



**HAL**  
open science

# Measuring sex-selective abortion: How many women abort?

Aditi Dimri, Véronique Gille, Philipp Ketz

► **To cite this version:**

Aditi Dimri, Véronique Gille, Philipp Ketz. Measuring sex-selective abortion: How many women abort?. *Journal of Development Economics*, 2024, 171, 10.1016/j.jdeveco.2024.103342. hal-04671748

**HAL Id: hal-04671748**

**<https://hal.science/hal-04671748v1>**

Submitted on 4 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



## Regular article

## Measuring sex-selective abortion: How many women abort? ☆

Aditi Dimri <sup>a</sup>, Véronique Gille <sup>b,\*</sup>, Philipp Ketz <sup>c</sup><sup>a</sup> Independent Researcher, Germany<sup>b</sup> Université Paris-Dauphine, Université PSL, LEDA, CNRS, IRD, France<sup>c</sup> Paris School of Economics - CNRS, France

## ARTICLE INFO

## JEL classification:

D10

J13

## Keywords:

Sex ratio

Sex-selective abortion

Missing girls

Measurement

India

## ABSTRACT

This paper demonstrates that sex-selective abortion induces a correlation between birth interval length and the sex of the next-born child. Using a statistical model, we show that shorter birth intervals for next-born girls indicate *repeated* sex-selective abortions between consecutive births. Analyzing data from India, we find evidence of repeated sex-selective abortions at birth order 2 when the first child is a girl, and strong evidence at birth order 3 when the first two children are girls. To quantify the extent of repeated abortions, we propose a maximum likelihood estimator that provides the number of women who abort and their likelihood of performing repeated abortions. Our estimation results reveal significant heterogeneity across birth orders, sibling compositions, and socio-demographic and geographic groups. Notably, literate and urban women who first had a girl rarely abort a second time, whereas women in northern India who first had two girls show a 13% likelihood of repeated sex-selective abortion. In this group, the estimated number of aborted female fetuses—the standard measure of sex-selective abortion—is 50% higher than the number of women who abort.

## 1. Introduction

Sex-selective abortion is an important societal issue. It results in male-biased sex ratios, leading to a series of undesirable societal consequences.<sup>1</sup> Additionally, there are concerns about morbidity and mortality for women, as sex-selective abortions are illegal in most countries and are often performed under unsafe conditions (Grimes et al., 2006; Singh et al., 2018a). Currently, the main measure of sex-selective abortion is the number of aborted female fetuses (see, e.g., Bongaarts and Guilmo, 2015; Bhalotra and Cochrane, 2010; Chao et al., 2019; Klasen and Wink, 2003). It is based on the difference between the observed proportion of male births and the natural probability of male birth.<sup>2</sup> While this measure provides information about the extent of sex-selective abortion, it does not address its prevalence. In other words, it tells us how many female fetuses are aborted but not how many women are involved. Is it 1000 women who abort once, or one woman who aborts 1000 times? Here, a key observation is that women may become

pregnant with a female fetus again after a sex-selective abortion and choose to abort again. Understanding how many women abort and how likely they are to do so repeatedly is not only interesting in its own right but also important from a public health perspective, particularly if there are increased health risks associated with repeated abortions.

To answer these questions, we start by pointing out that there is empirical content in the *joint distribution* of birth spacing and the sex of the next-born child. Fig. 1 shows the empirical distribution functions (*edfs*) of birth spacing by the sex of the next-born child in two different samples of women in India. In panel (a), where there is no evidence of sex-selective abortion—the observed proportion of male births, 0.517, is close to the natural probability of male birth, 0.513 (see e.g., Chao et al., 2019; Dubuc and Coleman, 2007)—the *edfs* of birth spacing for next-born girls and next-born boys are almost indistinguishable. In panel (b), however, where the proportion of male births is much higher at 0.568, indicative of sex-selective abortion, the *edfs* of birth spacing for next-born girls and next-born boys are clearly

☆ We thank the editor and two referees for helpful comments and suggestions, which have considerably improved the paper. We also thank François Maniquet and Jean-Marie Baland, as well as seminar participants at CORE and IRES, Université Catholique de Louvain, Delhi School of Economics, University of Namur, and University of Warwick, for their valuable feedback.

\* Corresponding author.

E-mail addresses: [aditidimri@gmail.com](mailto:aditidimri@gmail.com) (A. Dimri), [veronique.gille@ird.fr](mailto:veronique.gille@ird.fr) (V. Gille), [philipp.ketz@psemail.eu](mailto:philipp.ketz@psemail.eu) (P. Ketz).

<sup>1</sup> For example, they lead to unbalanced marriage markets, which may result in poverty among childless men in old age, especially in countries without a social security system (Das Gupta et al., 2010). Anukriti et al. (2022) list additional consequences in their introduction.

<sup>2</sup> For example, Chao et al. (2019) thus estimate that, between 1970 and 2017, 45 million female births were missing in the world and that China and India account for 51% and 46% of these, respectively. In a comprehensive policy report, Guilmo (2012) lists other countries with “distorted” sex ratios such as Albania, Armenia, Azerbaijan, Georgia, Montenegro, Pakistan, and Vietnam.

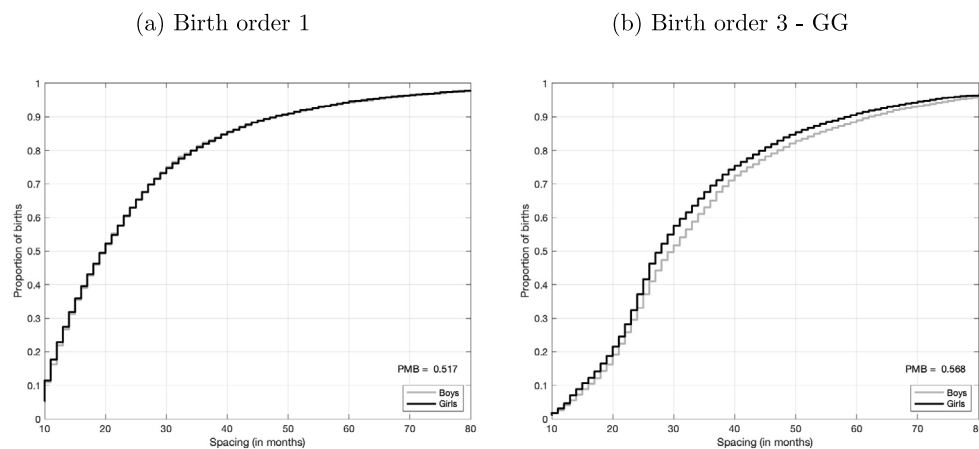


Fig. 1. Empirical distribution functions of birth spacing by sex of the next-born.

Note: Panel (a): Empirical distribution function (*edf*) of birth spacing at birth order 1, i.e., the time between marriage and birth of the first child, by sex of the firstborn. Panel (b): *edf* of spacing at birth order 3, i.e., the time between the births of the second and third child, when the first two children are girls (GG) by sex of the thirdborn. Both *edfs* pertain to literate women during the period 1988–1995; see Section 4 for a detailed data description. In both panels, the proportion of male births (PMB) is indicated in the bottom-right corner.

distinct. The correlation between birth interval length and the sex of the next-born child observed in panel (b) results from the fact that sex-selective abortions not only increase the proportion of male births but also prolong birth spacing, as it takes time to determine the sex of the fetus in utero and to conceive again.

In this paper, we propose a method to exploit the information contained in the differences in birth spacing between next-born girls and next-born boys. First, we introduce a test for the presence of (repeated) sex-selective abortion. Second, we propose two measures that complement the number of aborted female fetuses: the number of women who abort and the share of women at risk of multiple abortions, which measures the likelihood of women performing repeated sex-selective abortions. Since our methodology exploits birth interval lengths, our test and measures are only informative about (repeated) sex-selective abortion *between two consecutive births*, and our populations of interest are implicitly defined by birth order. Consequently, we cannot directly address repeated sex-selective abortion in the population as a whole. Nevertheless, we believe that our proposed methodology provides valuable insights that complement the information provided by the number of aborted female fetuses.

Our method relies on a simple statistical model for birth interval length in the presence of sex-selective abortion. The aforementioned test is an immediate by-product of our model and has the attractive feature that it relies solely on differences in average spacing between next-born girls and next-born boys, thus avoiding the need to specify the natural probability of male birth—a contested topic (Chao et al., 2019). Specifically, if birth spacing differs by the sex of the next-born child, this provides evidence of sex-selective abortion. Additionally, if the time to birth is shorter when the next-born is a girl, this provides evidence of repeated sex-selective abortion.

To obtain the number of women who abort and the share of women at risk of multiple abortions, we estimate our model using maximum likelihood. The main parameters of our model—assuming women abort at most twice—are the share of women who abort when the first pregnancy since the preceding birth is with a female fetus, and the share of women who abort a second time if again pregnant with a female fetus after a first abortion. These parameters readily provide us with the number of women who abort and the share of women at risk of multiple abortions. While we are not the first to note that sex-selective abortion impacts birth spacing (see, e.g., Pörtner, 2022; Anukriti et al., 2022), the observation that this impact depends on the sex of the next-born and can be used to measure the prevalence of sex-selective abortion is, to the best of our knowledge, new to the literature.

We use our proposed methodology to study the evolution of sex-selective abortion patterns in India over time, utilizing data from five rounds of the Demographic and Health Surveys (1992–93, 1998–99, 2005–06, 2015–16, and 2019–2021) and the 2002–04 round of the District Level Household & Facility Survey. While sex-selective abortion is practiced in several countries (Guilmoto, 2012), India has been estimated to account for 46% of the aborted female fetuses worldwide over the period 1970–2017 (Chao et al., 2019), making it an interesting case study. Our analysis is conducted at birth orders 1–3, for different sibling compositions, and is broken down according to socio-demographic and geographic factors. Specifically, we consider literate and illiterate women, women in urban and rural areas, and women in the North and South of India.

At birth order 1, consistent with the literature, we find little to no evidence of sex-selective abortion. At birth order 2, when the first child is a girl, our test provides some evidence of repeated sex-selective abortion. Strong evidence is found at birth order 3, when the first two children are girls. Our estimation results corroborate this, showing that the share of women who sex-selectively abort is increasing over time and is larger at birth order 3 than at birth order 2. We also find strong heterogeneity across socio-demographic and geographic groups and over time. For example, literate women and those in urban areas are more likely to abort than illiterate women and those in rural areas. Furthermore, literate and urban women do not abort a second time at birth order 2 but have a likelihood of around 9% to repeatedly abort at birth order 3, which appears to be stable over time (after 1995). Among women in the North, our estimates indicate that the likelihood of women aborting repeatedly has increased over time at both birth orders 2 and 3. At birth order 3, the likelihood is estimated to have doubled, from approximately 6.5% to around 13%. The latter number corresponds to an estimated number of aborted female fetuses per 1000 women of 99, which is around 50% larger than the estimated number of women who abort (65).

As noted earlier, our proposed measures should prove useful from a public health perspective. Abortions are often practiced in unsafe conditions, even in countries where they are legal. Singh et al. (2018b), for example, estimate that 5% of all abortions in India are practiced “outside of health facilities with methods other than medication abortion”. Nonmedical methods, such as inserting objects or liquids into the vagina (Singh et al., 2018a), can put the health of women at high risk. While abortion-related deaths are very rare (Yokoe et al., 2019), post-abortion complications are more frequent and can be severe, such as a perforated uterus or infections (Singh et al., 2018b). Given that

sex-selective abortions are illegal and more likely to be carried out in unsafe conditions, their repeated use can be extremely detrimental to maternal health both physically and mentally. Our methodology can help policymakers identify socio-demographic and geographic groups most at risk of multiple abortions, which may differ from groups with the largest number of aborted female fetuses.

Our model and findings also highlight the need for caution in interpreting sex ratios at birth. The sex ratio at birth is often used as a proxy for son preference, as seen in the literature on the origins of son preference (Alesina et al., 2018; Goli et al., 2022; Mavisakalyan and Minasyan, 2023; Qian, 2008) and its societal impact (Hwang et al., 2019). These studies typically assume that, for example, two geographical areas with the same sex ratio have (approximately) the same level of son preference. However, as this paper shows, the number of women who abort may very well differ across these geographical areas.

The rest of the paper is organized as follows. Section 2 provides the contextual background for sex-selective abortion in India. Section 3 introduces our model and illustrates its implications. This section also introduces our maximum likelihood estimator and our parameters of interest. Sections 4 and 5 present the data and our empirical findings, respectively. Finally, Section 6 concludes. Additional material, including proofs, can be found in [Appendices A–D](#).

