. Since sex-selective interventions that result in a live male birth would increase the *observed* inter-birth interval, one could infer sex-selection from birth intervals being longer when preceding male births. This increase in the observed inter-birth interval due to sex-selection is on account of inability to determine sex before 10-12 weeks of pregnancy, time needed for post-abortion recovery and the time taken till the next conception (Pörtner 2016, 2020). The dynamic nature of fertility choices made between two live births is also revealed as we study changes in the proportion of male births at varying lengths of the preceding interval, where we observe a tendency for parents to reverse their decision to sex-select after the inter-birth interval gets 'too long'.

We use a difference-in difference specification that exploits variations in birth history and the newborn's sex, to identify sex-selective behavior. The identification is based on the assumption that even if there are any unaccounted factors that could likely skew the natural probability of birthing a son *and* also influence the length of the preceding inter-birth interval; they should confound the relationship between sex and

4. Other signs of differential stopping — such as differences in parity progression following various histories; decrease in proportion of sons with increase in total number of children (Clark 2000); or girls being more likely born at lower birth orders and into families with larger number of children (Basu and Jong 2010; Baland, Cassan, and Woitrin 2020) — do not preclude instances of sex-selection.

preceding interval, if at all, for any birth history. This assumption is reasonable as long as the sex at conception is determined by nature independently of previous pregnancies. Since the incentives to sex-select vary with birth history but the effect of any confounding variables shouldn't, any significant difference across histories in the length of the preceding interval based on the sex of the newborn must be attributed to parental manipulation. An important caveat here, is that the direction of this relation would change when parents 'take-back' or revert on their decision to sex-select. Moreover parents may also anticipate that sex-selective methods take time and hence begin sooner. Nevertheless, any difference in length of birth preceding interval based on the sex of a newborn, spells manipulation.

To then be able to identify parental strategies or heuristics, we need to relate these specific instances of manipulations with parental preferences.

Therefore, to infer heuristics, we link parents' reported preferences for ideal parity and ideal sex-composition with their birth histories. Recall that parity here refers to the total children born alive to a mother. In particular, we redefine the birth orders of all children born to a mother in relation to her actual and ideal parity. The relative birth orders thus created, allow us to compare births at different birth orders for mothers who have different parities. Comparing the proportion of sons at various relative birth orders, we find a significant increase in the proportion of male births when mothers are at their ideal parity. Since parity preferences should theoretically be orthogonal to outcomes such as sex and birth history, that are determined by nature; high proportion of male births when mothers are at their ideal parity strongly implies use of sex-selection at the ideal parity to avoid exceeding it.

We therefore infer a heuristic whereby son-preferring parents who want to sex-select, do so at their ideal parity, while others continue childbearing in the hopes of birthing a son. To validate the same, we use an empirical specification that compares correlations between sex and length of preceding birth interval for sub-samples of parents who would sex-select according to our heuristic versus those who wouldn't. The findings from this empirical exercise confirm our heuristic. Extending this analysis to the first birth order, we find evidence of mothers with an ideal parity of one resorting to sex-selection as early as at the first birth order. We also find that parents who choose differential stopping but bear no sons at all, are likely resort to sex-selection at higher birth orders.

Our findings not only provide insights into the heuristics that have emerged among parents who continually choose between heeding to natural odds of sex-determination versus engaging in sex-selection, but also warn against expecting that the issue of sex-selection will resolve by its own with economic development or increased education. In fact, in light of the continuous decline in desired parity, this only become more pertinent.

In what follows, we first describe a conceptual framework in Section 2, which details the patterns and trends that parents' fertility choices would generate for observed birth-statistics such as proportions of sons and inter-birth intervals. We then use data on birth histories to infer the sex-manipulating strategies and the precise heuristics adopted by parents. Section 3 details the data-set used for our analyses and Section 4 describes our findings. Section 5 concludes.

2 Conceptual Framework

2.1 Setup

We model parents as single non-gendered agents who, given their fertility preferences and concurrent birth history, decide whether to conceive again, and whether to engage in sex-selection upon conception.[5] These decisions translate into a sex-sequence of live births for each parent that captures the order and sex of each child born, as well as the sex-composition up to each birth order. For example, `FFM` is the sex-sequence that represents a history of three births, youngest of which is of a boy.

2.1.1 Preferences

Preferences can be described for each parent as a complete and transitive ordering over all possible sex-sequences including an empty sequence representing no births. At the minimum, this ordering is based on parity i.e. the number of elements in the sequence. For several parents, son-preferring or otherwise, it would also depend on the sex-composition. A parent, for example, may prefer the sex sequence `FM` with two children the most, but conditional on birthing `FF`, could now want a third child if they preferred `FFM` over `FF`.

The ideal parity according to the parent would be the parity of their most preferred sex sequence. Likewise, the 'ideal' number of sons/daughters would be the respective frequencies of `M`/`F` in their most preferred sex sequence. While the ideal parity is well-defined for all parents, ideal composition is undefined for parents who are indifferent about the sex-composition of their children.

5. By modeling parents as single non-gendered agents, we abstain from modeling any gendered differences in fertility preferences and bargaining between spouses. Rasul (2008), for example, models household bargaining over fertility.

2.1.2 Actions

Until parents complete their fertility, they continually face two decisions: whether to have another child or not, and whether to sex-select at the successive pregnancy or not. Therefore, any parent with a given birth history h — where h represents the sex-sequence of children already born to the parent — could either stop childbearing and remain at h, or could transit to new history $(h, \texttt{F})$ following the birth of a girl, or to $(h, \texttt{M})$ following that of a boy; as is represented in Figure 1. Likewise, those who want to sex-select would either selectively abort their female fetus and return to h, or birth their male child and transit to $(h, \texttt{M})$.[6]

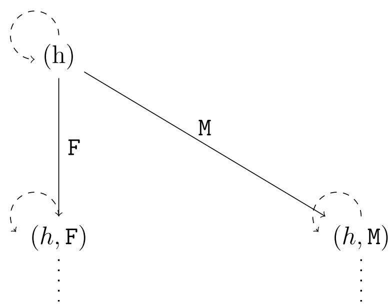

Figure 1: Birth histories and transitions

All parents therefore begin with a null history or an empty sequence, and based on their actions and realizations of sex of the children born to them, settle over time into various 'nodes' representing their birth histories. Figure 2 describes this transition across 'nodes' for a parent who after the birth of their first female child, aborts the female fetus conceived at their second attempt, and eventually births a male child before stopping childbearing altogether after a birth history of `FM`.

A parent with a given history h, faces these choices continually, *until* she either births a child and 'transits' to another history or ends her fertility. In any data, we only observe the transitions to various histories (via the sex-sequence of live births) and the time elapsed between these transitions (recorded as the inter-birth interval), but do not observe the actions chosen by the parent between these transitions. Since the actions aren't observable, we must rely on observable birth statistics — effected as a

6. By making the two decisions of having another child or not and sex-selecting or not, the parent essentially chooses from among three lotteries. The first is a degenerate lottery associated with not bearing another child and leading to the deterministic outcome h. The second, associated with the decision to bear another child without sex-selection, is a lottery over $(h, \texttt{F})$ and $(h, \texttt{M})$. Third, the decision to sex-select, is essentially a lottery over h and $(h, \texttt{M})$.

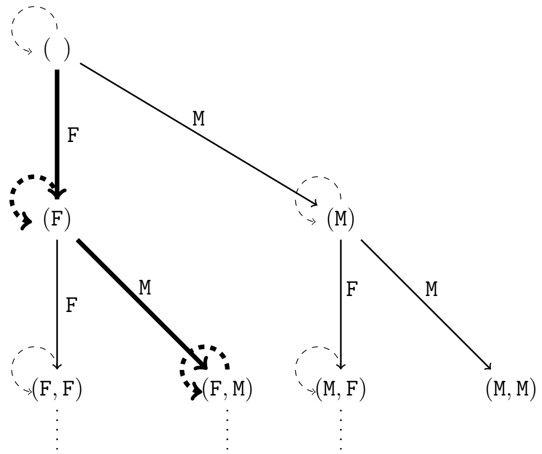

Figure 2: Example of transition across histories based on parent's actions. The highlighted path represents the transition across nodes (histories) for a parent who after the birth of their first female child, aborts the female fetus conceived at their second attempt and eventually births a male child in their third attempt, before stopping altogether at FM.

consequence of these actions — to back-out the fertility choices made by parents.

There are two ways in which parents manipulate the sex-composition of their children and transit to a preferred 'node' given any history. One is differential stopping behavior, whereby parents with few or no sons continue childbearing in order to achieve a more desirable sex-composition. The other is sex-selection, by which parents use technology to eliminate the transition to $(h, \texttt{F})$ by disallowing a successive live female birth. Based on this setup,

- if $\exists$ a pair of histories $h \neq h'$, with equal number of children (say, n) but dissimilar sex-composition, such that the tendency to stop childbearing differs following h compared to that following h', then it is indicative of *differential stopping behavior* at parity n
- if any positive measure of parents selectively eliminate female births given a concurrent history h, it is indicative of *sex-selection* following h

In the following Section we discuss some ways to identify differential stopping and sex-selective behavior based on observed realizations of birth histories. We will later employ these under Section 4 to identify sex-manipulating heuristics followed by parents.

2.2 Inferring parental manipulations from birth histories

2.2.1 Crude estimates of sex-selectors and fertility stoppers

A natural approach to inferring parental manipulations is to use the proportion of parents who continue childbearing after a given history (parity progression ratio) and the proportion of those who birth a successive male child (Jha et al. 2006; Sen 1992, 1990). If the natural probability of giving birth to a male child were known (let's denote this by π), we could, under some assumptions, deduce the proportion of sex-selectors based on excess male births following any given history.

Let N_h be the number of mothers whose birth history begins with the sequence h. Further we denote the proportion of mothers who stop childbearing following history h by μ_h, and the proportion of mothers who sex-selectively abort following history h by λ_h. If we assume that all mothers who selectively-abort their female fetuses conceive again up to $j < \infty$ times to be able to bear a male child, then following any history h, we should observe

$$\begin{aligned}
\text{Number of female successive births} &= \big[(1-\pi)(1-\lambda_h)\big](1-\mu_h)N_h \\
\text{Number of male successive births} &= \big[\pi + \lambda_h(1-\pi)(1-(1-\pi)^j)\big](1-\mu_h)N_h \\
\text{Total number of successive births} &= \big[1-\lambda_h(1-\pi)^{j+1}\big](1-\mu_h)N_h
\end{aligned}$$

Here, female successive births would occur only to parents who aren't sex-selectors and naturally conceive a female fetus. Male successive births would be to those who either conceive a male child at their first attempt, or selectively abort female fetuses enough times to conceive a male child in less than j attempts.

Using these, we can obtain estimates of μ_h and λ_h based on actual counts of births of either sex observed in the sample. (see Appendix for details)

Let $\hat{p_h}$ be the proportion of male successive births observed in the sample following h. Then, the proportion of sex-selectors (λ_h) can be calculated as

$$\hat{\lambda_h} = \frac{\hat{p_h} - \pi}{(1-\pi) - (1-\hat{p_h})(1-\pi)^{j+1}}$$

Likewise, the proportion of fertility stoppers (μ_h), by

$$\hat{\mu_h} = 1 - \frac{\#\ \text{Successive female births}}{(1-\pi)N_h(1-\hat{\lambda_h})}$$

Intuitively, it compares the number of female births observed in the sample with a case where all parents had continued childbearing (including a fraction of sex-selectors), to infer the proportion of fertility stoppers from the shortfall in female births. Recall that if for two distinct histories h and h' with the same parity, we observe $\mu_h \neq \mu_{h'}$, then it is suggestive of differential stopping behavior. In particular, we expect to observe higher proportion of fertility stoppers following histories that have a higher proportion of boys.

In using these estimators, we are making the following assumptions:

1. π is known and fixed for all mothers
2. Sex-selective abortion is the *only* method of sex-selection practiced. (Ignores methods such as sperm sorting, embryo selection etc.)
3. Sex-selective abortion is used exclusively for aborting female children.
4. *All* mothers willing to sex-selectively abort, make j such attempts to birth a boy and stop childbearing after j failed attempts. There is no allowance of take-back or reversal of the decision to sex-select.

This approach whilst intuitive makes too many assumptions. There is also no consensus on what the appropriate values of π and j should be, and whether they should be the same for all parents or not. Choice of a lower value for π, for example, would inflate the proportion of sex-selectors in the sample. Relative comparisons, however, are still meaningful.

2.2.2 Evidence for Sex-selection: Changing odds of male live births

The exercise described above estimates crudely the proportion of sex-selectors by comparing the proportion of male successive births against a given value of π (with suitable adjustments). There are two issues with relying on 'high' proportions of sons to infer sex-selection. The first, is that the natural birth sex ratio has been found to vary by age, geography, ethnicity, socio-economic status and environmental factors, among others (Chao et al. 2019; Anderson and Ray 2010); and hence there is no precise estimate of a *benchmark* probability of male birth.

Second, and more importantly, even if the natural probability of male births was known and fixed, we would never be able to infer sex-selection correctly *at the last birth order* based on 'high' proportions of sons relative to this known benchmark. This is because proportion of sons at the last birth order would be significantly high even among differential fertility stoppers.

To tackle this, we employ another observed variable: the inter-birth interval. Inter-birth intervals are typically effected by parents' *intended* birth-spacing and the mother's fecundity. The former typically depends on parents' opportunity costs of time and resources while the latter depends on maternal age, post-partum amenorrhea and contraceptive-use. With son preference, however, there is reason to believe that parents with less than desired number of sons may try to conceive sooner (Jayachandran and Kuziemko 2011). This would therefore reduce the length of birth intervals succeeding birth histories with fewer sons. This decrease in the length of intervals *succeeding* such histories would hold for all son-preferring mothers who try to conceive sooner, no matter their decision to birth naturally or through sex-selective methods. Studying succeeding birth intervals therefore would not help us distinctively identify instances of sex-selection.

Interestingly, sex-selective abortion uniquely affects the length of the observed birth interval *preceding* the birth at which sex-selection was resorted to. In particular, sex-selective abortion likely increases the observed inter-birth interval between the last birth and the successfully manipulated male birth which follows it. This increase in the length of the observed inter-birth interval is on account of inability to determine sex before 10-12 weeks of pregnancy, time for post-abortion recovery and the time taken till next conception (Pörtner 2020; Ebenstein 2010). More the number of abortions done by a parent to eliminate female fetuses, the longer would be the inter-birth interval. Pre-implantation sex-selecting methods that use assisted reproductive technologies (ART) would also increase the inter-birth interval due to multiple procedural cycles. In Table 1, we summarize the effects of various other sex-composition manipulating methods on observed birth intervals. Since sex-selective abortion and pre-implantation methods are unique in increasing the time to birth for a male child, we will attribute cases of longer birth interval preceding male births to be a consequence of sex-selection.[7]

Thus, any sex-selective method resulting in a live male birth recorded at birth order k would render the interval preceding the k^{th} birth to differ based on the sex of the child born at order k. Considering the temporal ordering of birth interval and the natural (and legal) revelation of the sex of the child at birth (at the end of this interval); for any observed correlation between the length of the interval and the sex of the child born at the end of it, the *direction* of effect is obvious: the sex recorded at

7. One caveat here is our inability to distinguish between instances of prenatal sex-selective abortions from unreported sex-specific post-birth infanticides. While respondents are asked to report all live births and the age of death of any children, it is possible that some respondents may omit reporting births subject to infanticide. The observed inter-birth interval in this case would also be longer and indistinguishable from a case of pre-natal abortions.

Expected Effect of Parental Manipulations on Preceding and Succeeding Birth Intervals

Preceding Interval			*Succeeding Interval*	
Effect on Interval Length	Depends on sex of child born at k?	Parental Behavior at Birth Order k	Effect on Interval Length	Depends on sex of child born at k?
-	-	Fertility Stopping (conceive sooner)	$\downarrow$	Depends on history
$\uparrow$	Yes	Sex-selective Abortion	-	-
$\uparrow$	Yes	Pre-implantation selection	-	-
$\uparrow$	Yes	Sex-selective infanticide (birth unreported) followed	-	-
		by another birth (reported as k^{th} birth)	-	-
-	-	Sex-selective infanticide (birth reported)	-	-
-	-	Sex-selective abortion/infanticide (birth unreported) and no	-	-
-	-	births thereafter	-	-

Table 1: Expected effect of parental sex-composition manipulating methods on reported preceding and succeeding birth intervals for birth at order k. Notes: Unreported births refer to live births subject to infanticide that the respondent doesn't report in her birth history.

the end of the interval is effected by behaviour *during* the time duration recorded as the inter-birth interval.

This is however only true if there is no reversal in the decision to sex-select (Pörtner 2016, 2020). If there were reversals in the decision to sex-select, then we would also observe girls being born after failed sex-selective attempts and hence after long intervals. The net effect on the direction of correlation between sex of a newborn and the preceding interval would then depend on the specific composition of the 'types' of sex-selectors: those who would reverse their decision after failed attempts, and those who wouldn't. Similarly, if parents who resort to sex-selective methods anticipate a long inter-birth interval and therefore begin their attempts at conception sooner than others, then also the net effect on the direction of correlation would be ambiguous. Therefore, correlation between sex and preceding birth interval confirms sex-selection, but an absence in correlation doesn't exactly deny use of sex-selection.

In Section 4 we will exploit this relation between sex and the length of the preceding birth interval — or rather, the lack of it — to identify sex-selecting behavior.

2.3 Identifying sex-manipulating heuristics

That some parents resort to sex-selection is known and evidenced in the literature. What is not known, is *when* parents choose to sex-select, if at all. The sex of a child is determined by nature independent of births at other orders and with slightly unequal odds in favor of males.[8] Knowing this, parents with preferences over sex-

8. Chances of birthing a boy are slightly higher than of birthing a girl, even in the absence of sex-selective practices (Orzack et al. 2015). The World Health Organisation (WHO) expects the

composition could either play these natural odds or manipulate the final composition of their children by using sex-selective methods or differential stopping. Playing the odds, although cost-less, runs the risk of birthing a child of an unwanted sex. Similarly, sex-selective methods, although significantly reducing the chances of birthing a child of an unwanted sex, have some costs associated with them (including illegality and hopefully morality). In the face of this, it is natural to expect that there emerge some rules of thumb in parents' minds to help decide if and when to engage in sex-selection. In order to identify these rules or heuristics that parents follow when engaging in sex-composition manipulating methods, it is important that we are able to tie back parents' actions to their preferences.

To be able account for preferences, we use the minimal characterization of parents having complete and transitive preference ranking over sex-sequences. We will further assume that parents are indifferent between sequences that have the same composition and parity, and so the specific ordering within a given sequence doesn't matter in the preference ranking. This is tantamount to preferences being defined over combinations of n and b; where n is the number of children in a sex-sequence and b is the number of boys in the sex-sequence. For parents who are indifferent about sex-composition, the preference ranking would only into account the value of n.

We then define n^* to be the ideal parity for a parent, which is equal to the total number of children in the most-preferred sex-sequence of the parent. Likewise, we can define b^* to be the ideal number of sons according to a parent, and equal to the total number of boys in the maximal sex-sequence in the parent's preference ranking. n^* is well-defined and unique for each parent, while b^* is undefined for parents who are indifferent about the sex-composition at n^*.[9]

Given these preference markers, we define *relative birth orders* for each child. Let us denote birth order, as defined in the usual sense, with k and the actual parity (or the total number of children born to a parent) with $\overline{n}$. Now consider a parent who desires two children, but birthed three (i.e $n^* = 2, \overline{n} = 3$). Their third child, born at order $k = 3$, is the youngest child relative to actual parity, but is 'one more than the parent's ideal number of children'. It is to capture *this* difference between births at any order k for parents with different values of n^* and $\overline{n}$, that we defined these *relative birth orders*.

natural birth sex-ratio to range between 103 to 107 boys per 100 girls. Based on this estimate, we can consider the natural probability of male births to range anywhere between 0.507 and 0.517.

9. Let P^* be the set of most preferred or maximal combinations of (n, b). We take $n^* = \min\{n|(n, b) \in P^*\}$ and let $B = \{b|(n^*, b) \in P^*\}$. We take $b^* \in B$ iff $|B| = 1$, otherwise, b^* is undefined.

For every birth order k, we define the relative birth order with respect to actual parity to be $j = k - \overline{n}$, and relative birth order with respect to ideal parity to be $j^* = k - n^*$. The third child in our example above is born at $j = 0$, but $j^* = 1$. For a parent who desires and births exactly two children, their youngest child is born at $j = j^* = 0$ and is hence well within their desired parity.

Since we only observe births at birth orders $k \leq \overline{n}$ for any parent; $j = k - \overline{n} \leq 0$ by construction, with older children born at lower values of j. Relative order with respect to ideal parity, given by $j^* = k - n^*$, could however take both negative and positive values. Observing births at relative order $j^* > 0$ is indicative of parents exceeding their ideal parity by at least j^* children.

We will deduce parents' sex-manipulating heuristics by studying the trends in the proportion of sons born at various *relative birth orders*..

2.3.1 Heuristics and Relative Birth Orders: Developing Intuition

Let us consider first, the example of parents who don't have any preferences over sex-composition. It is natural to expect that they would give birth to exactly n^* children and never sex-select. Thus, all children born to these parents at any birth order $k \leq \overline{n} = n^*$ would be born without any parental manipulations. The proportion of sons at relative orders with respect to actual parity (j) and at relative orders with respect to ideal parity (j^*) would then be equal to the natural probability of giving birth to a boy. (see Figure 3).

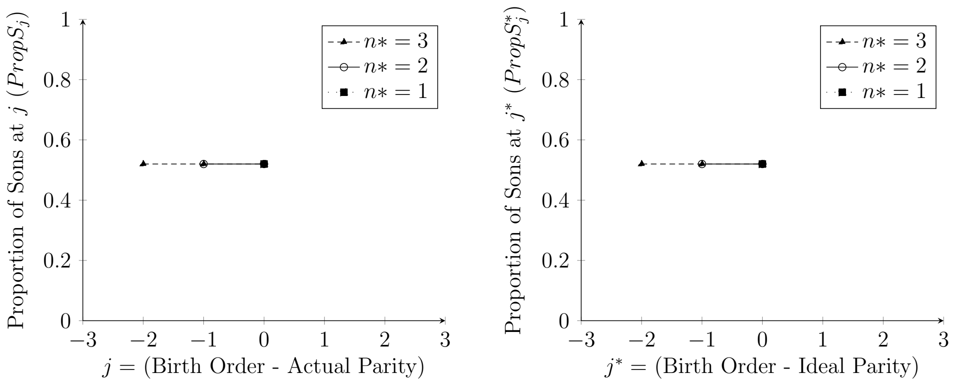

Figure 3: Proportion of sons at relative birth orders in the absence of son-preference and any parental manipulation. (Based on simulations with natural probability of conceiving a boy fixed at 0.52.) n^* denotes parents' ideal parity.

At the other extreme are parents who only want to secure b^* boys no matter how many children it takes. Even though it may sound preposterous, such parents would birth as many children as it takes to birth b^* boys. It is common to note such behavior among families that want at least one son ($b^* = 1$), since probabilistically it shouldn't take too many attempts. Parents following such a heuristic who couldn't naturally birth b^* sons in their first n^* attempts, would necessarily birth $\overline{n} > n^*$ children. For them, their youngest child would always be a boy — their b^{*th} son. For example, if $n^* = 1, b^* = 1$, all children except the youngest child must necessarily be girls, otherwise the parent would have stopped sooner. For $n^* = 2, b^* = 1$, however, parents with a male firstborn could have their youngest child born at $k = 2$ be a daughter. In general, as the relative order with respect to actual parity (denoted by j and represented in the left panel of Figure 4) decreases, the proportion of sons among son-preferring parents must fall, since we would be selecting on a pool of parents with higher $\overline{n}$ (i.e. those who had fewer sons at lower birth orders). In this example, since sex-selection is not resorted to, the proportion of sons at all relative orders with respect to ideal parity (j^*) would still be equal to the natural probability of conceiving boys. This is because sex is determined naturally and independently at each birth order no matter the value of n^* or k. (see Figure 4)

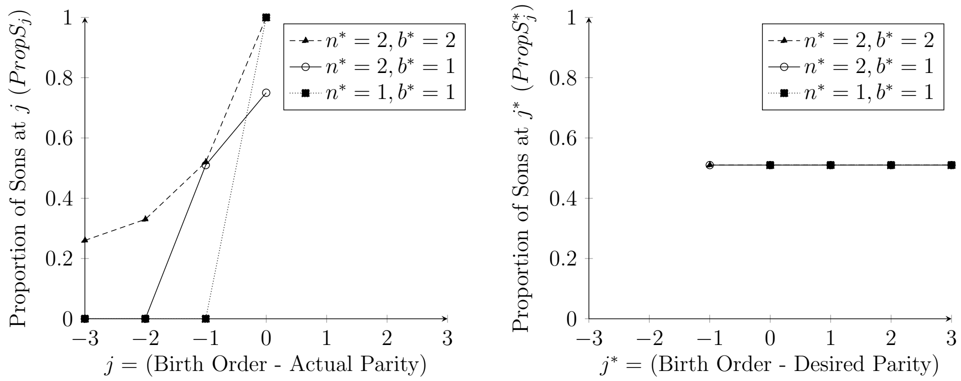

Figure 4: Proportion of sons at relative birth orders for parents who birth as many children as it takes to birth b^* boys. (Based on simulations with natural probability of conceiving a boy fixed at 0.52.) n^* denotes parent's ideal parity and b^* denotes parent's ideal number of boys.

A more realistic case is where parents care about both the number of children and the composition. Depending on the relative importance attributed to the two, one can think of several heuristics by which parents would be willing to tolerate some deviation from n^* (or b^*) in order to not deviate much from b^* (or n^*). For example, parents who are unable to birth b^* boys among their first n^* children, may be willing to extend their parity to n^*+1 just to birth another boy using sex-selecting technology.

If parents followed this heuristic, $\overline{n}$ would always be less than or equal to $n^* + 1$, and there would be no births at birth orders $k \geq n^* + 1$. Or to put differently, there would no no births at relative orders $j^* \geq 2$. Moreover, proportion of sons at relative order $j^* = 1$ would be equal to 1. This is because only those parents who couldn't achieve their desired target of boys by n^* would exceed their desired parity and birth at $k = n^* + 1$ using sex-selection (see Figure 5). All other births at relative orders $j^* \leq 0$ wouldn't be manipulated and hence the probability of male births among them would be equal to the natural probability.

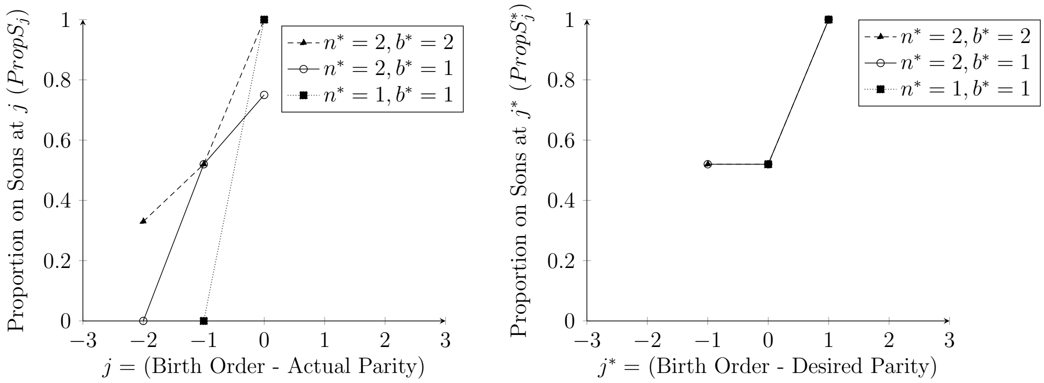

Figure 5: Proportion of sons at relative birth orders for parents who follow the following heuristic: sex-select only the $n^* + 1^{th}$ child if total sons among first n^* are less than b^*. (Based on simulations with natural probability of conceiving a boy fixed at 0.52.) n^* denotes parent's ideal parity and b^* denotes parent's ideal number of boys.