## 2. Sex-selective abortion in India

Abortion has been legal in India since the enactment of the *Medical Termination of Pregnancy Act, No. 34, Government of India (1971)*. This act allows pregnancies to be terminated up to 20 weeks gestation by registered medical practitioners at certified abortion facilities. However, sex-selective abortion is illegal and regulated under the *Pre-Natal Diagnostic Techniques (Regulation and Prevention of Misuse) Act, No. 57, Government of India (1994)*. This act prohibits the misuse of antenatal diagnostic tests for determining the sex of the fetus. It bans the advertisement of such tests, mandates the registration of all facilities that use them, and forbids revealing the sex of the fetus to the expectant parents.

Despite its prohibition, sex-selective abortion is widely practiced. [Chao et al. \(2019\)](#), for example, estimate the number of missing female births in India between 1970 and 2017 to be 20.7 million. While the 1994 ban on sex-selective abortion seems to have slowed down its increase among poorer households ([Rastogi and Sharma, 2022](#)), the imbalance in the sex ratio at birth has continued to rise. [Saikia et al. \(2021\)](#) estimate that the number of missing female births has increased by one million per decade, from 3.5 million between 1987 and 1996 to 5.5 million between 2007 and 2016.

Two main explanations have been proposed for this increase. First, sex determination during pregnancy has become more accessible and cheaper since the mid-1990s, following increases in imports and domestic production of ultrasound machines ([Anukriti et al., 2022](#)). Second, desired total fertility (i.e., the total number of children) has decreased ([Cassan et al., 2023](#)). As the probability of having at least one son decreases with lower total fertility, households that want to have at least one son are more likely to use sex-selective abortion. [Jayachandran \(2017\)](#) estimates that between one-third and one-half of the increase in the child sex ratio between 1981 and 2011 can be attributed to this change in desired total fertility.<sup>3</sup>

Furthermore, there is heterogeneity in the occurrence of sex-selective abortion across regions and demographic groups in India. Sex-selective abortion is more common in northern India ([Saikia et al., 2021](#)), urban areas ([Pörtner, 2022](#)), wealthier households, and among more educated women ([Bhalotra and Cochrane, 2010](#)). While differences

<sup>3</sup> [Jayachandran \(2017\)](#) uses the child sex ratio, i.e., the sex ratio among children aged 0–6 years, as a proxy for the sex ratio at birth.

across regions in terms of child sex ratios have decreased between 1991 and 2001 ([Diamond-Smith and Bishai, 2015](#)), differences between demographic groups in terms of the number of missing female births have increased over the same period ([Jha et al., 2011](#)).

The occurrence of sex-selective abortion also depends on birth order and the sex composition of previous children. The consensus in the literature is that there is no sex-selective abortion at birth order 1 ([Dahl and Moretti, 2008](#); [Milazzo, 2018](#); [Heath and Tan, 2018](#)), although [Saikia et al. \(2021\)](#) find some evidence for recent years. At higher birth orders, sex-selective abortion is more common in households that previously only had girls, reflecting a widespread preference for having at least one son ([Bhalotra and Cochrane, 2010](#)). While the number of missing female births was highest at birth order 3, [Saikia et al. \(2021\)](#) find that more female births have been missing at birth order 2 in recent years (2007–2016).

## 3. Birth interval length in the presence of sex-selective abortion

In the absence of sex-selective abortion, there is no correlation between birth interval length and the sex of the next-born child. While birth interval length may depend on, and thus be correlated with, the sexes of the preceding children, the sex of the next-born child, absent sex-selective abortion, is random and therefore independent of birth interval length. By contraposition, if birth interval length correlates with the sex of the next-born child, this provides evidence of sex-selective abortion. To the best of our knowledge, this observation is new to the literature. In what follows, we propose a simple model for birth interval length that (i) links the sign of the correlation between birth interval length and the sex of the next-born child to the presence of repeated sex-selective abortion, and (ii) in combination with certain distributional assumptions, allows us to estimate parameters of interest such as the share of women who abort when pregnant with a female fetus.

This section is organized as follows. Section 3.1 introduces our model, assuming fixed time intervals for ease of exposition. It also presents a proposition that can be used to test for the presence of (repeated) sex-selective abortion between two consecutive births. Section 3.2 relaxes the assumption of fixed time intervals, discussing the necessary assumptions for the previously obtained results to hold, which also underlie our maximum likelihood estimator introduced in Section 3.3. Finally, Section 3.4 introduces the number of aborted female fetuses, which corresponds to the number of missing women of [Sen \(1990\)](#) adapted to our setting, along with other parameters of interest.

Before introducing our model, an important comment is in order. As mentioned above, birth interval length may depend on the sexes of the preceding children. Assuming that households have a preference for sons, it is, for example, conceivable that households that first have a girl (rather than a boy) have shorter birth intervals, as they may decrease their waiting time before trying to have another child. This is consistent with evidence that girls are breastfed for shorter periods of time ([Jayachandran and Kuziemko, 2011](#)). Because we do not wish to model this difference in waiting times, we condition on the composition and birth order of the siblings of the next-born child.

### 3.1. A simple model for birth interval length

We start by assuming that each sex-selective abortion prolongs the birth interval by a fixed amount of time,  $D$ . We also assume that the time to birth in the absence of abortion is fixed. Letting  $T_b^a$  denote the length of the birth interval (time to birth) when  $a$  abortions are performed between the two births, we have  $T_b^a = T_b^0 + aD$ . The observed time to birth is  $T_b \equiv T_b^A$ , where  $A$  denotes the number of sex-selective abortions. Here and in what follows, we use the notational convention that uppercase and lowercase letters denote random variables and their possible realizations, respectively. Here,  $a \in \{0, 1, \dots, N_0\}$ , where  $N_0$

denotes the (known) maximum number of “possible” abortions. As mentioned above, we condition our analysis on the composition and birth order of the siblings. However, we omit this dependence for notational convenience.

When the first pregnancy since the preceding birth is with a female fetus, a household aborts with probability  $\alpha_1$ . More generally, a household aborts with probability  $\alpha_k$  when pregnant with a female fetus for the  $k$ th time having aborted  $k - 1$  times since the preceding birth.<sup>4</sup> Here, a key observation is that a household can become pregnant again with a female fetus after having aborted one or several female fetuses. In what follows, let  $Y \in \{0, 1\}$  denote the sex of the next-born child, with 0 indicating the birth of a girl and 1 indicating the birth of a boy. Letting  $\pi$  denote the natural probability of male birth, the probabilities of having a girl after  $a$  abortions and a boy after  $a$  abortions are given by

$$P(A = a, Y = 0) = (1 - \pi)^a (1 - \alpha_{a+1}) \prod_{k=0}^a \alpha_k, \tag{1}$$

and

$$P(A = a, Y = 1) = (1 - \pi)^a \prod_{k=0}^a \alpha_k \tag{2}$$

respectively, where  $\alpha_0 = 1$  and  $\alpha_{N_0+1} = 0$ . By definition,  $P(Y = 0) = \sum_{a=0}^{N_0} P(A = a, Y = 0)$  and  $P(Y = 1) = \sum_{a=0}^{N_0} P(A = a, Y = 1)$ .<sup>5</sup> The model implied by Eqs. (1) and (2) will be illustrated in the following sections, where we consider the cases where women abort at most once and at most twice. We note that the above model implicitly assumes no sex-selective abortion of male fetuses. We discuss this assumption in Section 3.2.

### 3.1.1. The case with at most one sex-selective abortion

We now show that if households abort at most once, i.e.,  $A \in \{0, 1\}$ , then the expected birth interval length is longer when the next-born child is a girl than when it is a boy, i.e.,  $E(T_b|Y = 0) > E(T_b|Y = 1)$ . This result follows from the fact that an abortion prolongs the birth interval (recall  $T_b^A = T_b^0 + AD$ ) and that an abortion is more likely to have taken place when the next-born child is a girl than when it is a boy, i.e.,  $P(A = 1|Y = 0) > P(A = 1|Y = 1)$ .

To see this, note that the (expected) time that elapses before the birth of a child of sex  $y$  is a weighted average of the time that elapses if no abortion takes place,  $T_b^0$ , and the time that elapses if one abortion takes place,  $T_b^1 = T_b^0 + D$ , weighted by the corresponding probabilities, which are given by  $P(A = 0|Y = y)$  and  $P(A = 1|Y = y)$ , respectively. That is,

$$\begin{aligned} E(T_b|Y = y) &= P(A = 0|Y = y) \cdot T_b^0 + P(A = 1|Y = y) \cdot (T_b^0 + D) \\ &= (1 - P(A = 1|Y = y)) \cdot T_b^0 + P(A = 1|Y = y) \cdot (T_b^0 + D) \\ &= T_b^0 + P(A = 1|Y = y) \cdot D \end{aligned} \tag{3}$$

Since the time that elapses if one abortion takes place is longer than the time that elapses if no abortion takes place, the time to birth is longer when the next-born child is a girl than when it is a boy if the probability

<sup>4</sup> One can think of the  $\alpha$ s as average probabilities for underlying types that differ in their willingness to abort (WTA). Different levels of WTA may reflect, for example, different strengths of the preference for the next-born child to be a boy. Imagine that there are two types of households, one with low WTA and one with high WTA. Letting  $p^L$  and  $p^H$  denote their respective probabilities of sex-selective abortion, we have, for example,  $\alpha_1 = \pi^H p^H + (1 - \pi^H) p^L$ , where  $\pi^H$  denotes the proportion of households with high WTA. Thus, even though the decision to perform an abortion is modeled as random, our model is compatible with non-random decisions ( $p^L = 0$  and  $p^H = 1$ ).

<sup>5</sup> Guilмото et al. (2020) consider a probabilistic model of sex-selective abortion behavior at a given parity similar to ours. However, they do not analyze the implications of their model for spacing.

**Table 1**  
Probabilities of sex and # abortions with at most one abortion.

$y$	$P(A = 0, Y = y)$	$P(A = 1, Y = y)$	$P(Y = y) = \sum_{a=0}^1 P(A = a, Y = y)$
0	$(1 - \pi)(1 - \alpha_1)$	$(1 - \pi)\alpha_1(1 - \pi)$	$(1 - \pi)[(1 - \alpha_1) + (1 - \pi)\alpha_1]$
1	$\pi$	$(1 - \pi)\alpha_1\pi$	$\pi[1 + (1 - \pi)\alpha_1]$

or weight of one abortion is higher when the next-born child is a girl than when it is a boy. That is,

$$E(T_b|Y = 0) = T_b^0 + P(A = 1|Y = 0) \cdot D > T_b^0 + P(A = 1|Y = 0) \cdot D = E(T_b|Y = 1)$$

as long as  $P(A = 1|Y = 0) > P(A = 1|Y = 1)$ .

In order to see that an abortion is more likely to have taken place when the next-born child is a girl, it is useful to consider Fig. 2, which plots all possible paths to birth when there is at most one abortion. Intuitively, an abortion is more likely to have occurred when the next-born child is a girl because almost as many girls as boys are born “after” an abortion (cf. Fig. 2 at time  $T_b^0 + D$ ), whereas there are fewer girls born than boys “without” an abortion (cf. Fig. 2 at time  $T_b^0$ ), precisely because some female fetuses are aborted.

To show this more formally, consider Table 1 which tabulates the probabilities in Eqs. (1) and (2) as well as their sums across the number of abortions when there is at most one abortion. From Table 1, it is easy to see that the probability that an abortion has occurred when the next-born child is a girl and when it is a boy are given by

$$P(A = 1|Y = 0) = \frac{P(A = 1, Y = 0)}{P(Y = 0)} = \frac{(1 - \pi)\alpha_1}{1 - \alpha_1 + (1 - \pi)\alpha_1}. \tag{4}$$

and

$$P(A = 1|Y = 1) = \frac{P(A = 1, Y = 1)}{P(Y = 1)} = \frac{(1 - \pi)\alpha_1}{1 + (1 - \pi)\alpha_1}, \tag{5}$$

respectively. Comparing Eqs. (4) and (5), we see that that an abortion is indeed more likely to have occurred when the next-born child is a girl, as the denominator in (4) is smaller than the denominator in (5) while the two numerators are equal.