Based on these examples, we can make the following intuitive observations:

1. In the absence of any manipulation, actual parity $\overline{n}$ equates ideal parity n^*, implying no births at relative orders $j^* > 0$. Further, proportion of sons at all relative orders $j = j^* \leq 0$ would be equal to the natural probability of birthing a male child.

2. Differential stopping behavior is sufficient to render the proportion of sons to be increasing with relative birth order j, since $\overline{n}$ becomes endogenous to sex of the last child.

3. In the absence of sex-selection, the proportion of sons at any relative order j^* will always equate the natural probability of conceiving a boy. However, if sex-selection is resorted to at any relative order j^*, we would observe a spike in proportion of sons at j^*.

In Section 4.3, we will generate relative birth orders for $1,030,609$ children in our sample with respect to their mother's actual and desired parity, to compare the proportion of sons across these relative orders. Any heuristic identified from this exercise would then be tested empirically by comparing lengths of preceding birth intervals.

3 Data

We use retrospective birth histories of Indian women aged between 15-49 years as recorded in the National Family Health Survey (2015-16).[10] Respondents report the details of each birth, including the inter-birth interval, birth order and sex.[11] We restrict our sample to $366,224$ women who report having completed their fertility and exclude those who ever had a multiple birth.[12] The birth histories of these mothers list $1,030,609$ children born live between 1970-2016.

The survey also collects other relevant information from respondents relating their age, level of education, household characteristics, contraceptive use, reproductive health as well as their ideal parity and sex-composition. For each respondent, the data also provides estimates of a wealth index — a composite measure of the household's living standard — based on the household's ownership of assets, access to living amenities, type of housing etc. The index is based on a continuous scale and is used in the data to categorize households into quintiles of relative wealth.

To record *ideal* parity and sex-composition, respondents are asked to report the ideal number and composition of children — number of sons, daughters and those of either sex — that *"the respondent would like to have in her whole life, irrespective of the number she already has"*.[13] This raises a question about how reliable these reported 'ideal parity' figures are and how must we interpret them? While it is fair to expect that the reported 'ideal' parity may be *ex-post* rationalized, we find that the correlation between actual and ideal parity isn't very high. Moreover, this correlation is even lower for respondents who have had atleast one instance of a multiple birth (see

10. NFHS-4 — International Institute for Population Sciences - IIPS/India and ICF. 2017

11. For all births reported, the survey confirms if the child is alive and if not, the age at death. In case the respondent fails to report births that were subject to infanticide, it would be difficult to separate instances of unreported infanticide from sex-selective abortions.

12. Respondents are asked if they would like to have another child; to which they reply if they would, wouldn't, haven't decided, or cannot because they are either sterilized or declared infecund. All women who report not wanting any more children or not being able to bear any more children are classified as having *completed* their fertility. A multiple birth is one where a mother delivers two or more children i.e. twins, triplets etc.

13. DHS Methodology, Standard Recode Manual

Table A1 for details), indicating that reported 'ideal' parity doesn't exactly tail actual parity.[14] In the context of our conceptual framework, we interpret the 'ideal' parity to be equal to total number of children in the respondent's *most preferred* sex-sequence of births.

Now it is possible that respondents' preference rankings aren't fixed and vary over time with birth experiences and history. Since we restrict our sample to women who complete their fertility, it is reasonable to believe that they report their ideal parity and composition based on the most recent preference ranking that would have been relevant for fertility choices made at the time of their last birth (and hence also relevant to their decision to complete fertility and stop childbearing). It is also likely that the preferences that guide fertility choices may not even be that of the respondent alone, but influenced by the preferences of her partner and other family members. Again, considering that respondents in our sample have completed their fertility, we expect them to at least be aware of their partner's fertility preferences. Roughly 84% of women in our sample believe that their partner "wants the same number of children[...]that she wants herself".[15]. Whether this alignment in fertility preferences between the two partners exists already or occurs after bargaining or deliberation is not a matter of concern for our analyses since we only concern ourselves with the need to capture preferences that ultimately influence fertility decisions.

4 Findings and Inferences

We first begin with crude estimates of sex-selectors and fertility stoppers, that we calculate based on the proportion of parents who continue childbearing after a given history and the proportion of those who birth a male child in particular (as described under Section 2.2.1). Table 2 reports estimates for the proportion of fertility stoppers and sex-selectors for a few combinations of parameter values assumed for π, the natural probability of conceiving a male child, and for j, the maximum number of sex-selective abortions a parent is willing to undergo.

The estimates are fairly consistent for the different parameter values. While proportions are ideally supposed to be between 0 and 1, our estimates — based on crude calculations under certain assumptions — sometimes take negative values when they are close to zero. Our objective here is not to take these estimates at face value, but to make relative comparisons across histories to confirm our intuitions about when

14. Jayachandran and Kuziemko (2011) also discuss some concerns that the 'ideal' parity measure may raise.

15. DHS Methodology, Standard Recode Manual

parents are most likely to resort to manipulating the sex-composition of their children.

	Proportion of parents who stop childbearing and sex-select					
Parameter Values	$\pi = 0.512$, $j = 3$		$\pi = 0.512$, $j = 1$		$\pi = 0.500$, $j = 3$	
Birth History	*Stoppers*	*Sex-selectors*	*Stoppers*	*Sex-selectors*	*Stoppers*	*Sex-selectors*
-		0.07		0.08		0.09
M	**0.11**	0.01	**0.11**	0.01	**0.11**	0.03
F	0.06	**0.11**	0.04	**0.14**	0.06	**0.14**
MM	**0.48**	-0.02	**0.48**	-0.02	**0.48**	0.01
MF	0.42	0.04	0.42	0.05	0.42	0.06
FM	0.44	0.05	0.43	0.06	0.43	0.08
FF	0.19	**0.15**	0.17	**0.18**	0.19	**0.18**
MMM	0.50	-0.02	0.50	-0.02	0.50	0.01
MMF	0.55	-0.02	0.56	-0.02	0.55	0.01
MFM	**0.59**	0.00	**0.58**	0.01	**0.58**	0.03
FMM	**0.60**	0.00	**0.60**	0.00	**0.60**	0.02
FFM	0.50	0.05	0.49	0.07	0.50	0.08
FMF	0.40	0.05	0.39	0.06	0.39	0.07
MFF	0.39	0.06	0.38	0.07	0.39	0.08
FFF	0.21	**0.15**	0.18	**0.18**	0.20	**0.18**

Note: Since these estimates are based on crude estimators, select values close to 0 *are negative.*

Table 2: Crude estimates of sex-selectors and fertility stoppers following select birth histories. Estimators used to calculate the proportion of fertility stoppers ($\hat{\mu_h}$) and the proportion of sex-selectors ($\hat{\lambda_h}$) are described in Section 2.2.1. π refers to the natural probability of conceiving male child and j refers to maximum number of abortions sex-selectors are willing to undergo. The highest proportion of sex-selectors and fertility stoppers at each birth order are highlighted in bold.

The estimates in Table 2 suggest that parents are least likely to stop childbearing and most likely to sex-select following histories with only female births i.e. F, FF, FF. Moreover, parents who complete their fertility at birth orders 1, 2 and 3, are most likely to do so following histories M, MM and MFM/FMM respectively. Our crude estimates for the proportion of sex-selectors are not far from previously reported estimates of 11.6% of second and 15.3% of third order births in 1995-2005 being subject to prenatal sex diagnosis (Bhalotra and Cochrane 2010).

These estimates are 'crude' because they follow from our assumption that the natural probability of birthing male children denoted by π is fixed and known, and can be used as a benchmark to compare the sample proportion of male births against. A choice of a lower value for π, for example, would increase the estimate of sex-selectors in the sample. However, a relative comparison of these estimates following various histories is still meaningful.

4.1 Evidence for Differential Fertility Stopping

Clearly, not all son-preferring parents are sex-selectors. The large differences in the proportion of fertility stoppers across histories varying in sex-composition is highly indicative of differential fertility stopping behavior (Table 2). That is, parents with less desirable compositions likely have another child in the hopes of birthing a boy. Figure 6 illustrates this.

Had fertility decisions been independent of sex-composition, then at any given level of total parity, the proportion of sons among all children born up to any birth order would be identical for all parities and equal to the natural probability of birthing sons. But the pattern in Figure 6, is one where the plots for parents with lower parity are stacked above those for higher parity. This means that parents at high parities tend to have few sons. Hence, it must be that parents who have few or no sons are the ones who *select into higher parity* (Clark 2000). Thus, parents who have few or no sons are most likely to continue to bear more children in the hopes of birthing a boy so as to achieve a composition that they rank higher than their current composition (albeit with higher parity).

Parents cannot however, engage in continued childbearing to achieve their desirable composition (due to childbearing and child-rearing costs). This is clear from the fall in proportion of parents with parity beyond three children (Figure 6). As described before, it is more realistic to think of son-preferring parents facing a trade off between deviations from their ideal parity and deviations from their ideal composition. Those who cannot tolerate deviations from their ideal parity have higher odds of giving up or resorting to sex-selection. Based on our estimates in Table 2, more than one-tenth of parents are sex-selecting at each birth order. To confirm the usage of sex-selective methods at these birth orders, we use another approach — one that compares preceding birth intervals.

4.2 Empirical Evidence for Sex-Selection

Considering that the natural and legal revelation of sex of a newborn can only occur *after* the child is born, that is, at the end of the birth interval preceding it; ideally, the sex of the newborn should be random and uncorrelated with the length of the

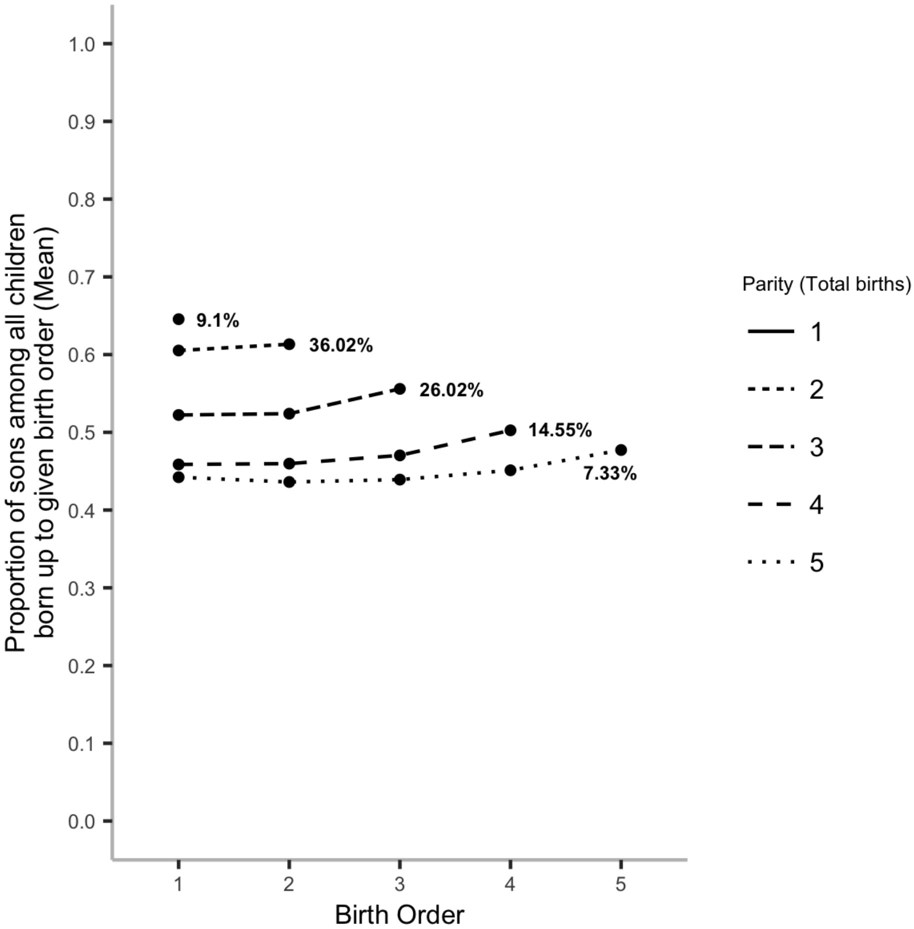

Figure 6: Evidence for parents selecting into higher parity. The figure plots the cumulative proportion of sons among children born up to various birth orders, for mothers with various total parities. Labels report percentage of mothers in our sample with the given total parities.

preceding birth interval. Sex-selective abortions, on the other hand, increase the observed interval between consecutive live births. Thus, if there is any sex-selective intervention, we would expect the birth interval preceding male births to be longer.[16] Note though, that in case of any reversal or take-back in the decision to sex-select, we

16. While this is true for post-implantation sex-selective methods such as abortion, it may not necessarily be true for pre-implantation sex-selective methods. Consider for example, parents who embryo-select using IVF to sex-select their succeeding child. They could anticipate a longer birth interval and try conceiving sooner. We don't anticipate this to be a major concern for our exercise since IVF technologies increase the chances of multiple births, ones which we exclude from our sample. Further, pre-implantation selection methods are generally expensive and not affordable for a large proportion of the population we are studying.

would also observe girls being born at the end of long intervals, thereby rendering the net effect on the direction of correlation between sex and the length of the preceding interval as ambiguous. In any case, a correlation between the sex of a newborn and the length of the birth interval preceding it, definitely suggests manipulation.

Figures 7 and 8 plot the proportion of male births conditional on the length of the preceding interval, for children born at the second and third birth orders, respectively. The Figures have separate subplots for births following different histories. If sex at consecutive births were determined independently and there were no sex-selecting interventions, we would expect these plots to be identical for all histories and horizontal (at the value π). Even if the relation between probability of birth a boy and the length of the preceding interval were non-linear, as long as sex of the consecutive child is determined independently, we would expect the plots to be identical for different histories.