**GRAPHICAL ILLUSTRATION:** Fig. 3 graphically illustrates the joint and conditional probabilities of our model for  $\alpha_1 = 0.8$  and  $\pi = 0.513$ . Panel (a) shows the joint probabilities  $P(A = a, Y = y)$  for zero abortions ( $a = 0$ ) and one abortion ( $a = 1$ ) from left to right and for girls ( $y = 0$ ) and boys ( $y = 1$ ) from top to bottom. One can see that there are more boys born than girls—the two bottom squares are larger than the two top squares. This is difficult to see with the naked eye for the two right squares—the bottom square is “only” 2.6% ( $= 0.513 - (1 - 0.513)$ ) larger—and reflects that there are almost as many girls born as boys “after” an abortion. Panel (b) shows the corresponding conditional probabilities  $P(A = 0|Y = y)$ . Here, one can see that one abortion has a larger weight for girls than for boys—the top-right circle is greater than the bottom-right circle. The black dots indicate the expected time to birth for girls in the first row and for boys in the second row, which are weighted averages of  $T_b^0$  and  $T_b^0 + D$  (see x-axis) where the weights are represented by the circles. This illustrates that expected birth interval length is longer when the next-born child is a girl than when it is a boy when there is at most one abortion.

Next, we show that when there are at most two abortions the expected birth interval length may be longer or shorter when the next-born child is a girl than when it is a boy.

### 3.1.2. The case with at most two sex-selective abortions

When there are at most two sex-selective abortions, i.e.,  $A \in \{0, 1, 2\}$ , the (expected) time to birth when the next-born child has sex  $y$  is a weighted average of the time that elapses if no abortion takes place,  $T_b^0$ , the time that elapses if one abortion takes place,  $T_b^1 = T_b^0 + D$ , and the time that elapses if two abortions take place,  $T_b^2 = T_b^0 + 2 \cdot D$ ,

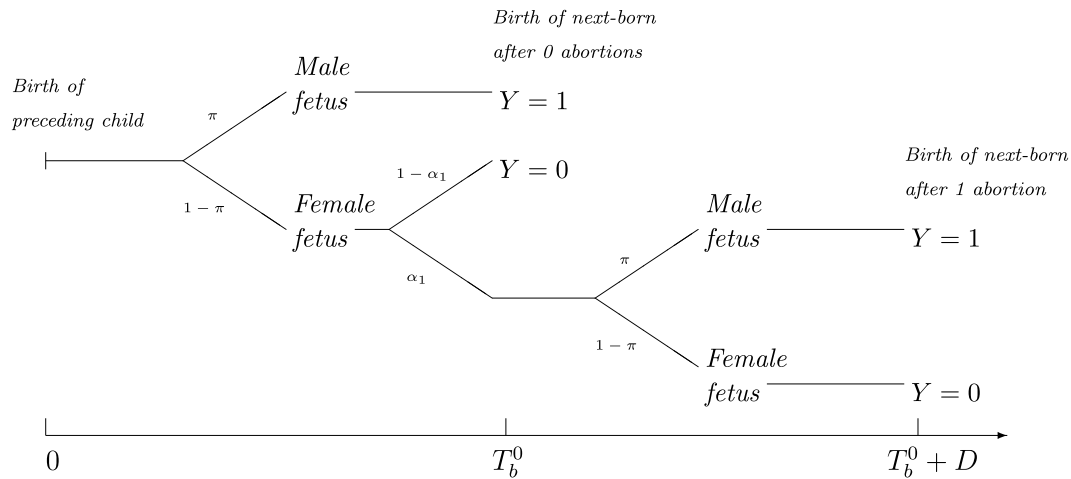


Fig. 2. Paths to birth with at most one abortion.

Note: The figure represents all possible outcomes in terms of sex and time to birth, as well as their respective probabilities, when households abort at most once between two consecutive births.

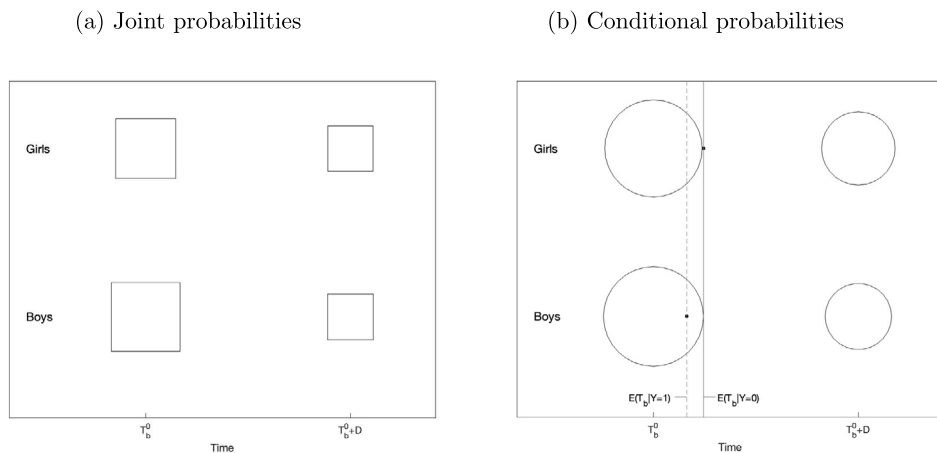


Fig. 3. Joint and conditional probabilities with at most one abortion.

Note: Panel (a) visualizes  $P(A = 0, Y = 0)$ ,  $P(A = 1, Y = 0)$ ,  $P(A = 0, Y = 1)$ , and  $P(A = 1, Y = 1)$  from left to right and from top to bottom, i.e., the areas of the squares represent probabilities that should be thought of as summing to one. Panel (b) visualizes  $P(A = 0|Y = 0)$ ,  $P(A = 1|Y = 0)$ ,  $P(A = 0|Y = 1)$ , and  $P(A = 1|Y = 1)$  from left to right and from top to bottom, i.e., the areas of the circles represent conditional probabilities that should be thought of as summing to one for each line. The black dots indicate  $E(T_b|Y = y)$  for  $y = 0$  (Girls) and  $y = 1$  (Boys). In both panels,  $\alpha_1 = 0.8$  and  $\pi = 0.513$ .

weighted by the corresponding probabilities  $P(A = a|Y = y)$  with  $a \in \{0, 1, 2\}$ , respectively. That is,

$$E(T_b|Y = y) = \sum_{a=0}^2 P(A = a|Y = y) \cdot (T_b^0 + a \cdot D) = T_b^0 + \sum_{a=1}^2 P(A = a|Y = y) \cdot a \cdot D. \tag{6}$$

Therefore, whether the expected time to birth is shorter or longer when the next-born child is a girl depends on those probabilities or weights. When there is at most one abortion, then one abortion is more likely to have occurred, i.e., it has a larger weight, if the next-born child is a girl than when it is a boy. Similarly, when there are at most two abortions, then two abortions are more likely to have occurred, i.e., they have a larger weight, when the next-born child is a girl than when it is a boy.

To gain some intuition for this, consider Fig. 4, which plots all possible paths to birth when there are at most two abortions. The intuition is similar to before (when there was at most one abortion): Whereas almost as many girls as boys are born “after” two sex-selective abortions (cf. Fig. 4 at time  $T_b^0 + 2D$ ), no male but some female fetuses have been aborted before that. Consequently, more boys than girls are

born “without” an abortion (cf. Fig. 4 at time  $T_b^0$ ) and “with” one abortion (cf. Fig. 4 at time  $T_b^0 + D$ ).

Given Table 2, which is equivalent to Table 1 except that it considers the case with at most two abortions, it is easy to see that the probability that two abortions have occurred when the next-born child is a girl and when it is a boy are given by

$$P(A = 2|Y = 0) = \frac{P(A = 2, Y = 0)}{P(Y = 0)} = \frac{(1 - \pi)^2 \alpha_1 \alpha_2}{1 - \alpha_1 + (1 - \pi)(1 - \alpha_2)\alpha_1 + (1 - \pi)^2 \alpha_1 \alpha_2} \tag{7}$$

and

$$P(A = 2|Y = 1) = \frac{P(A = 2, Y = 1)}{P(Y = 1)} = \frac{(1 - \pi)^2 \alpha_1 \alpha_2}{1 + (1 - \pi)\alpha_1 + (1 - \pi)^2 \alpha_1 \alpha_2}, \tag{8}$$

respectively. While the numerators in Eqs. (7) and (8) are equal, the denominator in (7) is smaller and the conclusion follows.

Given that two abortions are more likely to have occurred when the next-born child is a girl and that two abortions prolong the time to birth by twice the time that one abortion does, the only way for the time to birth to be shorter when the next-born child is a girl is for one abortion (as opposed to two abortions) to be considerably less likely when the

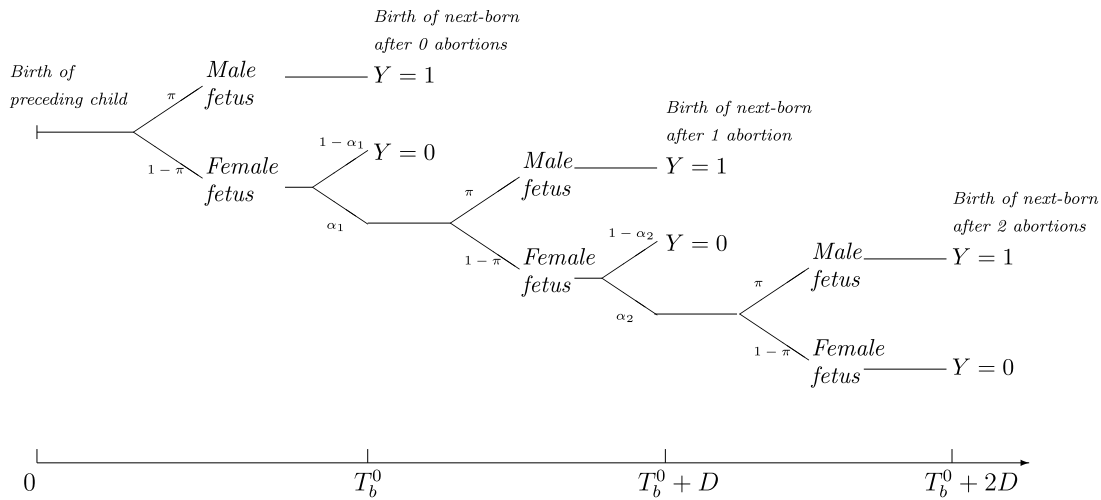


Fig. 4. Paths to birth with at most two abortions.

Note: The figure represents all possible outcomes in terms of sex and time to birth, as well as their respective probabilities, when households abort at most twice between two consecutive births.

**Table 2**  
Probabilities of sex and # abortions with at most two abortions.

$y$	$P(A = 0, Y = y)$	$P(A = 1, Y = y)$	$P(A = 2, Y = y)$	$P(Y = y) = \sum_{a=0}^2 P(A = a, Y = y)$
0	$(1 - \pi)(1 - \alpha_1)$	$(1 - \pi)^2 \alpha_1 (1 - \alpha_2)$	$(1 - \pi)^3 \alpha_1 \alpha_2$	$(1 - \pi)[1 - \alpha_1 + (1 - \pi)\alpha_1(1 - \alpha_2) + (1 - \pi)^2 \alpha_1 \alpha_2]$
1	$\pi$	$(1 - \pi)\alpha_1 \pi$	$(1 - \pi)^2 \alpha_1 \alpha_2 \pi$	$\pi[1 + (1 - \pi)\alpha_1 + (1 - \pi)^2 \alpha_1 \alpha_2]$

next-born child is a girl than when it is a boy. This occurs when a large share of female fetuses conceived after a first sex-selective abortion are aborted, i.e., when  $\alpha_2$  is large, as this results in a low probability of a girl being born “with” one abortion (cf. Fig. 4 and Table 2). Exact calculations, detailed in Appendix B, show that the difference in mean spacing between girls and boys,  $E(T_b|Y = 0) - E(T_b|Y = 1)$ , is negative, positive, or zero if and only if  $\alpha_2$  is greater than, less than, or equal to<sup>6</sup>

$$\frac{1 - 2(1 - \pi)\alpha_1 - \sqrt{1 - 4(1 - \pi)\alpha_1}}{2\alpha_1(1 - \pi)^2}, \tag{9}$$

respectively.

Fig. 5 plots Eq. (9) for  $\pi = 0.513$ . Values of  $\alpha_2$  above (below) the black line imply that time to birth is shorter (longer) when the next-born child is a girl. We see that  $\alpha_2$  indeed needs to be large for the time to birth to be shorter when the next-born child is a girl. In particular,  $\alpha_2$  needs to be greater than  $\alpha_1$ .

**GRAPHICAL ILLUSTRATION:** Fig. 6 graphically illustrates the conditional probabilities of our model for two different sets of parameter values when there are at most two sex-selective abortions. In panel (a),  $\alpha_1 = 0.8$  and  $\alpha_2 = 0.1$ , while in panel (b),  $\alpha_1 = 0.25$  and  $\alpha_2 = 0.99$ . In both panels, we observe that two abortions have a larger weight for girls than for boys—the top-right circles are larger than the bottom-right circles. In panel (a), the weight of one abortion is also larger for girls than for boys—the top-center circle is larger than the bottom-center circle. Consequently, the expected time to birth, represented by the black dot and obtained as a weighted average of the times of zero, one, and two abortions (see x-axis), is shorter for boys than for girls. In panel (b), however, the weight of one abortion is considerably larger for boys than for girls—the top-center circle is smaller than the bottom-center circle. As a result, the expected time to birth is shorter for girls than for boys.

<sup>6</sup> For  $\alpha_1 > \frac{1}{4(1-\pi)}$ ,  $E(T_b|Y = 0) - E(T_b|Y = 1)$  is strictly positive for any  $\alpha_2 \in [0, 1]$ .

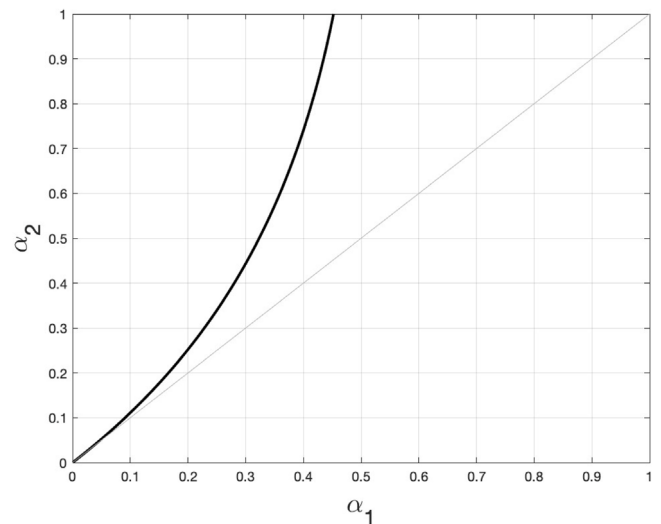


Fig. 5. Eq. (9).  
Note: The black line plots Eq. (9) for  $\pi = 0.513$ .

### 3.1.3. A useful proposition

If there is no sex-selective abortion, the time to birth is equal regardless of the sex of the next-born child. Furthermore, our model implies that if households abort at most once, the time to birth is longer when the next-born child is a girl (cf. Section 3.1.1). By contraposition, if the time to birth is shorter when the next-born child is a girl, this provides evidence of repeated sex-selective abortion between two consecutive births. In other words, there are households that repeatedly abort when repeatedly pregnant with a female fetus. The following proposition formalizes this result as well as the observation made at the beginning of Section 3. A proof is provided in Appendix A.

**Proposition 1.** (i) If  $E(T_b|Y = 0) \neq E(T_b|Y = 1)$ , then  $\alpha_1 > 0$ , i.e., there is sex-selective abortion.

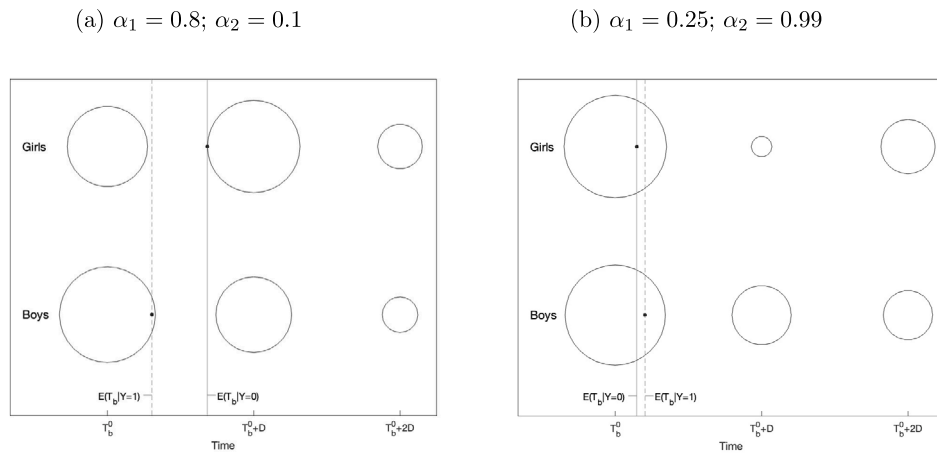


Fig. 6. Conditional probabilities with at most two abortions.

Note: In both panels, the top and bottom lines visualize  $P(A = a|Y = 0)$  and  $P(A = a|Y = 1)$  for  $a = 1, 2,$  and  $3$  from left to right, respectively, i.e., the areas of the circles represent conditional probabilities that should be thought of as summing to one for each line in each panel. The black dots indicate  $E(T_b|Y = y)$  for  $y = 0$  (Girls) and  $y = 1$  (Boys). In panel (a)  $\alpha_1 = 0.8; \alpha_2 = 0.1$  and in panel (b)  $\alpha_1 = 0.25; \alpha_2 = 0.99$ .

(ii) If  $E(T_b|Y = 0) < E(T_b|Y = 1)$ , then  $\alpha_1, \alpha_2 > 0$ , i.e., there is repeated sex-selective abortion.

Proposition 1(i) can be used to detect sex-selective abortion and therefore complements the sex ratio.<sup>7</sup> An advantage of using mean spacing as an indicator of sex-selective abortion is that it is independent of the natural sex ratio.<sup>8</sup> On the other hand, it is only informative about sex-selective abortion between two consecutive births and thus specific to a given birth interval.

Proposition 1(ii) can be used to detect repeated sex-selective abortion between two consecutive births. If the mean spacing for girls is lower than the mean spacing for boys, then there is repeated sex-selective abortion, i.e., a strictly positive fraction of the population aborts a second time if again pregnant with a female fetus. We note that Proposition 1 provides sufficient, but not necessary, conditions. For example, it is possible that repeated sex-selective abortion is present in the population, but that the mean spacing for girls is greater than the mean spacing for boys as illustrated in panel (a) of Fig. 6. In that sense, tests based on Proposition 1 are only partial.

### 3.2. Extension of the above model to varying times

So far, we have assumed that  $T_b^0$  and  $D$  are fixed for expositional purposes. In what follows, we model  $T_b^0$  and  $D$  as random variables, while maintaining that the probabilities  $P(A = a, Y = y)$  are given by Eqs. (1) and (2). This implicitly assumes that the decision to abort is independent of  $T_b^0$  and  $D$  (and vice versa). In order to facilitate the discussion of this assumption, we impose some structure on  $T_b^0$  and  $D$ . This structure also underlies our estimation procedure and ensures that our previous findings, including Proposition 1, continue to hold.

We assume that

$$T_b^0 = T_w + T_p,$$

where  $T_w$  denotes the waiting time until conception, which includes the time that the household waits after the birth of a child before trying to conceive again, plus the time it takes to conceive.  $T_p$  denotes the time

<sup>7</sup> We note that Proposition 1(i) also holds without conditioning on the sibling composition.

<sup>8</sup> More generally, our model implies that if  $P(T_b < t|Y = 0) \neq P(T_b < t|Y = 1)$  for at least some  $t$ , then  $\alpha_1 > 0$ . Therefore, an alternative, potentially more powerful test for the presence of sex-selective abortion would consist in testing whether the two conditional distributions are equal. Here, we focus on the “difference-in-means” for its simplicity.

of pregnancy, which we assume to be fixed and equal to nine months. Any potential variation in  $T_p$  is assumed to be independent of the sex of the fetus and is subsumed in  $T_w$ .

We assume that the  $k$ th abortion prolongs the birth interval by

$$T_s + T_{c,k}.$$

Here,  $T_s$  denotes the time to screening, which is the duration from conception until the sex of the fetus is known.<sup>9</sup> It represents the waiting period before a household may potentially perform a sex-selective abortion. We do not index  $T_s$  by the abortion cycle because we assume it is fixed and equal to five months.  $T_{c,k}$  denotes the time to conception after the  $k$ th abortion, which is the duration until a woman conceives again after having aborted  $k$  times. We assume  $T_{c,k}$  is i.i.d. across  $k$ , with  $T_{c,k} \sim T_c$ .<sup>10</sup>

Given the above assumptions, the time to birth with  $a$  abortions (having taken place) is given by

$$T_b^a = T_w + \sum_{k=1}^a (T_s + T_{c,k}) + T_p.$$

There are several assumptions underlying the above structure worth discussing. First, there is the assumption that possible variation in  $T_p$  is independent of the sex of the fetus. Possible variation in  $T_p$  might, for example, be due to premature births, which could be more likely for girls if women rest less when they know that they are expecting a girl. Similarly, one may fear that miscarriages and abortions not performed for the purpose of sex-selection are not equally likely for male and female fetuses (see e.g., Orzack et al., 2015). While such concerns may be theoretically justified, panel (a) of Fig. 1 suggests that they are not of first-order empirical relevance: when the observed proportion of male births is close to the natural probability of male birth, there is no difference in spacing by sex of the next-born child. Second, we implicitly assume that, after an abortion, the household tries to conceive again without any additional waiting time. This assumption seems plausible given that the decision to have another child has already been made before the abortion and that ovulation typically

<sup>9</sup> For simplicity, we assume that every household knows the sex of the fetus. However, the implications of the model remain the same if we assume that a household that does not, or would not, sex-selectively abort does not know the sex of the fetus.

<sup>10</sup> This assumption is supported by Oster (2022), who finds that time to conception at birth order 1 only explains 11% of the variation in time to conception at higher birth orders.



returns within one month after an abortion (see, e.g., the “Abortion Care guideline” of the [World Health Organization, 2022](#)).

Given the above structure, the aforementioned independence assumption implies that the decision to abort is independent of  $T_w$  and  $T_c$ , which constitute the random parts of  $T_b^0$  and  $D$ , respectively. This assumption is not innocuous because  $T_w$  includes the time to conception, and it is plausible that a household is less likely to abort a female fetus if it took a long time to conceive.<sup>11</sup> Similarly, the waiting time component of  $T_w$  may be correlated with the decision to abort.

For instance, households with a son preference but low willingness to abort (WTA) may follow a fertility stopping rule (see e.g., [Cassan et al., 2023](#)). These households may have shorter waiting times in anticipation of potentially needing a larger number of children to have at least one boy. Other dependencies through WTA are also conceivable. If abortions impact the time to conception (contained in  $T_w$  and the  $T_{c,k,s}$ ), then a household with a higher WTA (and thus a greater likelihood to abort) may have a differential time to conception, as it is more likely to already have performed an abortion during a previous birth interval.

While we leave the task of relaxing the independence assumption for future research, we estimate the parameters of our model across birth orders and sibling compositions, as well as separately for different socio-demographic and geographic groups. This approach ensures that the women in our estimation samples are more homogeneous in terms of waiting times and unobserved factors.

As mentioned earlier, we implicitly assume no sex-selective abortion of male fetuses. However, as long as female fetuses are more likely to be aborted than male fetuses, the above model remains applicable with the understanding that the  $\alpha$ s represent differences in the probabilities of aborting female and male fetuses. The additional time to birth that is due to sex-selective abortion and that is common to fetuses of both sexes can be subsumed in  $T_w$  and  $T_c$ .

To see that our previous findings, including [Proposition 1](#), continue to hold under varying times, note that, given the above assumptions, we have

$$E(T_b|Y = y) = T_p + E(T_w) + \sum_{a=1}^{N_0} P(A = a|Y = y) \cdot a \cdot (T_s + E(T_{c,k})),$$

which has the same structure as Eqs. (3) and (6) (with  $T_b^0 = T_p + E(T_w)$  and  $D = T_s + E(T_{c,k})$ ) that underlie the proof of [Proposition 1](#) and the analysis in [Section 3.1.2](#).