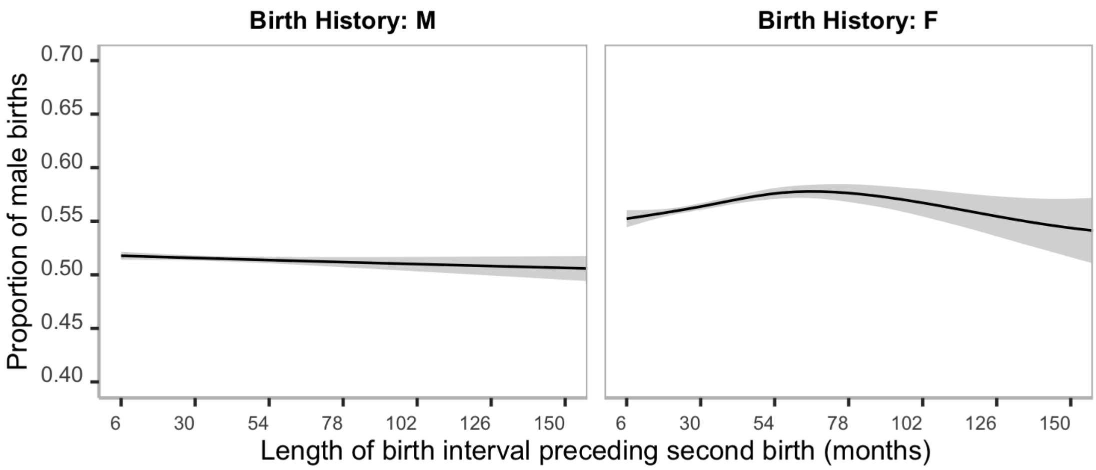

Figure 7: (Smoothed) Proportion of male births at the **second** birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories.

Our observations suggest that mothers with a firstborn girl are disproportionately more likely to give birth to sons at the second order, especially after long birth-intervals. This trend however reverses after the inter-birth interval exceeds 5-7 years. Hence, given birth history `F`, some parents engage in sex-selection, with many of these parents taking-back their decision to sex-select as the inter-birth interval gets

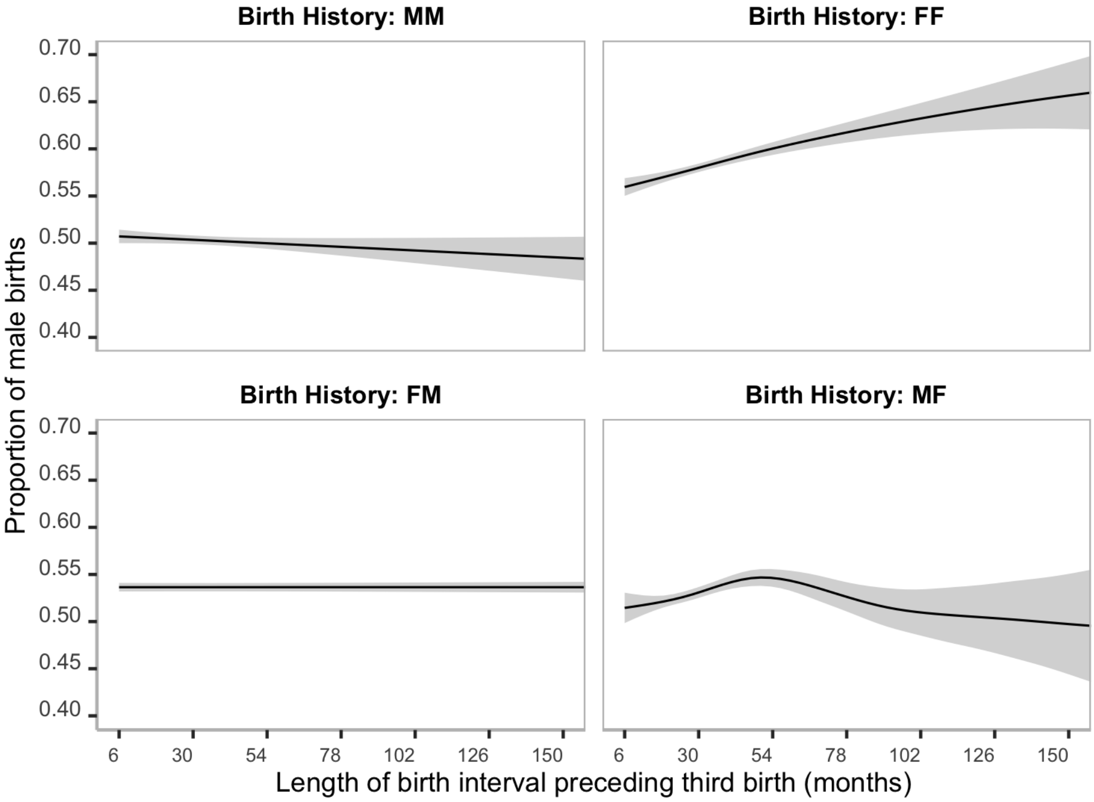

Figure 8: (Smoothed) Proportion of male births at the **third** birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories.

'too long'. A similar hump-shaped curve, albeit with a smaller hump, is also observed following history `MF` at the third birth order. Following `FF` however, the proportion of male newborns increases steadily with the length of the preceding interval with no indications of any reversal in the decision.

At the third birth order, studying histories `MF` and `FM` separately illustrates that parental incentives to manipulate may not be based on composition alone. Possible reasons why parental incentives to sex-select following `MF` may differ from those following `FM` are: (1) recency of male birth in `FM` may discourage parents from sex-selecting immediately after, or make them mistakenly optimistic about birthing sons in their next attempt; or alternatively, (2) the sequence `FM` may have itself been achieved by sex-selecting at the second order, and parents might be unwilling to sex-select at consecutive orders. That parental incentives do not depend on composition alone, is an important take-away.[17]

Figure A1 plots the probability of male births at the fourth birth order conditional on the length of the preceding birth interval. Consistent with our crude estimates, we find evidence of sex-selection following history `FFF`. We also note a tendency of girls being born after long intervals following histories such as `FMM`, `MMF` and `MMM`.

A comparison by geographical region — also synonymous with cultural divisions in India — echoes the well documented regional disparity in sex-ratios and gender relations in India (Dyson and Moore 1983). North-western states including Rajasthan, Punjab and Haryana, along with states in Western and Central India, such as Gujarat, exhibit strong non-reversing sex-selecting behavior following histories `F` and `FF`; as is also reflected in the low female-to-male ratios in these regions (see Figures A2 and A3). In contrast are the Southern and North-Eastern states — typically with more balanced sex-ratios and higher agency for women — that display, if anything, a marginally decreasing trend in proportion of sons.

Unlike our crude estimates of sex-selectors in Table 2 which are based on the total proportion of sons born after any given history; studying how proportion of male births varies with the length of the preceding interval gives insights into the dynamic nature of fertility choices made between two live births. Here, we can observe that following histories such as `F`, `FM` and `FFF`, parents also tend to take-back their decision to sex-select. Whether this is driven by conscience or by a conscious choice to not space children too far apart is hard to comment on.

To support these inferences with an empirically sound comparison, we employ a difference-in-difference specification to account for biases from any omitted variables. For each birth order k, we compare correlations between the sex of the child and the interval preceding its birth (denoted by '$preI_k$'), for various birth histories $h \in H_k$. Here, H_k denotes the set of all sex-sequences until birth order k. For example, for births at the third birth order: $k = 3$ and $H_k = \{\texttt{MM}, \texttt{FF}, \texttt{MF}, \texttt{FM}\}$.

17. Note that here we are discussing *actions* of parents following specific birth histories and this is why the specific ordering of births matters. This is not the same as parents' *preferences* being dependant on the specific ordering within a sex sequence.

We estimate the following specification:

$$\begin{aligned} preI_{mk} = const &+ \sum_{h \in H_k \setminus \{h_r\}} \alpha_h \{\text{hist}_{mk}\text{:h}\} + \beta\{\text{sex}_{mk}\text{:M}\} \\ &+ \sum_{h \in H_k \setminus \{h_r\}} \gamma_h \{\text{hist}_{mk}\text{:h} \times \text{sex}_{mk}\text{:M}\} \\ &+ \theta' X_m + u_{mk} \end{aligned}$$

where

m, k : indices for mother, birth order, respectively
$preI_{mk}$: length of birth interval preceding k^{th} birth for mother m
h : birth history until order $k-1$
H_k : set of all possible birth histories until order $k-1$
h_r : birth history with $k-1$ male children
'hist_{mk}:h' $= 1$ if birth history until order $k-1$ for mother m is h, 0 otherwise
'sex_{mk}:M' $= 1$ if k^{th} child of mother m is male, 0 otherwise
X_m : vector of covariates such as mother's age at $k-1^{th}$ birth, wealth quintile, urban/rural, etc.

Our difference-in-differences specification estimates the difference between the length of the birth interval preceding male births and that preceding female births for a given history h, and then differences this difference relative to that following the reference history h_r. The identification here is based on the assumption that *even if* there are any unaccounted unobservables such as geography, ethnicity, socio-economic or environmental factors, that could likely skew the natural probability of birthing a son *and* also influence the length of the preceding inter-birth interval; they should confound the relationship between sex and preceding interval, if at all, for *any* birth history.[18] Since the incentives to sex-select vary with birth history but the effect of any confounding variables shouldn't, any significant difference-in-difference in length of the preceding interval by the sex of the newborn, must be attributed to parental manipulation.

Our coefficients of interest, therefore, are γ_h — the coefficient of interaction between indicator for history h and indicator for male birth i.e. 'hist:h $\times$ sex:`M`' — for all histories $h \in H_k \setminus \{h_r\}$ at any given birth order k.

18. This is akin to a Parallel Trends Assumption in standard Difference-in-Differences models.

Table 3 reports the difference-in-differences estimates for our sample at the second and third birth order. Specifications (2) and (4) include controls such as maternal age at previous birth, wealth quintile of the household, and urban/rural place of residence. Our estimates for the coefficient(s) of interest are significant and indicate that boys are typically birthed after longer birth intervals following birth histories `F, FF` and `MF` (relative to reference histories), confirming sex-selection at the second and third birth orders. These findings are robust and not driven by the tails of the distribution of the length of preceding birth intervals (see Table A2).

Table A3 reports the difference-in-difference estimates for the fourth, fifth and sixth birth orders. We don't report estimates for higher birth orders because roughly 97% of our sample has fewer than 6 children. We find evidence of sex-selection following histories `FFF` and `FFFF`, as well as for `MFFF`. It is important to exercise caution when interpreting results for specific histories at high birth orders because the sample size driving these results falls dramatically. Moreover, as we analyse higher birth orders ($k \geq 4$), our sub-sample selects on parents who are willing to have atleast k children. Thus, as k increases, we would be selecting on parents who either have a large ideal parity, or for whom exceeding their ideal parity is probably not much of a concern.

Whether to sex-select and when, are not trivial decisions. Aside from needing access and the means to use sex-selecting technologies, it also involves continually comparing two options against each other: that of playing the natural odds of sex-determination at no cost, versus reducing the odds of a female birth at some cost (monetary, legal and psychological). It is unlikely that parents take on as much cognitive load to choose their optimal and most efficient decision to this dynamic problem. Rather, it is quite possible that parents rely on simple rules of thumb or heuristics to decide when and if it is best for them to sex-select. To identify such a heuristic that summarizes these choices for parents with different parity preferences, we look for patterns of sex-selective behavior across relative birth orders.

4.3 Heuristics of Parental Manipulations

We generated relative birth orders for $1,030,069$ children in our sample based on their mother's reported total parity and ideal parity. Figure 9 plots the proportion of sons among children born at various relative birth orders in our sample. The panel on the left plots these proportions for relative orders defined with respect to actual parity, and the panel on the right plots the same against relative orders defined with respect to ideal parity. The grayscale gradient in the color of the markers represents the share of births in the sample, with darker shades representing larger shares.

	Dependent variable: Birth Interval preceding k^{th} birth			
	$k = 2$		$k = 3$	
	(1)	(2)	(3)	(4)
sex:M	-0.199^{**}	-0.186^{*}	-0.368^{**}	-0.323^{*}
	(0.099)	(0.099)	(0.184)	(0.184)
hist:F	-1.450^{***}	-1.393^{***}		
	(0.107)	(0.107)		
hist:F × sex:M	0.560^{***}	0.455^{***}		
	(0.145)	(0.145)		
hist:FF			-1.415^{***}	-1.404^{***}
			(0.187)	(0.187)
hist:FM			-0.840^{***}	-0.822^{***}
			(0.186)	(0.186)
hist:MF			-1.266^{***}	-1.254^{***}
			(0.185)	(0.185)
hist:FF × sex:M			1.644^{***}	1.418^{***}
			(0.254)	(0.254)
hist:FM × sex:M			0.367	0.295
			(0.259)	(0.258)
hist:MF × sex:M			0.590^{**}	0.531^{**}
			(0.258)	(0.257)
Constant	34.985^{***}	35.820^{***}	34.779^{***}	32.451^{***}
	(0.071)	(0.243)	(0.130)	(0.334)
Controls		Yes		Yes
Observations	332,656	332,656	200,824	200,824
R^2	0.001	0.005	0.001	0.005
F Statistic	88.6^{***}	186.6^{***}	16.7^{***}	71.3^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01	

Table 3: Difference-in-Difference estimates for second and third birth orders. Specifications (1) and (2) report estimates at the second birth order while (3) and (4) do the same for the third. Specifications (2) and (4) include controls such as maternal age at the previous birth, wealth quintile and place of residence (rural/urban).