### 3.3. Maximum likelihood estimation

In order to estimate the model and its main parameters,  $\alpha = (\alpha_1, \dots, \alpha_{N_0})'$ , we rely on the principle of maximum likelihood. We assume that  $T_w$  and  $T_c$  are independent and gamma-distributed, with shape parameters  $\gamma_w$  and  $\gamma_c$  and scale parameters  $\beta_w$  and  $\beta_c$ .<sup>12</sup> This distributional assumption is not only flexible but also allows us to obtain an analytical expression for the likelihood function, avoiding computationally expensive simulation-based estimation. In particular, the individual likelihood function is given by

$$L(\theta; t, y) = \sum_{a=0}^{N_0} f_{T_b^a}(t; \gamma_w, \beta_w, \gamma_c, \beta_c) P(Y = y, A = a; \alpha),$$

<sup>11</sup> For example, a long time to conceive may increase the emotional attachment to the female fetus. It may also impact beliefs about future times to conception, including the  $T_{c,k,s}$ .

<sup>12</sup> The independence assumption seems plausible given that times to conception are only weakly auto-correlated within households (see footnote 10), making them difficult to predict. This also suggests that households cannot effectively adjust their waiting times.

where  $\theta = (\alpha', \gamma_w, \beta_w, \gamma_c, \beta_c)'$  and where

$$f_{T_b^a}(t; \gamma_w, \beta_w, \gamma_c, \beta_c) = \begin{cases} f_{\Gamma}(t - t_p; \gamma_w, \beta_w) & \text{if } a = 0 \\ f_{\Sigma\Gamma}(t - t_p - a \cdot t_s; \gamma_w, \beta_w, \gamma_c, \beta_c) & \text{if } a > 0 \end{cases}$$

denotes the probability density function (pdf) of  $T_b^a$  for a given  $a$ . Here,  $f_{\Gamma}(\cdot; \gamma, \beta)$  and  $f_{\Sigma\Gamma}(\cdot; \gamma_1, \beta_1, \gamma_2, \beta_2)$  denote the pdf of a gamma-distributed random variable (with shape parameter  $\gamma$  and scale parameter  $\beta$ ) and the pdf of the sum of two independent gamma-distributed random variables (with shape parameters  $\gamma_1$  and  $\gamma_2$ , and scale parameters  $\beta_1$  and  $\beta_2$ ), respectively; see [Appendix C](#) for details. Note that  $P(Y = y, A = a; \alpha)$ , where we have made the dependence on  $\alpha$  explicit, is given in (1) and (2) for  $y = 0$  and  $y = 1$ , respectively.

With  $L(\theta; t, y)$  thus defined, we could, in principle, estimate  $\theta$  by maximizing the log-likelihood function,  $\sum_{i=1}^n \log L(\theta; t_i, y_i)$ , over the parameter space  $[0, 1]^{N_0} \times [0, C]^4$  for some large  $C < \infty$ , where  $i = 1, \dots, n$  indexes the households in the estimation sample. To aid estimation, we instead calibrate  $\gamma_c$  and  $\beta_c$  using auxiliary data; see [Appendix C](#) for details. Furthermore, we impose that  $N_0 = 2$  and, with a slight abuse of notation, let  $\theta = (\alpha_1, \alpha_2, \gamma_w, \beta_w)'$  with the understanding that  $L(\theta; t_i, y_i)$  denotes  $L((\theta', \tilde{\gamma}_c, \tilde{\beta}_c)'; t_i, y_i)$ , where  $\tilde{\gamma}_c$  and  $\tilde{\beta}_c$  denote the calibrated values of  $\gamma_c$  and  $\beta_c$ , respectively. Formally, our maximum likelihood estimator is defined as

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \sum_{i=1}^n \log L(\theta; t_i, y_i),$$

where  $\Theta = [0, 1]^2 \times [0, C]^2$ .

An important comment about our estimation problem is in order:  $\alpha_2$  is not identified when  $\alpha_1$  is equal to zero, in the sense that  $\sum_{i=1}^n \log L(\theta; t_i, y_i)$  does not depend on  $\alpha_2$  when  $\alpha_1 = 0$ ; cf. Assumption A in [Andrews and Cheng \(2012\)](#) (AC hereafter). Intuitively, we cannot identify the share of women who abort a second time if none abort a first time. By continuity,  $\alpha_2$  is only weakly identified when  $\alpha_1$  is close to zero relative to the sample size,  $n$ . More precisely, the identification strength of  $\alpha_2$  is governed by  $\sqrt{n}\alpha_1$ :  $\alpha_2$  is weakly identified if  $\sqrt{n}\alpha_1 \rightarrow h$  (as  $n \rightarrow \infty$ ) with  $h < \infty$ , while  $\alpha_2$  is (semi-) strongly identified if  $\sqrt{n}\alpha_1 \rightarrow h$  with  $h = \infty$ , using the terminology in AC.<sup>13</sup> The possibility that  $\alpha_2$  may only be weakly identified implies that “standard” standard errors may not be reliable.

In this paper, we use an identification-category-selection (ICS) procedure to address the problem of weak identification (see, e.g., AC). To select an identification category (i.e.,  $h < \infty$  or  $h = \infty$ ), an ICS statistic is compared to a tuning parameter,  $\kappa$ , that satisfies  $\kappa \rightarrow \infty$  and  $\kappa/\sqrt{n} \rightarrow 0$  as  $n \rightarrow \infty$ ; a standard choice is  $\kappa = \sqrt{\ln n}$ . If  $\text{ICS} \leq \kappa$ , one concludes that identification is weak; if  $\text{ICS} > \kappa$ , one concludes that identification is (semi-) strong. An ICS procedure is consistent if the ICS statistic is  $O_p(1)$  under weak identification and diverges under (semi-) strong identification. Here, we use the standard t-statistic for testing  $H_0 : \alpha_1 = 0$ , say  $t_1$ , as the ICS statistic. Our ICS procedure involves only reporting estimates of and making inferential statements about  $\alpha_2$  when  $t_1 > \kappa$ .<sup>14,15</sup> In order to be conservative, we take  $\kappa = 6$ .<sup>16</sup>

<sup>13</sup>  $\alpha_2$  is semi-strongly identified if, in addition,  $\alpha_1 \rightarrow 0$  and strongly identified otherwise. Here, we implicitly consider (drifting) sequences of true parameters whose dependence on the sample size is omitted for notational convenience; see, e.g., [Andrews and Guggenberger \(2010\)](#) for the importance of the asymptotic behavior of estimators and test statistics under drifting sequences of true parameters for determining the asymptotic size of a test.

<sup>14</sup> Constructing tests that are valid under weak identification ( $t_1 \leq \kappa$ ) is beyond the scope of this paper.

<sup>15</sup> From results in [Ketz \(2019, 2022\)](#), it can be deduced that (the confidence interval based on) the standard t-statistic for  $\alpha_2$  is valid under (semi-) strong identification, in that its asymptotic size does not exceed the nominal level.

<sup>16</sup> We do not use  $\kappa = \sqrt{\ln n}$  because we are concerned that this choice might lead us to concluding that identification is (semi-) strong even when it is not. Even for our largest sample (cf. [Table 4](#)), we “only” have  $\sqrt{\ln(232,935)} \approx$

### 3.4. Parameters of interest

In this section, we introduce the number of aborted female fetuses, which corresponds to the number of missing women of Sen (1990) adapted to our setting. We then show how  $\alpha_2$  captures the relationship between the number of aborted female fetuses and the number of women who abort when there are at most two sex-selective abortions between two consecutive births. Finally, we note that the product of  $\alpha_1$  and  $\alpha_2$  measures how likely women are to perform repeated sex-selective abortion between two consecutive births.

As underlined in several places in this paper, we are interested in (repeated) sex-selective abortion *between two consecutive births*. A population of interest is therefore defined by birth order (and other criteria such as sibling composition). The birth interval associated with a given birth order refers to the interval between the preceding birth and the birth indicated by the birth order. For example, at birth order 2, we consider the time that elapses between the birth of the first and second child. In what follows, a “sample” should be thought of as a random sample from a population of interest.

In a sample of  $n$  households, the expected number of pregnancies with a female fetus (ENFF) is equal to

$$ENFF = n \times PMB \frac{1 - \pi}{\pi},$$

where PMB denotes the proportion of male births observed in the sample. This number is defined analogous to the expected number of women in Sen (1990) and the expected number of female births in Guilmoto et al. (2020). It tells us how many female fetuses we would expect to be conceived (but not necessarily to be born) given the observed number of male births,  $n \times PMB$ . The observed number of female births (ONFB) is given by

$$ONFB = n \times (1 - PMB)$$

and defined analogous to the observed number of women in Sen (1990) and the observed number of female births in Guilmoto et al. (2020). The difference between ENFF and ONFB gives us the (expected) number of aborted female fetuses (ENAFF), i.e.,

$$ENAFF = ENFF - ONFB = n \frac{PMB - \pi}{\pi}.$$

We note that ENAFF, in fact, constitutes an *estimator* of the (expected) number of aborted female fetuses. The corresponding population analogue is obtained by replacing the observed proportion of male births, PMB, by the unknown probability of male births,  $P(Y = 1)$ . When there are at most two sex-selective abortions, the latter is equal to

$$\pi[1 + (1 - \pi)\alpha_1 + (1 - \pi)^2\alpha_1\alpha_2],$$

see Table 2. Plugging in and dividing by the number of women who abort,  $n(1 - \pi)\alpha_1$ , we obtain

$$1 + (1 - \pi)\alpha_2. \tag{10}$$

We have thus shown that  $\alpha_2$  provides a direct link between the number of aborted female fetuses—the standard measure of sex-selective abortion—and the number of women who abort. When  $\alpha_2$  equals zero, no woman aborts more than once, and the number of aborted female fetuses is equal to the number of women who abort. When  $\alpha_2$  equals one, the number of aborted female fetuses is  $2 - \pi \approx 1.5$  times larger than the number of women who abort, i.e., there are approximately 50% more aborted female fetuses than women who abort.

3.5. The value 6 is inspired by Elliott et al. (2015), who use this number in a related context; see their running example. In unreported Monte Carlo simulations (available upon request), we have also found that the t-test for testing hypotheses about  $\alpha_2$  has good size properties under realistic choices of the parameter values, such that  $t_1$  is approximately normally distributed across simulations, and less than 0.5% of the simulations have  $t_1 > 6$ .

While  $\alpha_1$  and  $\alpha_2$  are defined as probabilities, they can equally be thought of as shares in the population.  $\alpha_1$  is the share of women who abort if the first pregnancy since the preceding birth is with a female fetus, and  $\alpha_2$  is the share of women who abort a second time if again pregnant with a female fetus after having performed a sex-selective abortion. Since the women who abort a second time have necessarily aborted a first time,  $\alpha_1$  can also be interpreted as the share of women who abort at least once if “necessary”. For brevity, we refer to  $\alpha_1$  as the “share of women who abort”. Similarly, we refer to  $\alpha_2$  as the “share of women who abort a second time”. Additionally,  $\alpha_1 \times \alpha_2$  measures the likelihood of women performing *repeated* sex-selective abortion between two consecutive births. It can be thought of as the share of women who abort twice if “necessary”. This means they abort twice if repeatedly pregnant with a female fetus but only once if pregnant with a male fetus after a first sex-selective abortion, and never if the first pregnancy is with a male fetus. In what follows, we refer to  $\alpha_1 \times \alpha_2$  as the “share of women at risk of multiple abortions”.

### 4. Data description

We use data from the five rounds of the Demographic and Health Surveys (DHS 1–5) for India (1992–93, 1998–99, 2005–06, 2015–16, and 2019–2021)<sup>17</sup> and the 2002–2004 round of the District Level Household & Facility Survey (DLHS 2). The DHS and DLHS provide retrospective but precise information on the fertility (full birth history) of each woman aged between 15 and 49 years at the time of the survey.

We investigate the presence of (repeated) sex-selective abortion between marriage and the birth of the third child, i.e., at birth orders 1–3. As misreporting of births has been shown to increase with the recall period (Schoumaker, 2014; Pörtner, 2022), we exclude birth intervals that started more than 15 years before the survey date. To mitigate sample selection issues due to censoring, we further restrict our analysis to birth intervals that began more than five years before the survey date. Table 3 illustrates how the sample size evolves as we progressively apply these criteria.

An implicit assumption of our analysis is that the probability of observing a next-born child is not influenced by sex-selective abortion. However, women who perform sex-selective abortions are likely over-represented among those missing from our sample due to very long birth intervals, as illegally performed abortions can lead to fertility issues, including a delayed restart of ovulation, and even maternal death. Consequently, our reported estimates of the shares of women who abort may underestimate the corresponding population shares of interest. However, we believe these biases are relatively small. For instance, abortion-related mortality constitutes “only” 5% of all maternal deaths (Meh et al., 2022).

In order to study the evolution of sex-selective abortion patterns over time, we divide our sample into three 10-year periods, 1985–1994, 1995–2004, and 2005–2014.<sup>18</sup> The first period is pre-ban, with low availability of ultrasounds, and the last two periods are post-ban, with high availability of ultrasounds; see Section 2 for details.

### 5. Empirical analysis

#### 5.1. Results based on Proposition 1

First, we use Proposition 1 to explore what can be learned from differences in average spacing regarding the presence of (repeated) sex-selective abortion in various groups over time. To define our groups of interest, we rely on existing literature that has identified different patterns of sex-selective abortion across education levels and geographical

<sup>17</sup> DHS 1–4 were downloaded from IPUMS DHS (Boyle et al., 2022) and DHS 5 from the DHS website.

<sup>18</sup> The years refer to the start of the birth interval.

**Table 3**  
Sample description.

Data source	Survey year	# women <sup>a</sup>	# births at birth orders 1–3				
			Total <sup>b</sup>	5–15 years <sup>c</sup>	Spacing <sup>d</sup>	Our sample <sup>e</sup>	in %
DHS 1	1992–93	78,287	113,273	48,293	48,241	12,181	10.75
DHS 2	1998–99	79,759	192,660	76,385	68,355	62,698	32.54
DLHS 2	2002–04	500,767	609,146	293,345	292,913	292,913	48.09
DHS 3	2005–06	83,342	193,893	75,091	71,177	71,177	36.71
DHS 4	2015–16	469,881	1,022,017	377,914	369,638	369,638	36.17
DHS 5	2019–21	494,019	1,073,352	384,408	376,444	370,749	34.54
Total		1,706,055	3,204,341	1,255,436	1,226,768	1,179,356	36.80

<sup>a</sup> We exclude women who had twins.

<sup>b</sup> We only keep births with complete information on sex, sex of previous sibling, and spacing.

<sup>c</sup> We only keep birth intervals that started between 5 and 15 years before the survey date.

<sup>d</sup> We only keep birth intervals that are between 9 months and 10 years and 9 months.

<sup>e</sup> We only keep birth intervals that started after 1984 and before 2015.

locations (Bhalotra and Cochrane, 2010; Saikia et al., 2021; Pörtner, 2022). Specifically, we consider literate and illiterate women, women living in urban and rural areas, and women living in the North and South of India.<sup>19</sup> In the main body of the paper, we focus on “only girls” sibling compositions, as sex-selective abortion is more likely to occur in households where no boys have been born yet (Jayachandran, 2017). The corresponding results for other sibling compositions are reported in Appendix D.

Table 4 reports the proportions of male births (PMB) along with the differences in average spacing between girls and boys (DAS), i.e., average spacing when the next-born is a girl minus average spacing when the next-born is boy, for our different groups by period, birth order, and sibling composition. For birth order 1, PMB and DAS largely agree on the presence of sex-selective abortion. For most groups in the first two periods, PMB is not significantly different from the natural probability of male birth,  $\pi$ , which we take as 0.513 (cf. Chao et al., 2019; Dubuc and Coleman, 2007), and DAS is not significantly different from zero. In the last period, PMB suggests the presence of sex-selective abortion, consistent with the findings in Saikia et al. (2021). However, DAS does not provide additional evidence of sex-selective abortion during that period.

For birth orders 2 and 3 with “only girls” sibling compositions, PMB provides evidence of sex-selective abortion for almost all periods and all groups, with a greater imbalance at birth order 3. DAS complements this picture by showing some evidence of repeated sex-selective abortion at birth order 2 in the first two periods, and strong evidence of repeated sex-selective abortion at birth order 3, being statistically different from zero and negative for many periods and groups.

Table D.1 in Appendix D presents the corresponding results for sibling compositions that include boys. Both PMB and DAS suggest some evidence of sex-selective abortion at birth order 2 in the first period when the firstborn is a boy. At birth order 3, PMB suggests some evidence of sex-selective abortion for sibling compositions that include one girl and one boy. However, the values of PMB are generally smaller than those in Table 4 for the “only girls” sibling compositions, and DAS shows very little evidence of (repeated) sex-selective abortion.

For the “only boys” sibling compositions at birth order 3, PMB is significantly smaller than  $\pi = 0.513$  for many groups. Since in most cases DAS is not significantly different from zero, a potential explanation could be that the natural probability of female births increases with the age of the parents, consistent with findings in the

<sup>19</sup> The North of India is defined as the collection of the following states: Chandigarh, Delhi, Haryana, Himachal Pradesh, Jammu and Kashmir, Ladakh, Punjab, Rajasthan, Uttarakhand, Uttar Pradesh, Bihar, Jharkhand, Madhya Pradesh, West Bengal, and Gujarat. The remaining states make up the South of India.

literature (Jacobsen et al., 1999; Mathews and Hamilton, 2005; Matsuo et al., 2009; Nicolich et al., 2000). Given that our estimation procedure uses  $\pi = 0.513$  regardless of the parents’ age, the subsequent results may therefore potentially underestimate the incidence of sex-selective abortion at higher birth orders.

## 5.2. Estimation results

To gain further insights into the prevalence of sex-selective abortion, we apply our maximum likelihood estimator. Estimation is conducted on the samples with the highest proportions of male births, specifically at birth orders 2 and 3 when all previous children are girls (see Tables 4 and D.1 in Appendix D).

### 5.2.1. Shares of women who abort

We first present our estimates of the shares of women who abort,  $\hat{\alpha}_1$ . Fig. 7 shows the estimates at birth order 2 when the firstborn is a girl, and Fig. 8 shows the estimates at birth order 3 when the first two children are girls. Both figures include 95% confidence intervals for  $\alpha_1$ . The corresponding t-statistics for testing  $H_0 : \alpha_1 = 0$ ,  $t_1$ , which serve as our ICS statistics, are provided in Table D.2 in Appendix D. We note that the confidence intervals in Figs. 7 and 8 should be interpreted with caution, especially for small values of  $t_1$ , as they do not account for the possibility that  $\alpha_2$  may be weakly identified.<sup>20</sup>

Panel (a) of Fig. 7 shows that the share of women who abort at birth order 2 gradually increases over time from approximately 5% in the first period to around 8% in the last period. Panels (b) to (d) reveal significant heterogeneity across socio-demographic and geographical groups. Specifically, literate women and those in urban areas are much more likely to abort than illiterate women and those in rural areas. The share of women who abort among illiterate women remains small and relatively constant over time, while there is a slight increase among rural women, reaching about 6% in the last period compared to approximately 16% among urban women. Lastly, women in the North are estimated to be more likely to abort than women in the South, though the estimates are relatively close.

At birth order 3, the picture is qualitatively similar but with higher shares of women who abort (see Fig. 8). There is an increase in the share of women who abort over time, from approximately 7% in the first period to about 14% in the last period, which may be partly

<sup>20</sup> The underlying standard errors are based on a numerical approximation of the second-order derivative matrix of  $\sum_{i=1}^n \log L(\theta; t_i, y_i)$  evaluated at  $\hat{\theta}$ . This numerical approximation becomes unreliable or even infeasible when  $\hat{\alpha}_1$  is very close to zero. Nonetheless, we were able to compute it for all estimation results reported in Figs. 7 and 8 using the DERIVESTsuite package for Matlab.

**Table 4**  
Proportions of male births and differences in average spacing.

	Pooled	By education		By urban-rural status		By region	
		Illiterate	Literate	Rural	Urban	South	North
<b>Birth order 1</b>							
[1985–1994] PMB	0.513 (0.003)	0.509 (0.004)	0.517 (0.004)	0.511 (0.004)	0.518 (0.005)	0.509 (0.004)	0.517 (0.004)
DAS	0.10 (0.23)	0.13 (0.34)	–0.01 (0.30)	–0.05 (0.29)	0.30 (0.37)	0.50 (0.31)	–0.19 (0.32)
<i>n</i>	[28,602]	[14,009]	[14,564]	[18,673]	[9,929]	[13,288]	[15,314]
[1995–2004] PMB	0.515 (0.002)	0.510 (0.003)	0.517 (0.002)*	0.513 (0.002)	0.519 (0.003)*	0.514 (0.003)	0.515 (0.002)
DAS	0.17 (0.14)	0.44 (0.25)*	–0.02 (0.16)	0.28 (0.17)*	–0.12 (0.23)	–0.33 (0.20)*	0.52 (0.19)***
<i>n</i>	[75,806]	[27,459]	[48,170]	[51,532]	[24,274]	[30,558]	[45,248]
[2005–2014] PMB	0.519 (0.001)***	0.519 (0.002)***	0.520 (0.001)***	0.517 (0.001)***	0.526 (0.002)***	0.517 (0.002)***	0.521 (0.001)***
DAS	0.05 (0.07)	0.03 (0.14)	0.06 (0.08)	0.01 (0.08)	0.13 (0.13)	–0.05 (0.10)	0.15 (0.10)
<i>n</i>	[232,935]	[65,206]	[167,353]	[172,778]	[60,157]	[95,094]	[137,841]
<b>Birth order 2 - G</b>							
[1985–1994] PMB	0.527 (0.002)***	0.519 (0.002)**	0.535 (0.003)***	0.523 (0.002)***	0.534 (0.003)***	0.520 (0.003)***	0.531 (0.002)***
DAS	–0.41 (0.13)***	–0.28 (0.17)*	–0.46 (0.20)**	–0.32 (0.15)**	–0.54 (0.25)**	–0.13 (0.21)	–0.67 (0.16)***
<i>n</i>	[77,934]	[40,349]	[37,553]	[53,029]	[24,905]	[33,157]	[44,777]
[1995–2004] PMB	0.532 (0.002)***	0.517 (0.003)	0.544 (0.003)***	0.527 (0.002)***	0.544 (0.004)***	0.525 (0.003)***	0.537 (0.003)***
DAS	–0.36 (0.14)**	–0.63 (0.19)***	0.04 (0.20)	–0.34 (0.16)**	–0.26 (0.28)	0.08 (0.23)	–0.73 (0.18)***
<i>n</i>	[65,967]	[28,854]	[36,994]	[46,004]	[19,963]	[27,072]	[38,895]
[2005–2014] PMB	0.538 (0.002)***	0.522 (0.003)***	0.546 (0.002)***	0.533 (0.002)***	0.551 (0.003)***	0.532 (0.002)***	0.541 (0.002)***
DAS	0.02 (0.12)	0.06 (0.18)	0.16 (0.15)	0.03 (0.13)	0.20 (0.26)	0.27 (0.20)	–0.21 (0.14)
<i>n</i>	[109,392]	[37,608]	[71,582]	[83,369]	[26,023]	[43,434]	[65,958]
<b>Birth order 3 - GG</b>							
[1985–1994] PMB	0.539 (0.003)***	0.520 (0.004)*	0.568 (0.004)***	0.532 (0.003)***	0.557 (0.005)***	0.531 (0.004)***	0.545 (0.004)***
DAS	–1.15 (0.20)***	–0.36 (0.23)	–1.89 (0.35)***	–0.79 (0.22)**	–1.83 (0.42)***	–0.50 (0.33)	–1.63 (0.25)***
<i>n</i>	[31,258]	[18,525]	[12,721]	[22,162]	[9,096]	[12,410]	[18,848]
[1995–2004] PMB	0.550 (0.003)***	0.529 (0.004)***	0.575 (0.005)***	0.539 (0.004)***	0.581 (0.006)***	0.540 (0.005)***	0.555 (0.004)***
DAS	–1.27 (0.23)***	–0.75 (0.29)***	–1.42 (0.37)***	–0.89 (0.26)***	–1.95 (0.51)***	–0.47 (0.40)	–1.78 (0.28)***
<i>n</i>	[24,807]	[13,383]	[11,383]	[18,266]	[6,541]	[9,030]	[15,777]
[2005–2014] PMB	0.561 (0.003)***	0.533 (0.004)***	0.587 (0.003)***	0.556 (0.003)***	0.585 (0.006)***	0.556 (0.004)***	0.564 (0.003)***
DAS	–0.98 (0.20)***	–0.63 (0.26)**	–0.75 (0.29)***	–0.85 (0.21)***	–1.10 (0.50)**	–0.30 (0.35)	–1.36 (0.23)***
<i>n</i>	[38,418]	[17,768]	[20,575]	[31,040]	[7,378]	[12,917]	[25,501]

The table shows the proportions of male births (PMB) and the differences in average spacing between girls and boys (DAS), i.e., average spacing when the next-born is a girl minus average spacing when the next-born is a boy, for different samples: G and GG denote “first child is a girl” and “first two children are girls”, respectively. Standard errors are reported in parentheses. *p*-values are for two-sided *t*-tests for testing that the probability of male birth equals  $\pi$  and that the difference in mean spacing equals 0.