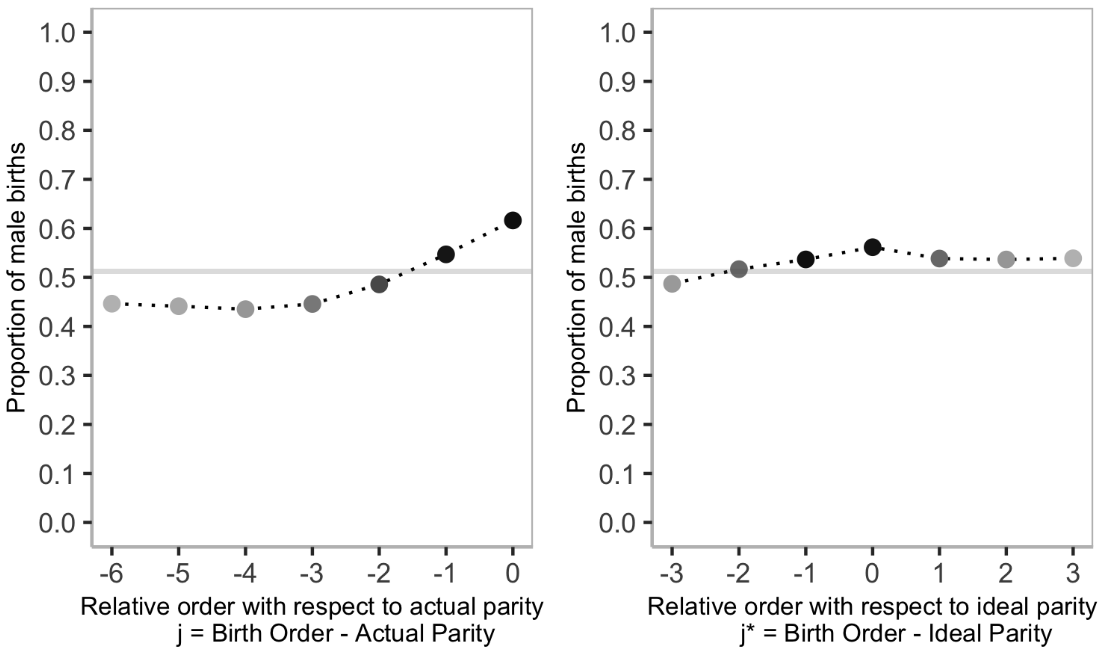

Figure 9: Proportion of male births at relative birth orders. Left panel plots proportion of sons born at various relative orders with respect to actual parity i.e birth order - actual parity. Right panel plots proportion of sons born at various relative orders with respect to ideal parity i.e birth order - ideal parity. The grayscale gradient in the color of the markers indicates the share of births at the given relative order in the sample, with darker shades representing larger shares.

The left panel in Figure 9 shows that proportion of sons is increasing in relative order with respect to actual parity. This means that younger children in any family, no matter the family size, are more likely to be male.[19] The right panel plots the proportion of sons among children born at various relative orders with respect to ideal parity. Parity preferences should, in theory, be orthogonal to natural outcomes such as sex and birth history. However, a spike in the proportion of sons among children born at $j^* = 0$ i.e. those born at a parent's ideal parity, can be noted in the right panel.[20] This suggests that parents who resort to sex-selective methods, typically do so at their n^{*th} birth order, perhaps to avoid exceeding n^*, their ideal parity.

For a more detailed break-down, Figure 10 plots the proportion of sons among children born at the second, third and fourth birth orders separately, for mothers with different ideal parity. As before, we can note that the proportion of male births at a given birth order is high for mothers whose ideal parity matches the birth order.

19. Table A4 reports the proportion of sons at various relative birth orders by mother's total parity.
20. Table A5 reports the proportion of sons at various relative birth orders by mother's ideal parity.

Looking at these proportions separately for parents with different birth histories confirms that the increase in proportion of male births is only for parents who have birthed either few or no sons and hence have an incentive to sex-select. The exception to this rule are mothers who have already exceeded their ideal parity and have birthed no sons at all i.e. those with histories `FF, FFF`, who sex-select at third and fourth birth orders respectively after exceeding their ideal parity. We will discuss this case of differential fertility stoppers - turned - sex-selectors under the next sub-section.

We thus hypothesize a heuristic whereby parents resort to sex-selection when they are at their ideal parity. This means that a parent with an ideal parity of n^* who wants to sex-select, most likely does so at birth order $k = n^*$ (i.e. at $j^* = 0$). For example, under this hypothesis, a parent who desires exactly two children will sex-select, if needed, at the second birth order.
We validate this heuristic of sex-selecting at ideal parity by confirming that those who sex-select at second and third birth orders are indeed those with an ideal parity of two and three, respectively. Table 4 reports the estimates of the difference-in-differences specification for two sub-samples: those who the heuristic suggests would sex-select (i.e. those with $n^* = k$), and a placebo sub-sample which according to our heuristic shouldn't be sex-selecting at the given order (i.e. those with $n^* > k$). We find the difference-in-difference estimates to be statistically significant following histories `F` and `FF`, only for the sub-sample that our heuristic suggests would sex-select.

Since well-educated mothers and mothers from high-income households are more likely to desire a smaller parity, earlier findings of sex-selection among the educated could be reconciled with their low ideal parity (Bhalotra and Cochrane 2010; Pörtner 2020, 2016). Among mothers with the same ideal parity, we find no noticeable difference in sex-manipulating behavior based on levels of maternal education and household wealth (see Tables A6 and A7).

4.4 Fertility Stoppers who later Sex-Select

In Figure 10 we can see that at the third birth order, proportion of male births is high not only for parents with an ideal parity of two children (except those with history `MM`), but also for parents with an ideal parity of one and with the history `FF`. This is also true at the fourth birth order, for parents with history `FFF` and an ideal parity of two or three. This suggests that parents who use differential fertility stopping and exceed their ideal parity may also sex-select at higher birth orders if they still haven't given birth to any sons. Figure 11 shows the Box-Whisker plot for birth intervals preceding male births at various relative orders with respect to ideal parity. The panels graph these plots for parents who have no sons, parents who have one son and

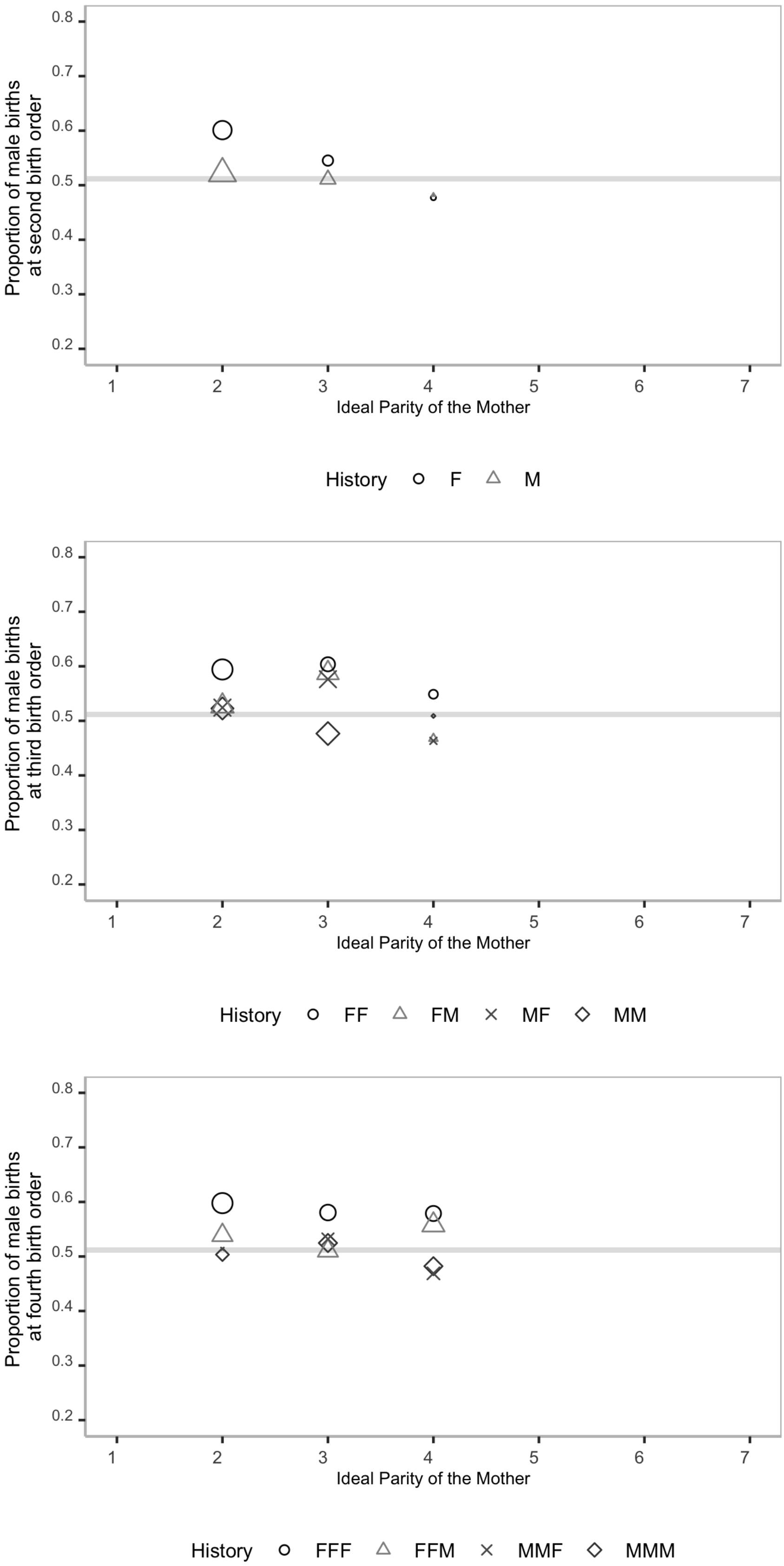

Figure 10: Proportion of male births at second, third and fourth birth orders, following select birth histories, to mothers with different ideal parity. The size of the markers represents the share of mothers in the sample with a given ideal parity.

	Dependent variable: Birth Interval preceding k^{th} birth			
	$k = 2$		$k = 3$	
	Heuristic	Placebo	Heuristic	Placebo
	$n^* = 2$	$n^* > 2$	$n^* = 3$	$n^* > 3$
	(1)	(2)	(3)	(4)
sex:M	-0.419^{***}	0.094	-0.341	0.269
	(0.136)	(0.147)	(0.299)	(0.352)
hist:F	-1.078^{***}	-1.263^{***}		
	(0.152)	(0.151)		
hist:F×:sex:M	0.525^{***}	-0.280		
	(0.203)	(0.211)		
hist:FF			-1.618^{***}	-1.859^{***}
			(0.329)	(0.350)
hist:FM			-0.027	-0.941^{***}
			(0.312)	(0.347)
hist:MF			-0.964^{***}	-0.990^{***}
			(0.307)	(0.347)
hist:FF×:sex:M			1.441^{***}	-0.148
			(0.446)	(0.482)
hist:FM×:sex:M			-0.437	-0.451
			(0.427)	(0.494)
hist:MF×:sex:M			0.228	-0.00001
			(0.423)	(0.493)
Constant	36.653^{***}	36.959^{***}	30.742^{***}	33.108^{***}
	(0.346)	(0.355)	(0.572)	(0.625)
Observations	187,513	124,459	66,843	47,058
R^2	0.006	0.003	0.004	0.002
F Statistic	119.374^{***}	47.917^{***}	22.671^{***}	7.279^{***}
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4: Difference-in-difference estimates at second and third birth orders for mothers who satisfy the heuristic at the birth order i.e. those whose ideal parity is equal to the given birth order ($n^* = k$) and a placebo sample. Specifications (1) and (3) report estimates for the sample of mothers who according to our heuristic are expected to sex-select, while (2) and (4) do the same for a placebo sample.

parents who have two sons, respectively.

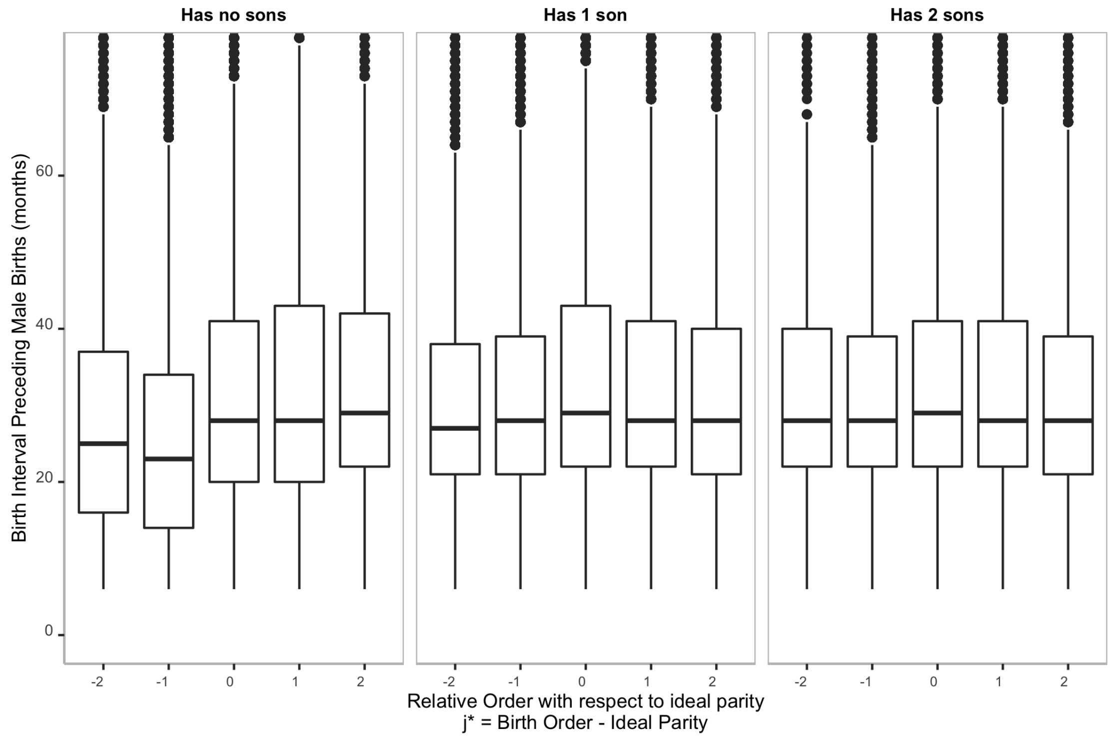

Figure 11: Box-plot of length of birth interval preceding male births at various relative orders with respect to ideal parity. Each panel plots the box-plot for parents who have already birthed given number of sons.