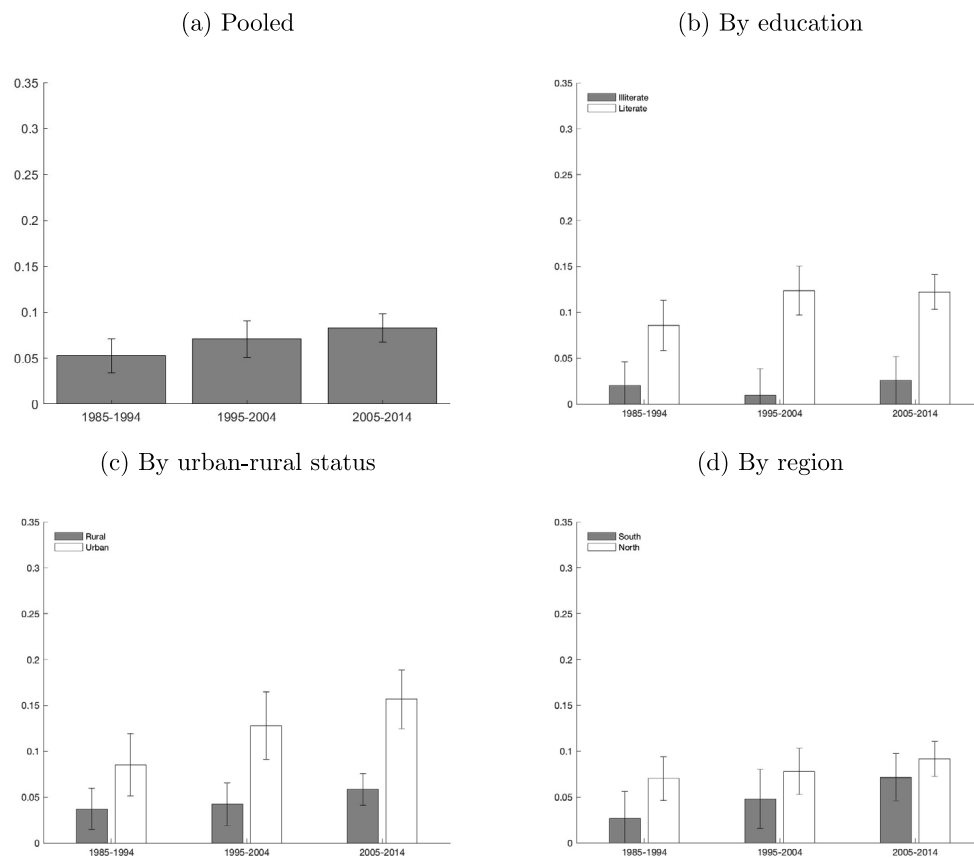
\*  $p < 0.1$ .

\*\*  $p < 0.05$ .

\*\*\*  $p < 0.01$ .

due to the increased availability of ultrasounds. Again, the decomposition by socio-demographic and geographical groups reveals significant heterogeneity. Literate women and those in urban areas are more likely to abort, and in both groups, the share of women who abort has

increased by more than 9 percentage points from the first to the last period, rising from around 14% (12%) to approximately 23% (25%) for literate women (urban women). As before, there is little difference in the share of women who abort between the North and the South.



**Fig. 7.** Shares of women who abort at birth order 2 when the firstborn is a girl.  
 Note: Each panel shows the estimated shares of women who abort,  $\hat{\alpha}_1$ , together with 95% confidence intervals for the three time periods.

**5.2.2. Shares of women who abort a second time and shares of women at risk of multiple abortions**

Next, we present our estimates of the shares of women who abort a second time,  $\hat{\alpha}_2$ , and the shares of women at risk of multiple abortions,  $\hat{\alpha}_1 \times \hat{\alpha}_2$ . As discussed in Section 3.3, we only consider samples for which  $t_1 > 6$ . Specifically, we focus on literate women, women in urban areas, and women in the North during the last two periods, at birth orders 2 and 3 (see Table D.2). The results are graphically depicted in Fig. 9, with a tabulated version, including standard errors, provided in Table D.3 in Appendix D.

For each sample, Fig. 9 shows three quantities: the estimated shares of women who abort ( $\hat{\alpha}_1$ , gray bar), already reported in Figs. 7 and 8; the estimated shares of women who abort a second time ( $\hat{\alpha}_2$ , white bar); and the estimated shares of women at risk of multiple abortions ( $\hat{\alpha}_1 \times \hat{\alpha}_2$ , black bar). The first striking result is that at birth order 2, women in some groups do not abort a second time ( $\hat{\alpha}_2 \approx 0$ ). In other words, there is no repeated sex-selective abortion. This is the case for literate women and those in urban areas, providing evidence against the assumption, sometimes made in the literature, that women abort indefinitely until they have a boy (see, e.g., Guilmoto et al., 2020).

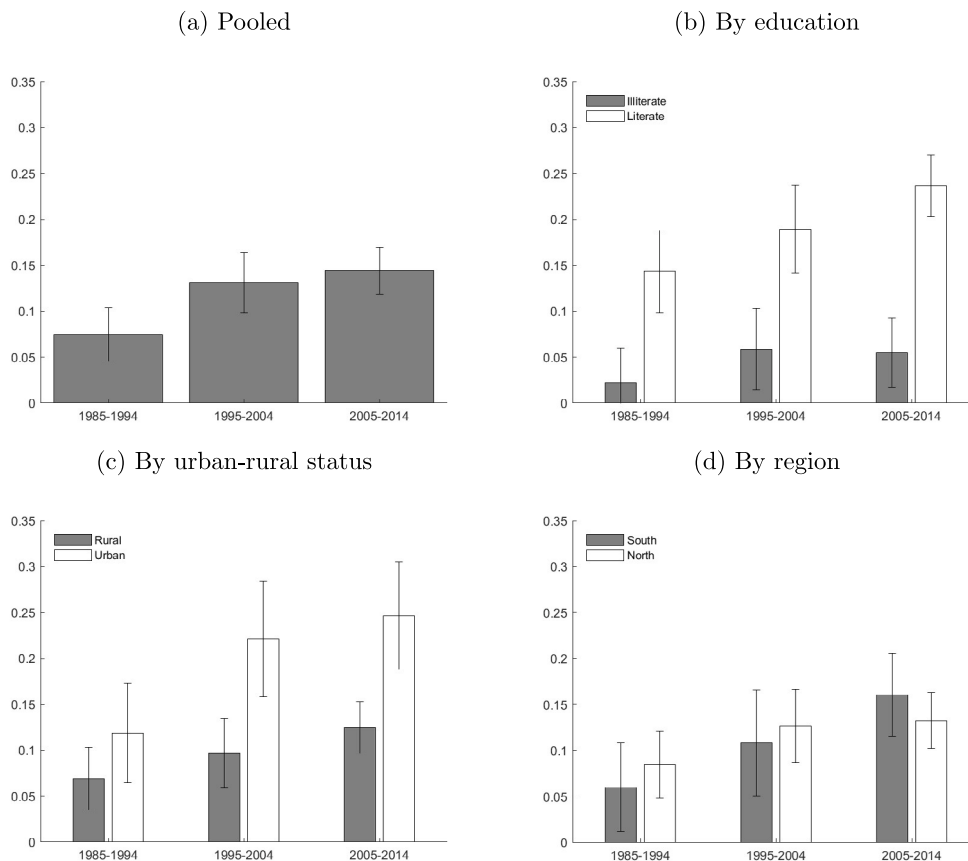
Second, we find strong evidence of repeated sex-selective abortion at birth order 2 for women in the North and at birth order 3 for all groups, i.e., literate women, urban women, and women in the North. In both periods, the estimated shares of women who abort a second time are large and significantly different from zero at the 10% significance level. While the estimates suggest that the shares of women who abort a second time have been stable over time for literate and urban women, there appears to be an increase for women in the North at birth orders 2 and 3.

Relatedly, we observe that the estimated shares of women at risk of multiple abortions have remained stable over time for literate and urban women at birth order 3, at approximately 9%. Among women in the North, our estimates indicate that the shares of women at risk of multiple abortions have increased over time both at birth orders 2 and 3. At birth order 3, the share is even estimated to have doubled, from approximately 6.5% to around 13%.

Taken together, these estimates show considerable variation across groups in the shares of women who abort, as well as in the shares of women who abort a second time and those at risk of multiple abortions. Moreover, the estimates are relevant from a public health perspective. For example, if there are increased health risks associated with repeated sex-selective abortions, women in the North are at a higher risk than literate and urban women. While the latter groups have larger estimated shares of women who abort, women in the North have a larger estimated share at risk of multiple abortions, notably in the last period, both at birth orders 2 and 3.

As shown in Eq. (10),  $\alpha_2$  captures the difference between the number of aborted female fetuses and the number of women who abort. To provide a sense of the magnitudes of these numbers, Table 5 reports the corresponding estimates per 1000 women for the same samples considered in Fig. 9. These estimates are obtained as ENAFF (see Section 3.4) and  $n(1 - \pi)\hat{\alpha}_1$  with  $n = 1000$ , respectively. The table also reports the ratio of the two estimates.

The estimated numbers of aborted female fetuses vary between 47 and 143, while the estimated numbers of women who abort range from 38 to 120. For some groups, such as literate women or urban women at birth order 2, the estimates are not statistically significantly different from each other. However, for other groups, the number of aborted



**Fig. 8.** Shares of women who abort at birth order 3 when the first two children are girls.  
 Note: Each panel shows the estimated shares of women who abort,  $\hat{\alpha}_1$ , together with 95% confidence intervals for the three time periods.

**Table 5**  
 Numbers of aborted female fetuses and numbers of women who abort.

	Birth order	Period	# aborted female fetuses (I)	# women who abort (II)	Ratio (I/II)
Literate	2	1995–2004	61	60	1.02
		2005–2014	64	59	1.08
	3	1995–2004	121	92	1.32**
		2005–2014	143	115	1.24***
Urban	2	1995–2004	61	62	0.98
		2005–2014	75	76	0.99
	3	1995–2004	132	108	1.22*
		2005–2014	141	120	1.18*
North	2	1995–2004	47	38	1.24*
		2005–2014	55	45	1.22***
	3	1995–2004	83	62	1.34**
		2005–2014	99	65	1.52***

The table shows the estimated numbers of aborted female fetuses (I) and the estimated numbers of women who abort (II) per 1000 women for selected samples. It also reports the ratio of the two (I/II).  $p$ -values are for the one-sided t-test of  $H_0 : \alpha_2 = 0$  vs.  $H_1 : \alpha_2 > 0$ . Given Eq. (10), this test also tests the null hypothesis that the ratio (I/II) is equal to 1.

\*  $p < 0.1$ .  
 \*\*  $p < 0.05$ .  
 \*\*\*  $p < 0.01$ .

female fetuses is found to be statistically significantly greater than the number of women who abort. This is the case, for example, for women in the North at birth order 3, for whom we obtain the largest ratio. In

the last period, the estimated number of aborted female fetuses is equal to 99, approximately 50% larger than the estimated number of women who abort (65).