According to our heuristic for sex-selection, male births (via parental manipulations) at the ideal parity ($j^* = 0$) would be preceded by long birth intervals. We observe exactly this in Figure 11. At the same time, in line with our inferences from Figure 10, we also observe long birth intervals preceding male births at $j^* \in 1, 2$ for parents who don't have any sons yet.

We therefore conclude the following: Those who do not engage in sex-selection at their ideal parity, exceed their ideal parity by bearing more children. If at these higher parities they still don't have any male offspring, they also may resort to sex-selection. There is however, no consistent rule of thumb that these differential stoppers - turned - sex-selector parents follow.

4.5 Sex-Selection at the First Birth Order

So far we only discussed sex-selective behavior at birth orders greater than one. This was because there are no histories before the first birth to be able to exploit our difference-in-difference identification strategy. Moreover, the time interval between marital union and the first birth is conceptually very different from inter-birth intervals. Previous research that relied on sex of the firstborn being random for its analyses, defended this assumption based on sex-ratios at the first order being within a 'normal' range. Contrary to these claims, Figure 12 shows that mothers with an ideal parity of one are disproportionately more likely to birth a male child at the first birth order.

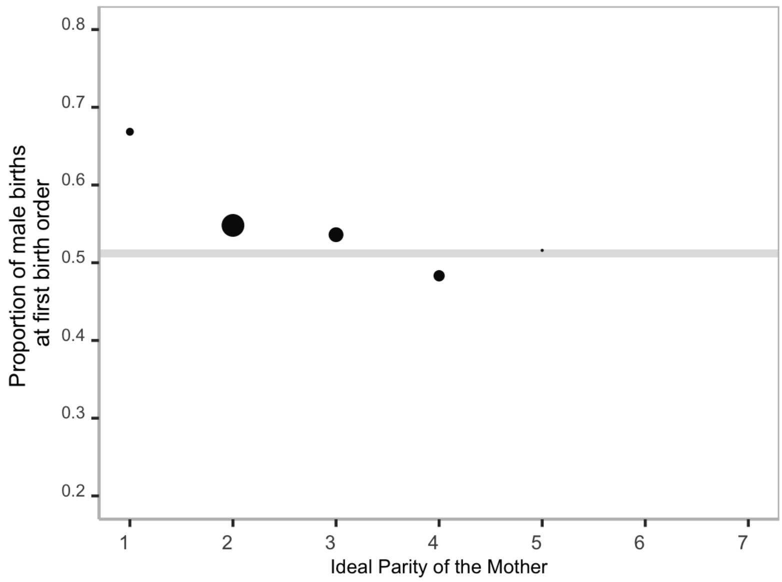

Figure 12: Proportion of male births at first birth orders to mothers with different ideal parity. The size of the markers represents the share of mothers in the sample with the given ideal parity.

To detect sex-selective behaviour at the first birth order we therefore utilise the heuristic that we deduced. Specifically, according to our heuristic we should expect only those parents to sex-select at the first birth order whose ideal parity is a single male child. Hence, we estimate the following specification:

$$I_m = const + \alpha\{n_m^* = 1\} + \beta\{\text{sex:M}\} + \gamma\{n_m^* = 1\} \times \{\text{sex:M}\} + \theta' X_m + u_m$$

where

$$
\begin{aligned}
m &: \text{index for mother} \\
I_m &: \text{interval between marital union and first birth for } m \\
n_m^* &: \text{ideal parity as reported by } m \\
\text{sex:M} &= 1 \text{ if first child is male, 0 otherwise} \\
X_m &: \text{vector of covariates such as mother's age at union, wealth quintile,etc.}
\end{aligned}
$$

We're simply comparing the length of the interval between marriage and first birth for mothers with firstborn boys versus mothers with firstborn girls, and then comparing this difference for a sub-sample that our heuristic predicts will sex-select with one that won't. Table 5 reports the estimates for our specification.

According to our findings, there is significant correlation between the sex of the first child and the length of the time interval preceding the first birth for mothers with an ideal parity of one (relative to others). No such correlation is observed for the placebo group of mothers with an ideal parity of two or three. The negative coefficient for the interaction term in Specification (1) suggests that there is likely some reversal in the decision to sex-select after failed attempts. We confirm that some parents do reverse their decision to sex-select by noting a hump shaped curve in the proportion of male births against the length of time interval between marital union and first birth (see Figure A4).

5 Conclusion

The prevalence of son-preference in India, as well as the practise of differential stopping and pre-natal sex-selection, are well documented in the literature. This paper makes a significant contribution to this literature by characterising parental sex-manipulation strategies. In particular, we find evidence that son-preferring parents who don't have their desired number of boys, resort to sex-selection when they are at their ideal parity.

We overcome two major challenges in identifying this heuristic: not being able to observe parental fertility decisions and not being able to tie these decisions with parents' preferences. To tackle the first challenge, we exploit the non-relation between the sex of a newborn and the length of the birth interval preceding its birth to be able to identify use of pre-natal sex-selection. To account for any potential bias from unobserved factors, we adopt a difference-in-difference specification that compares the correlation between sex and preceding interval for parents with different birth

	Dependent variable: Interval preceding first birth	
	Heuristic	Placebo
	(1)	(2)
sex:M	-0.224^{***}	0.141
	(0.087)	(0.180)
Heuristic: $\mathbb{1}\{n^* = 1\}$	6.620^{***}	
	(0.342)	
sex:M $\times$ Heuristic: $\mathbb{1}\{n^* = 1\}$	-1.123^{***}	
	(0.417)	
Placebo: $\mathbb{1}\{n^* \in \{2, 3\}\}$		-1.141^{***}
		(0.150)
sex:M $\times$ Placebo: $\mathbb{1}\{n^* \in \{2, 3\}\}$		-0.333
		(0.204)
Constant	68.871^{***}	69.154^{***}
	(0.246)	(0.269)
Controls	Yes	Yes
Observations	322,727	322,727
R^2	0.103	0.101
F Statistic	$4{,}109.503^{***}$	$4{,}021.981^{***}$
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 5: Difference-in-Difference estimates for first birth order. The coefficient of the interaction term in (1) captures the difference in length of interval preceding first birth by the sex of the newborn for mothers with ideal parity one ($n^* = 1$), relative to others. Specification (2) estimates the same for a placebo sub-sample of mothers with ideal parity equal to two or three.

histories. Since the incentives to sex-select vary with birth history but the effect of omitted variables don't, any significant difference in correlation in length of the preceding interval by the sex of the newborn (relative to a reference history), must be attributed to parental manipulation.

Using retrospective data on birth histories of a representative sample of 366,224 Indian women, we not only find evidence of sex-selection, but also note a tendency among mothers to reverse their decision to sex-select following unsuccessful attempts.

In order to identify a heuristic that describes *when* parents choose to sex-selection, we used a novel approach of studying probability of male births at *relative birth orders*. We re-describe the order of births in our sample relative to the parent's reported ideal and actual parity. This allows us to compare birth outcomes across families of different size and also resolves the issue of endogeneity of male-biased sex-ratio at the last birth order precluding identification of sex-selection. We observe a significant increase in the proportion of male births among parents with fewer sons when they are at their ideal parity, suggestive of a tendency to resort to sex-selection to avoid exceeding one's ideal parity. Identification of this heuristic also makes way in finding evidence for sex-selection at the first birth order. Those that don't sex-select at their ideal parity, continue childbearing based on differential stopping behavior. If at higher parities they are still unable to birth a male child, then they consider sex-selection.

The goal of this paper has been to improve our understanding of why and when parents sex-select. This is pertinent in light of policy initiatives that have mostly failed to address the issue of missing women and the use of sex-selective methods; and especially so now since as we experience a declining trend in desired parity.

Appendix

	Dependent variable: Ideal parity	
Total parity	0.351***	0.356***
	(0.001)	(0.001)
Total parity $\times \mathbb{1}$(multiple births)		-0.052^{***}
		(0.002)
Constant	1.340***	1.330***
	(0.003)	(0.003)
Observations	371,869	371,869
R^2	0.255	0.256
F Statistic	127,209.500***	63,994.910***
Note:		*p<0.1; **p<0.05; ***p<0.01

Table A1: Correlation between ideal and actual total parity. Multiple births refers to pregnancies with more than one baby i.e. twins, triplets, etc.

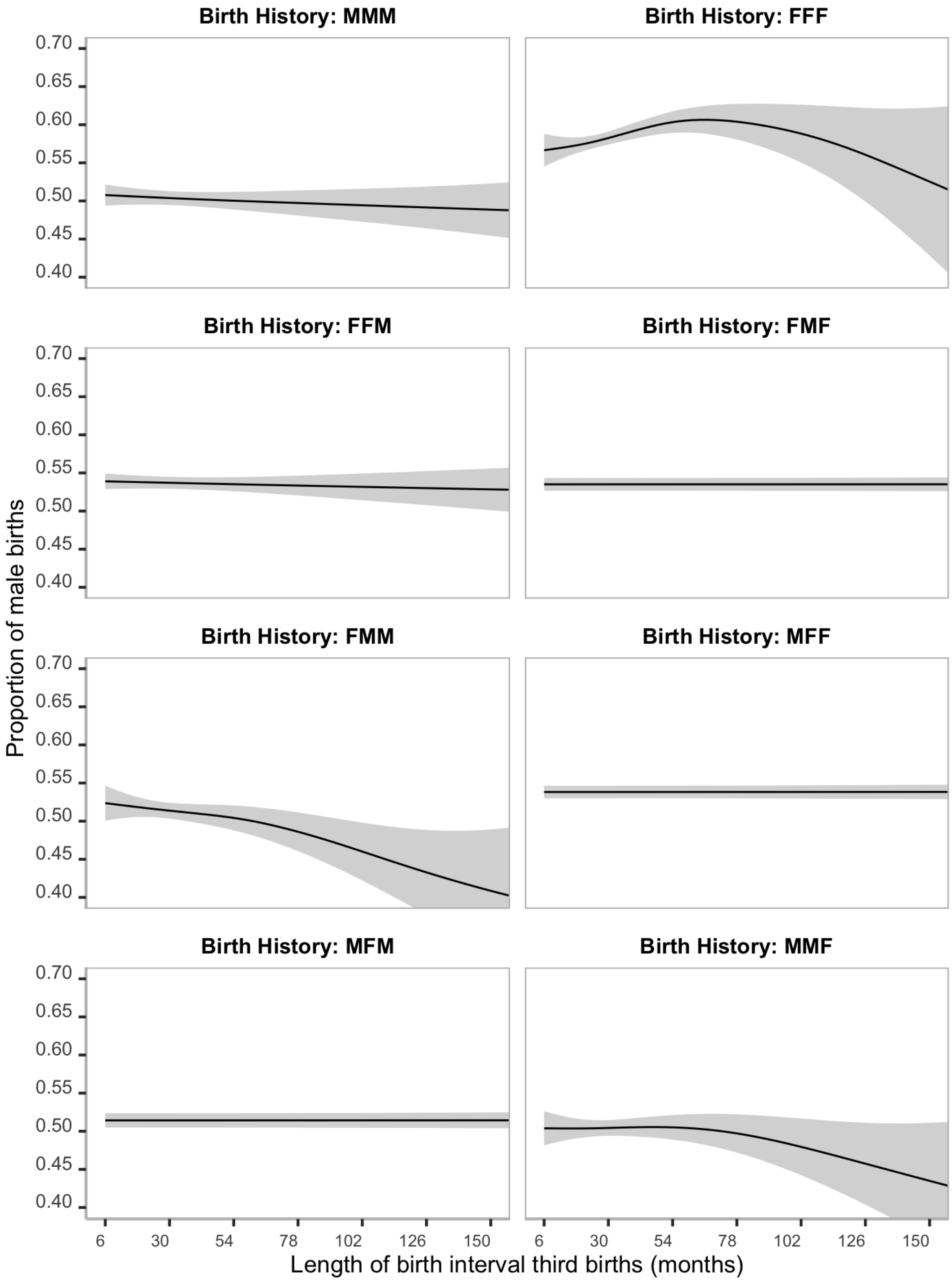

Figure A1: (Smoothed) Proportion of male births at the **fourth** birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions following different birth histories.

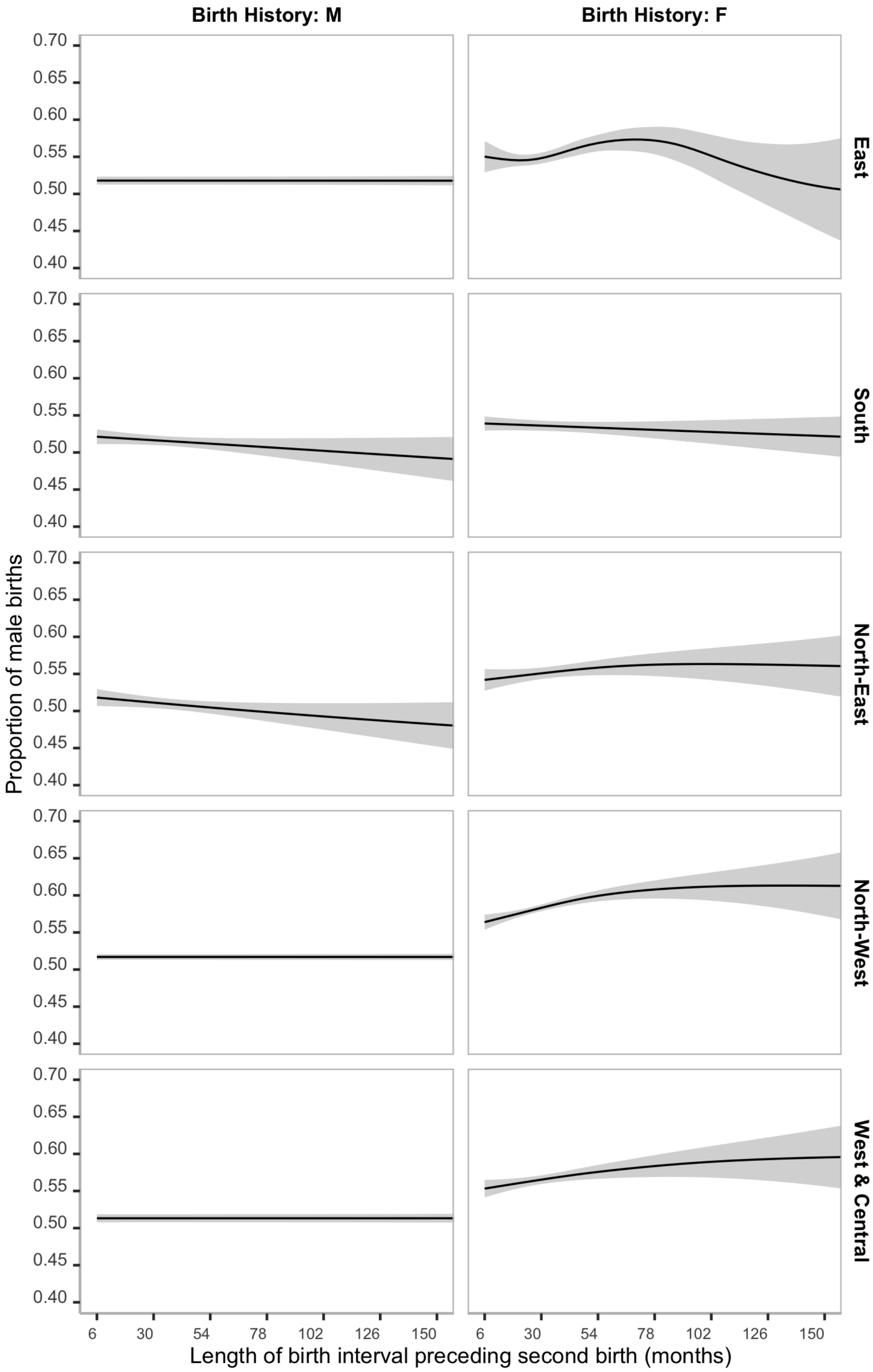

Figure A2: (Smoothed) Proportion of male births at the **second** birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions among births following specific birth histories (columns) in various geographical regions of India (rows)

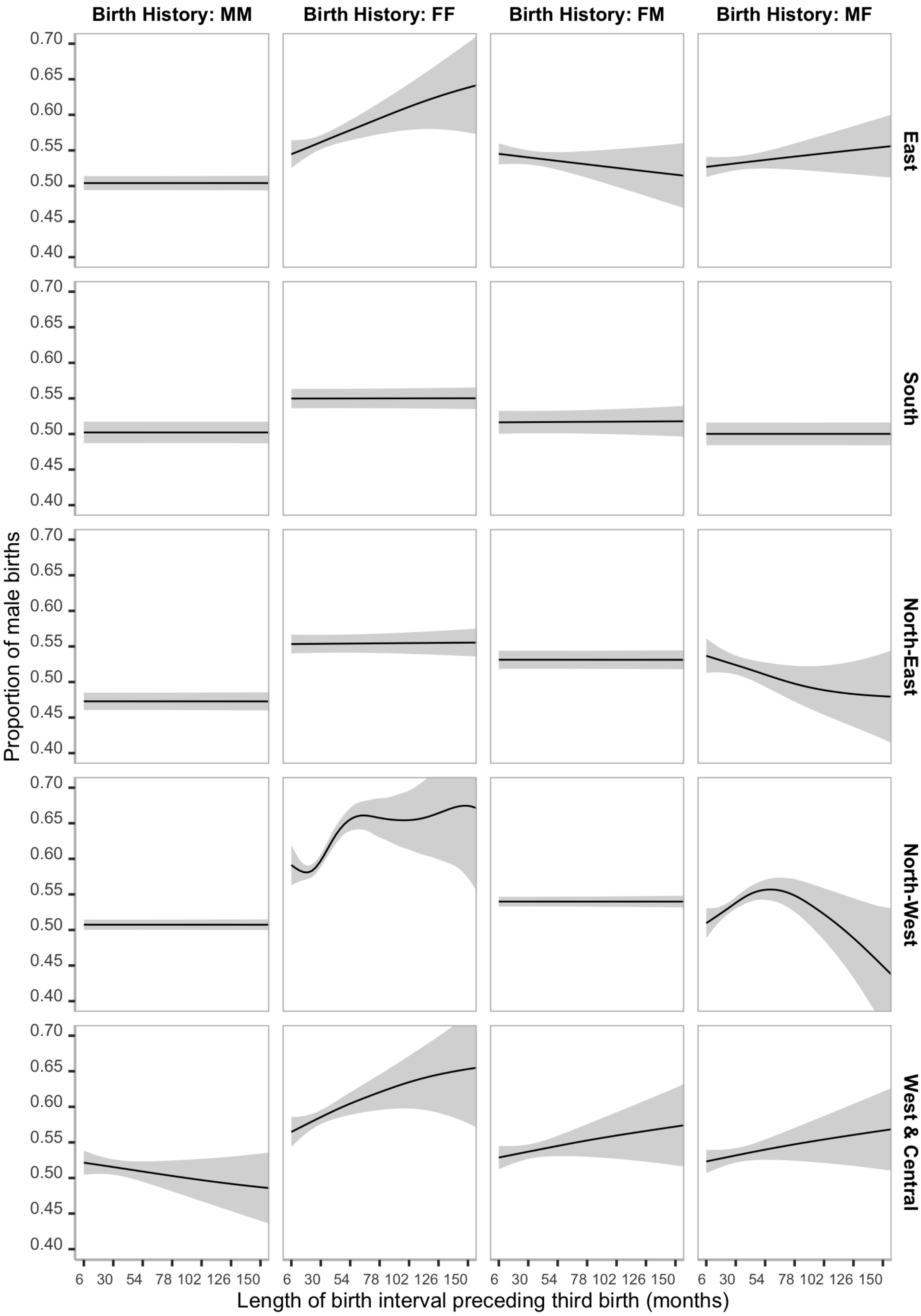

Figure A3: (Smoothed) Proportion of male births at the **third** birth order, conditional on the length of the preceding interval. Each panel plots the conditional proportions among births following specific birth histories (columns) in various geographical regions of India (rows)

	Dependent variable: Birth Interval preceding k^{th} birth (I_k)			
	$\log(I_2)$	I_2	$\log(I_3)$	I_3
	(1)	(2)	(3)	(4)
sex:M	-0.005^{**}	-0.128	-0.003	-0.062
	(0.002)	(0.087)	(0.005)	(0.165)
hist:F	-0.037^{***}	-1.212^{***}		
	(0.003)	(0.094)		
hist:F × sex:M	0.015^{***}	0.475^{***}		
	(0.004)	(0.128)		
hist:FF			-0.032^{***}	-0.972^{***}
			(0.005)	(0.168)
hist:FM			-0.018^{***}	-0.614^{***}
			(0.005)	(0.168)
hist:MF			-0.033^{***}	-1.036^{***}
			(0.005)	(0.167)
hist:FF × sex:M			0.030^{***}	0.993^{***}
			(0.007)	(0.229)
hist:FM × sex:M			0.004	0.163
			(0.007)	(0.232)
hist:MF × sex:M			0.013^{**}	0.356
			(0.007)	(0.232)
Constant	3.392^{***}	32.817^{***}	3.311^{***}	30.380^{***}
	(0.006)	(0.214)	(0.009)	(0.301)
Controls	Yes	Yes	Yes	Yes
Observations	332,656	330,144 (R)	200,824	199,574 (R)
R^2	0.005	0.006	0.004	0.005
Adjusted R^2	0.005	0.006	0.004	0.005
F Statistic	188.427^{***}	222.801^{***}	65.979^{***}	74.703^{***}

Note: *p<0.1; **p<0.05; ***p<0.01
(R): Sample restricted to those with interval $\leq$ 120 months

Table A2: Estimates for alternate Difference-in-Differences specifications

	Dependent variable: Birth Interval preceding k^{th} birth		
	k=4	k=5	k=6
	(1)	(2)	(3)
sex:M	−0.526	−1.106*	−0.468
	(0.353)	(0.666)	(1.257)
hist:FFF×sex:M	1.041**		
	(0.460)		
hist:FFM×sex:M	0.154		
	(0.470)		
hist:FMF×sex:M	0.605		
	(0.484)		
hist:FMM×sex:M	−0.552		
	(0.515)		
hist:MFF×sex:M	0.638		
	(0.481)		
hist:MFM×sex:M	0.650		
	(0.512)		
hist:MMF×sex:M	0.098		
	(0.516)		
hist:FFFF×sex:M		2.279***	
		(0.837)	
hist:FFFM×sex:M		0.013	
		(0.855)	
hist:FFMF×sex:M		0.759	
		(0.893)	
hist:FFMM×sex:M		0.164	
		(0.962)	
hist:FMFF×sex:M		1.388	
		(0.908)	
hist:FMFM×sex:M		0.523	
		(0.984)	
hist:FMMF×sex:M		0.888	
		(1.003)	

continued . . .

. . . continued

	Dependent variable: Birth Interval preceding k^{th} birth		
	k=4	k=5	k=6
	(1)	(2)	(3)
hist:FMMM×sex:M		0.815	
		(1.004)	
hist:MFFF×sex:M		2.354***	
		(0.896)	
hist:MFFM×sex:M		0.698	
		(0.960)	
hist:MFMF×sex:M		1.671*	
		(0.980)	
hist:MFMM×sex:M		0.459	
		(1.001)	
hist:MMFF×sex:M		0.783	
		(0.981)	
hist:MMFM×sex:M		0.315	
		(1.005)	
hist:MMMF×sex:M		1.224	
		(0.989)	
hist:FFFFF×sex:M			1.052
			(1.533)
hist:FFFFM×sex:M			−2.051
			(1.555)
hist:FFFMF×sex:M			2.124
			(1.673)
hist:FFFMM×sex:M			−0.347
			(1.762)
hist:FFMFF×sex:M			1.591
			(1.683)
hist:FFMFM×sex:M			−1.208
			(1.793)
hist:FFMMF×sex:M			−0.620
			(1.895)

continued . . .

. . . continued

	Dependent variable: Birth Interval preceding k^{th} birth		
	k=4	k=5	k=6
	(1)	(2)	(3)
hist:FFMMM×sex:M			0.583
			(1.853)
hist:FMFFF×sex:M			1.370
			(1.708)
hist:FMFFM×sex:M			−1.271
			(1.815)
hist:FMFMF×sex:M			0.952
			(1.891)
hist:FMFMM×sex:M			0.143
			(1.938)
hist:FMMFF×sex:M			1.181
			(1.906)
hist:FMMFM×sex:M			−2.620
			(1.948)
hist:FMMMF×sex:M			2.561
			(1.938)
hist:FMMMM×sex:M			0.245
			(1.927)
hist:MFFFF×sex:M			−0.585
			(1.649)
hist:MFFFM×sex:M			1.117
			(1.791)
hist:MFFMF×sex:M			2.858
			(1.858)
hist:MFFMM×sex:M			1.176
			(1.876)
hist:MFMFF×sex:M			1.135
			(1.848)
hist:MFMFM×sex:M			0.297
			(1.919)
hist:MFMMF×sex:M			0.953

continued . . .

. . . continued

	Dependent variable: Birth Interval preceding k^{th} birth		
	k=4	k=5	k=6
	(1)	(2)	(3)
			(1.955)
hist:MFMMM×sex:M			4.562**
			(1.914)
hist:MMFFF×sex:M			1.569
			(1.843)
hist:MMFFM×sex:M			−0.712
			(1.911)
hist:MMFMF×sex:M			−1.353
			(1.978)
hist:MMFMM×sex:M			−1.222
			(1.906)
hist:MMMFF×sex:M			1.643
			(1.859)
hist:MMMFM×sex:M			−0.894
			(1.956)
hist:MMMMF×sex:M			1.031
			(1.845)
Constant	31.559***	34.197***	33.354***
	(0.497)	(0.773)	(1.237)
Controls	Yes	Yes	Yes
Observations	105,587	52,378	25,529
R^2	0.003	0.002	0.003
F Statistic	13.024***	2.730***	1.252*
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A3: Difference-in-Difference estimates for fourth, fifth and sixth birth orders. The table only reports coefficients for variable sex:M and it's interactions with hist:h.

Mother's Total Parity	Proportion of male births — Relative Birth Order (with respect to actual parity), youngest ⟶ oldest						
	0	-1	-2	-3	-4	-5	**All**
1	0.646	-	-	-	-	-	**0.646**
2	0.621	0.605	-	-	-	-	**0.613**
3	0.620	0.526	0.522	-	-	-	**0.556**
4	0.599	0.492	0.461	0.459	-	-	**0.503**
5	0.582	0.487	0.445	0.430	0.442	-	**0.477**
6	0.567	0.489	0.436	0.434	0.420	0.440	**0.464**

Table A4: Proportion of male births at relative birth orders (with respect to actual parity), for mothers with different total parity. Relative birth order with respect to actual parity, denoted by j is calculated as $j =$ Birth Order - Total Parity. Thus, the youngest child in any family, is born at $j = 0$, and older siblings at lower values of j. The last column reports proportion of sons among all children. The sample is restricted to mothers who report completion of fertility.

Mother's Ideal Parity	Sample Share	Proportion of male births — Relative Birth Order (with respect to ideal parity)					
		- 3	-2	-1	0	1	2
1	0.05	-	-	-	**0.669**	0.550	0.587
2	0.56	-	-	0.548	**0.558**	0.546	0.544
3	0.21	-	0.536	0.527	**0.559**	0.529	0.529
4	0.11	0.483	0.479	0.498	**0.533**	0.518	0.517

Table A5: Proportion of male births at relative birth orders (with respect to ideal parity), for mothers with different ideal parity. Relative birth order with respect to ideal parity, denoted by j^* is calculated as $j^* =$ Birth Order - Ideal Parity. Children born at $j^* \leq 0$ are within the total number of children the parent thinks is ideal, while $j^* > 0$ indicates births beyond ideal number. The sample is restricted to mothers who have non-zero ideal parity and report completion of fertility. The relative orders at which the proportion of sons is the highest are highlighted in bold.

Wealth Quintile	Mother's Ideal Parity	Sample Share	Proportion of male births: Relative Birth Order (with respect to ideal parity) - 3	-2	-1	0	1	2
Poorest	**1**	0.01	-	-	-	**0.649**	0.545	0.527
	2	0.31	-	-	0.552	**0.558**	0.533	0.533
	3	0.31	-	0.542	0.530	**0.554**	0.530	0.527
	4	0.25	0.488	0.481	0.496	**0.531**	0.517	0.512
	5	0.04	0.510	0.516	0.513	**0.529**	0.510	-
Poorer	**1**	0.01	-	-	-	**0.650**	0.531	0.584
	2	0.41	-	-	0.552	**0.556**	0.535	0.538
	3	0.28	-	0.542	0.531	**0.557**	0.526	0.520
	4	0.19	0.487	0.476	0.504	**0.533**	0.523	0.524
	5	0.03	0.485	0.508	0.518	**0.552**	0.495	-
Middle	**1**	0.02	-	-	-	**0.659**	0.542	0.573
	2	0.50	-	-	0.545	**0.550**	0.539	0.544
	3	0.24	-	0.533	0.524	**0.556**	0.524	0.535
	4	0.14	0.475	0.484	0.493	**0.531**	0.509	0.519
	5	0.02	0.520	0.531	0.490	**0.529**	0.558	-
Richer	**1**	0.03	-	-	-	**0.650**	0.528	0.585
	2	0.57	-	-	0.547	**0.552**	0.550	0.547
	3	0.21	-	0.528	0.523	**0.564**	0.529	0.541
	4	0.11	0.469	0.474	0.505	**0.546**	0.524	0.499
	5	0.02	0.503	0.526	0.556	0.553	**0.589**	-
Richest	**1**	0.05	-	-	-	**0.697**	0.585	-
	2	0.66	-	-	0.546	0.571	**0.580**	0.570
	3	0.16	-	0.524	0.517	**0.580**	0.551	0.551
	4	0.07	0.495	0.481	0.489	0.528	0.515	**0.547**
	5	0.01	0.545	0.539	0.555	**0.576**	0.521	0.504

Table A6: Proportion of male births at relative birth orders (with respect to ideal parity), for mothers from households in different wealth quintiles. Relative birth order with respect to ideal parity, denoted by j^* is calculated as j^* = Birth Order - Ideal Parity. Children born at $j^* \leq 0$ are within the total number of children the parent thinks is ideal, while $j^* > 0$ indicates births beyond ideal number. The sample is restricted to mothers who have non-zero ideal parity and report completion of fertility. The relative orders at which the proportion of sons is the highest are highlighted in bold.

Mother's Education	Mother's Ideal Parity	Sample Share	Proportion of male births: Relative Birth Order (with respect to ideal parity) - 3	-2	-1	0	1	2
None	**1**	0.01	-	-	-	**0.648**	0.528	0.559
	2	0.34	-	-	**0.565**	0.560	0.535	0.530
	3	0.31	-	0.551	0.536	**0.554**	0.528	0.529
	4	0.23	0.495	0.483	0.504	**0.532**	0.516	0.513
	5	0.03	0.506	0.526	0.510	**0.548**	0.523	-
Primary	**1**	0.02	-	-	-	**0.655**	0.546	0.594
	2	0.50	-	-	0.553	**0.554**	0.535	0.546
	3	0.25	-	0.534	0.524	**0.559**	0.525	0.511
	4	0.14	0.463	0.473	0.487	**0.534**	0.528	0.529
	5	0.02	0.526	0.510	**0.514**	0.512	0.487	-
Secondary	**1**	0.04	-	-	-	**0.678**	0.555	0.612
	2	0.62	-	-	0.540	0.556	0.559	**0.568**
	3	0.18	-	0.512	0.510	**0.569**	0.542	0.556
	4	0.09	0.458	0.471	0.489	**0.538**	0.518	0.533
	5	0.02	0.497	0.506	**0.541**	0.540	0.524	-
Higher	**1**	0.10	-	-	-	**0.666**	0.572	0.558
	2	0.74	-	-	0.523	0.568	**0.632**	0.627
	3	0.09	-	0.465	0.500	**0.610**	0.477	0.333
	4	0.03	0.492	0.489	0.483	0.522	0.571	**0.714**

Table A7: Proportion of male births at relative birth orders (with respect to ideal parity), for mothers with different levels of education. Relative birth order with respect to ideal parity, denoted by j^* is calculated as j^* = Birth Order - Ideal Parity. Children born at $j^* \leq 0$ are within the total number of children the parent thinks is ideal, while $j^* > 0$ indicates births beyond ideal number. The sample is restricted to mothers who have non-zero ideal parity and report completion of fertility.The relative orders at which the proportion of sons is the highest are highlighted in bold.

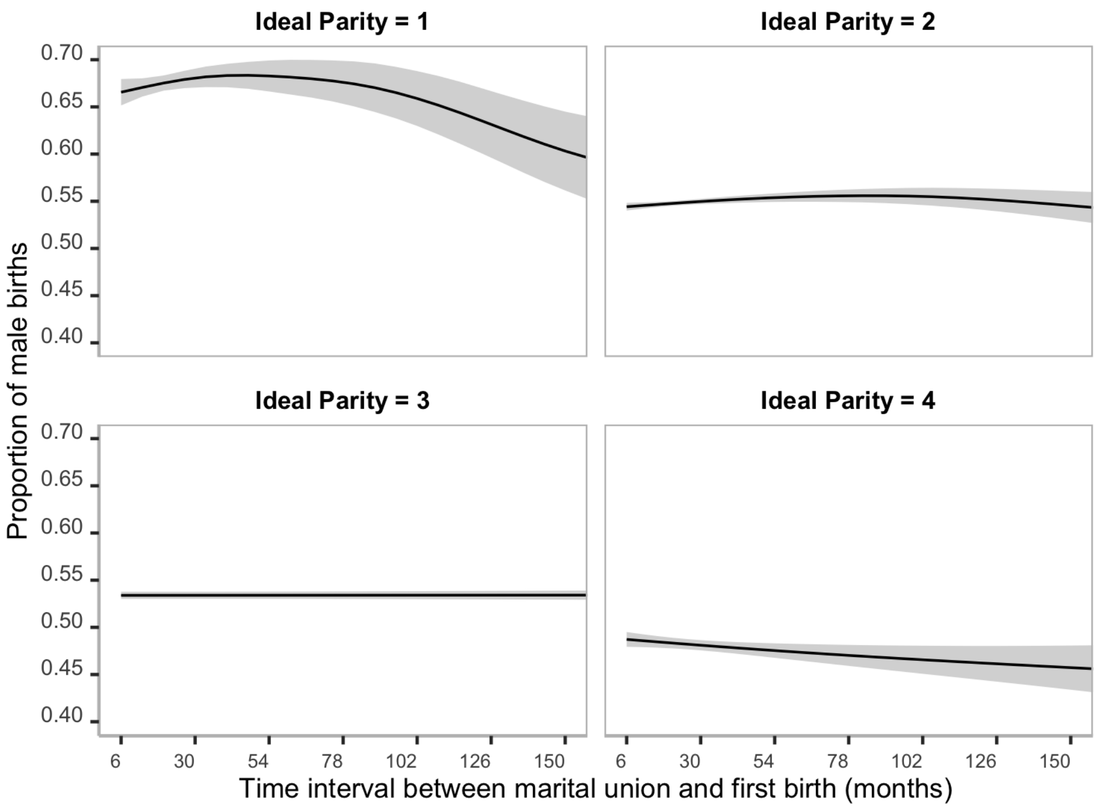

Figure A4: (Smoothed) Proportion of male births at the **firth** birth order, conditional on time interval between marital union and first birth. Each panel plots the conditional proportions for mothers with different ideal parity.

Crude Estimators for Fertility Stoppers and Sex-Selectors

Let N_h be the number of mothers whose birth history begins with the sequence h. If we denote the proportion of mothers who stop childbearing following history h by μ_h, then only $(1 - \mu_h)N_h$ mothers would conceive again. They could either birth their successive child irrespective of its sex or selectively abort a female fetus and try again. If π is the natural probability of giving birth to a male child and λ_h denotes the proportion of mothers who sex-selectively abort following history h, then π proportion of parents would conceive a male child, and $(1 - \pi)(1 - \lambda_h)$ would conceive a female child and not abort it. As a consequence, there would only be $\big[\pi + (1 - \pi)(1 - \lambda_h)\big](1 - \mu_h)N_h$ successive births, of which $\pi(1 - \mu_h)N_h$ would be male and the remaining $(1 - \pi)(1 - \lambda_h)(1 - \mu_h)N_h$ would be female.

Of the $\lambda_h(1 - \pi)(1 - \mu_h)N_h$ mothers who abort and try again, a proportion π would birth male children in their second attempt, followed by a proportion $\pi(1 - \pi)$ who would birth a male child in their third attempt and so on, until finite $j < \infty$ attempts.[21]

Therefore, following any given history h,

$$\begin{aligned}
\text{Number of female successive births} &= \big[(1 - \pi)(1 - \lambda_h)\big](1 - \mu_h)N_h \\
\text{Number of male successive births} &= \big[\pi + \lambda_h(1 - \pi)(1 - (1 - \pi)^j)\big](1 - \mu_h)N_h \\
\text{Total number of successive births} &= \big[1 - \lambda_h(1 - \pi)^{j+1}\big](1 - \mu_h)N_h
\end{aligned}$$

Here, female successive births would occur only to parents who aren't sex-selectors and naturally conceive a female fetus. Male successive births would be to those who either conceive a male child at their first attempt, or selectively abort female fetuses enough times to conceive a male child in less than j attempts.

Using these, we can obtain estimates of μ_h and λ_h based on actual counts of births of either sex observed in the sample.

Let $\hat{p_h}$ be the proportion of male successive births observed in the sample following h. Then, equating $1 - \hat{p_h}$ to the theoretical *proportion* of female successive births, we get

$$1 - \hat{p_h} = \frac{\big[(1 - \pi)(1 - \lambda_h)\big](1 - \mu_h)N_h}{\big[1 - \lambda_h(1 - \pi)^{j+1}\big](1 - \mu_h)N_h}$$

21. We are assuming *all* mothers who sex-selectively abort would try to conceive again up to a maximum of j times. We also assume that no one would reverse their decision to sex-select and choose to birth a female child after consecutive sex-selective abortions.

which can be solved to obtain an estimator for proportion of sex-selectors, given by

$$\hat{\lambda_h} = \frac{(1-\pi) - (1-\hat{p_h})}{(1-\pi) - (1-\hat{p_h})(1-\pi)^{j+1}} = \frac{\hat{p_h} - \pi}{(1-\pi) - (1-\hat{p_h})(1-\pi)^{j+1}}$$

Likewise, equating the count of successive female births in the sample to the theoretical *counts* of female successive births we obtain an estimator for the proportion of fertility stoppers

$$\hat{\mu_h} = 1 - \frac{\#\text{ Female successive births}}{(1-\pi)N_h(1-\hat{\lambda_h})}$$

References

Anderson, Siwan, and Debraj Ray. 2010. “Missing Women: Age and Disease.” *Review of Economic Studies* 77:1262–1300.

Baland, Jean-Marie, Guilhem Cassan, and Francois Woitrin. 2020. “The Stopping Rule and Gender selective mortality: World Evidence.” *CEPR Discussion Papers.*

Basu, Deepankar, and Robert De Jong. 2010. “Son targeting fertility behavior: Some consequences and determinants.” *Demography* 47 (2): 521–536.

Bhalotra, Sonia, and Tom Cochrane. 2010. “Where Have All the Young Girls Gone? Identification of Sex Selection in India.” *IZA Discussion Papers, No. 5381.*

Chao, Fengqing, Patrick Gerland, Alex R. Cook, and Leontine Alkema. 2019. “Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (19): 9303–9311.

Clark, Shelley. 2000. “Son Preference and Sex Composition of Children: Evidence from India.” *Demography* 37 (1): 95–108.

Dyson, Tim, and Mick Moore. 1983. “On Kinship Structure, Female Autonomy, and Demographic Behavior in India.” *Population and Development Review,* 35–60. https://doi.org/10.2307/1972894.

Ebenstein, Avraham. 2010. “The ”Missing Girls” of China and the Unintended Consequences of the One Child Policy.” *The Journal of Human Resources* 45 (1): 87–115.

International Institute for Population Sciences - IIPS/India and ICF. 2017. “National Family Health Survey NFHS-4, 2015-16: India. Mumbai: IIPS.”

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

Jha, Prabhat, Rajesh Kumar, Priya Vasa, Neeraj Dhingra, Deva Thiruchelvam, and Rahim Moineddin. 2006. “Low Male-to-Female Sex Ratio of Children Born in India: National Survey of 1· 1 Million Households.” *The Lancet* 367:211–218.

Orzack, Steven Hecht, J. William Stubblefield, Viatcheslav R. Akmaev, Pere Colls, Santiago Munné, Thomas Scholl, David Steinsaltz, and James E. Zuckerman. 2015. “The human sex ratio from conception to birth.” *Proceedings of the National Academy of Sciences* 112 (16): E2102–E2111.

Pörtner, Claus C. 2016. "Sex-Selective Abortions, Fertility and Birth Spacing." *World Bank Policy Research Working Paper, no. WPS 7189.*

———. 2020. "Birth Spacing and Fertility in the Presence of Son Preference and Sex-Selective Abortions: India's Experience Over Four Decades." *Working Paper.*

Rasul, Imran. 2008. "Household Bargaining Over Fertility: Theory and Evidence From Malaysia." *Journal of Development Economics,* 215–41.

Sen, Amartya. 1990. "More Than 100 Million Women Are Missing." *The New York Review of Books.*

———. 1992. "Missing Women." *British Medical Journal* 304:587–588.

Tables created in R using stargazer v.5.2.2 by Marek Hlavac, Harvard University.