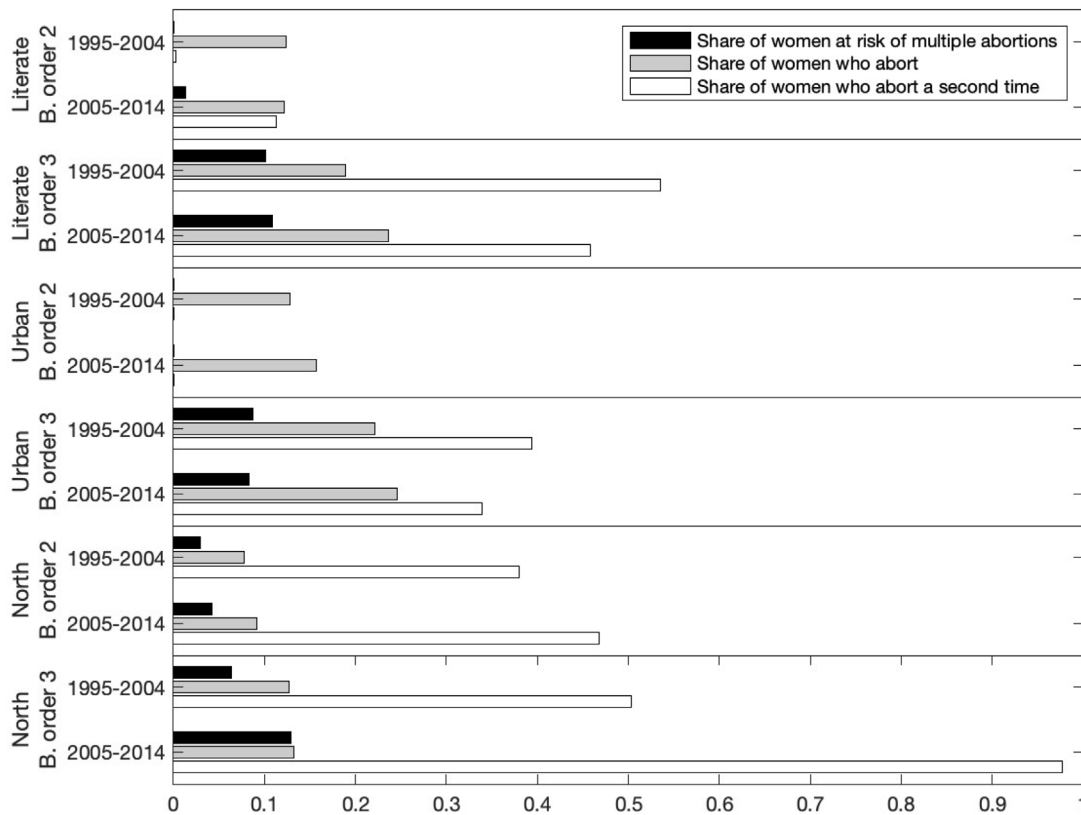


Fig. 9. Shares of women at risk of multiple abortions, shares of women who abort, and shares of women who abort a second time for selected samples. Note: The figure shows the estimated shares of women at risk of multiple abortions,  $\hat{a}_1 \times \hat{a}_2$ , the estimated shares of women who abort,  $\hat{a}_1$ , and the estimated shares of women who abort a second time,  $\hat{a}_2$ , for selected samples.

This showcases that there can be significant and *sizeable* differences between the number of aborted female fetuses—the standard measure of sex-selective abortion—and the number of women who abort.

### 6. Conclusion

In this paper, we observe that sex-selective abortion introduces a correlation between birth interval length and the sex of the next-born child. Using a simple statistical model, we establish a link between the sign of this correlation and the presence of repeated sex-selective abortion between two consecutive births. Specifically, our model suggests that if the time to birth is shorter when the next-born is a girl rather than a boy, this provides evidence of repeated sex-selective abortion.

Using Indian data, we thus find some evidence of repeated sex-selective abortion at birth order 2 when the first child is a girl, and strong evidence of repeated sex-selective abortion at birth order 3 when the first two children are girls. However, this analysis does not address how many women sex-selectively abort or how likely they are to do so repeatedly.

To answer these questions, we make distributional assumptions that allow us to estimate our model using maximum likelihood. Our estimation results reveal significant heterogeneity across socio-demographic and geographic groups, birth orders, and sibling compositions.

For instance, some groups of women exhibit high abortion levels but do not abort a second time, such as literate and urban women at birth order 2 who first had a girl. In contrast, in other groups, the likelihood of repeated abortions is estimated to be as high as 13%, observed among women in the North at birth order 3 whose first two children are girls. In this group, the estimated number of aborted female fetuses is

approximately 50% larger than the number of women who abort. This underscores the empirical relevance of our proposed methodology for assessing the prevalence of sex-selective abortion.

Our proposed methodology should prove valuable for policymakers, particularly in identifying populations at high risk of multiple abortions. Since sex-selective abortions often occur under unsafe conditions, our methodology can aid in identifying women likely to experience health complications due to multiple abortions.

Our methodology only requires information on women’s birth histories (i.e., the sexes and dates of birth of their children), which can be an advantage or a disadvantage. This information is widely available in low and middle-income countries, where large household surveys are regularly conducted. Therefore, our methodology should prove useful for studying sex-selective abortion patterns in such countries. However, our methodology cannot be applied in countries where such information is not available. Additionally, our methodology measures (repeated) sex-selective abortion between two consecutive births and is thus specific to a particular birth order. Consequently, it does not directly provide information on (repeated) sex-selective abortion in the population as a whole. Another disadvantage of our methodology is that it requires large sample sizes, which may limit its suitability for studying sex-selective abortion patterns in countries with small populations.

Lastly, this paper makes several assumptions that could be restrictive. For example, we assume there is no correlation between the length of time women or households wait before trying to conceive and their willingness to abort if pregnant with a female fetus. Relaxing this assumption is left for further research.

**CRedit authorship contribution statement**

**Aditi Dimri:** Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Véronique Gille:** Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Philipp Ketz:** Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing.

**Data availability**

The Data are freely available, and we will share the code.

**Appendix A. Proof of Proposition 1**

**Proof of Proposition 1.** We first prove (i). Note that  $T_b = T_b^0$  in the absence of sex-selective abortion (i.e.,  $\alpha_1 = \alpha_2 = \dots = \alpha_{N_0} = 0$ ). Therefore, without sex-selective abortion,  $T_b$  is constant and, thus, independent of  $Y$ . In particular,  $E(T_b|Y = 0) = T_b = E(T_b|Y = 1)$ , and the results follows by contraposition.

Next, we prove (ii). In Section 3.1.1, we have shown that  $E(T_b|Y = 0) > E(T_b|Y = 1)$  if there is at most one sex-selective abortion (i.e.,  $\alpha_1 > 0$  and  $\alpha_2 = \alpha_3 = \dots = \alpha_{N_0} = 0$ ). Combining this with the proof of (i), we have  $E(T_b|Y = 0) \geq E(T_b|Y = 1)$  if no woman aborts a second time (i.e.,  $\alpha_2 = \alpha_3 = \dots = \alpha_{N_0} = 0$ ). Again, the desired results follows by contraposition.  $\square$

**Appendix B. Details for Section 3.1.2**

Using  $T_b = T_b^0 + AD$ , we have

$$E(T_b|Y = 0) = T_b^0 + (1 - \pi)\alpha_1 \frac{(1 - \alpha_2) + 2(1 - \pi)\alpha_2}{1 - \pi\alpha_1 - \pi(1 - \pi)\alpha_1\alpha_2} D.$$

and

$$E(T_b|Y = 1) = T_b^0 + (1 - \pi)\alpha_1 \frac{1 + 2(1 - \pi)\alpha_2}{1 + (1 - \pi)\alpha_1 + (1 - \pi)^2\alpha_1\alpha_2} D.$$

when  $N_0 = 2$ . Subtracting  $E(T_b|Y = 1)$  from  $E(T_b|Y = 0)$ , we obtain

$$E(T_b|Y = 0) - E(T_b|Y = 1) = \frac{(1 - \pi)\alpha_1 D}{(1 + (1 - \pi)\alpha_1 + (1 - \pi)^2\alpha_1\alpha_2)(1 - \pi\alpha_1 - \pi(1 - \pi)\alpha_1\alpha_2)} \times ((1 - \pi)^2\alpha_1\alpha_2^2 + (2(1 - \pi)\alpha_1 - 1)\alpha_2 + \alpha_1).$$

As the first factor is strictly positive, the sign of  $E(T_b|Y = 0) - E(T_b|Y = 1)$  only depends on the second factor, which is a polynomial of degree 2 with respect to  $\alpha_2$ .

For  $\alpha_1 > \frac{1}{4(1-\pi)}$  no real root exists and the sign of  $E(T_b|Y = 0) - E(T_b|Y = 1)$  is strictly positive for any value of  $\alpha_2$ . For  $\alpha_1 \leq \frac{1}{4(1-\pi)}$  the two real roots are

$$\frac{1 - 2(1 - \pi)\alpha_1 \pm \sqrt{1 - 4(1 - \pi)\alpha_1}}{2\alpha_1(1 - \pi)^2}.$$

However, for  $\alpha_1 \leq \frac{1}{4(1-\pi)}$ , we have that

$$\frac{1 - 2(1 - \pi)\alpha_1 + \sqrt{1 - 4(1 - \pi)\alpha_1}}{2\alpha_1(1 - \pi)^2} \geq \frac{1 - 2(1 - \pi)\alpha_1}{2\alpha_1(1 - \pi)^2} \geq \frac{1}{1 - \pi} > 1$$

for any  $\pi \in (0, 1)$  such that we only need to consider the first root. Denoting the first root by  $\alpha_2^{\text{sign}}(\alpha_1)$  the conclusion in the main text concerning the sign of  $E(T_b|Y = 0) - E(T_b|Y = 1)$  follows.

**Appendix C. Estimation details**

*C.1. Definition of  $f_{T_b^a}(\cdot)$*

First, we define  $f_{\Gamma}(\cdot)$  for sake of completeness. In particular,

$$f_{\Gamma}(t; \gamma, \beta) = \begin{cases} \frac{t^{\gamma-1} e^{-t/\beta}}{\beta^{\gamma} \Gamma(\gamma)} & \text{if } t > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where  $\Gamma(\cdot)$  denotes the gamma function. Next, we define  $f_{\Sigma\Gamma}(\cdot)$ . To that end, assume without loss of generality that  $\beta_1 < \beta_2$ . Then,

$$f_{\Sigma\Gamma}(t; \gamma_1, \beta_1, \gamma_2, \beta_2) = \begin{cases} C \sum_{k=0}^{\infty} \delta_k \frac{t^{\gamma_1+\gamma_2+k-1} e^{-t/\beta_1}}{\beta_1^{\gamma_1+\gamma_2+k} \Gamma(\gamma_1+\gamma_2+k)} & \text{if } t > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where  $C = \left(\frac{\beta_1}{\beta_2}\right)^{\gamma_2}$  and where  $\delta_k$  is defined recursively by

$$\delta_{k+1} = \frac{\gamma_2}{k+1} \sum_{i=1}^{k+1} \left(1 - \frac{\beta_1}{\beta_2}\right)^i \delta_{k+1-i}$$

for  $k = 0, 1, 2, \dots$  with  $\delta_0 = 1$ ; see Moschopoulos (1985) for more details. To numerically evaluate the infinite series in  $f_{\Sigma\Gamma}(\cdot)$ , we rely on a finite series approximation. In particular, we truncate the series at 30 terms. The resulting approximation error is immaterial.

**Table D.1**

Proportions of male births and differences in average spacing for remaining sibling compositions.

	Pooled	By education		By urban-rural status		By region	
		Illiterate	Literate	Rural	Urban	South	North
<b>Birth order 2 - B</b>							
[1985–1994] PMB	0.520 (0.002)***	0.525 (0.002)***	0.515 (0.003)	0.522 (0.002)***	0.517 (0.003)	0.519 (0.003)**	0.522 (0.002)***
DAS	0.37 (0.13)***	0.58 (0.17)***	0.08 (0.20)	0.39 (0.15)***	0.30 (0.25)	0.42 (0.20)**	0.33 (0.16)**
<i>n</i>	[82,748]	[43,361]	[39,354]	[56,480]	[26,268]	[35,163]	[47,585]
[1995–2004] PMB	0.508 (0.002)**	0.510 (0.003)	0.507 (0.003)**	0.510 (0.002)	0.504 (0.003)**	0.507 (0.003)*	0.509 (0.002)
DAS	0.19 (0.14)	0.25 (0.20)	0.12 (0.20)	0.31 (0.16)*	-0.12 (0.28)	0.23 (0.23)	0.16 (0.18)
<i>n</i>	[68,287]	[30,125]	[38,053]	[47,656]	[20,631]	[27,809]	[40,478]
[2005–2014] PMB	0.501 (0.001)***	0.509 (0.003)*	0.497 (0.002)***	0.502 (0.002)***	0.500 (0.003)***	0.498 (0.002)***	0.503 (0.002)***
DAS	0.20 (0.12)*	0.46 (0.17)***	-0.02 (0.15)	0.24 (0.13)*	0.07 (0.26)	-0.13 (0.19)	0.39 (0.14)***
<i>n</i>	[111,423]	[39,591]	[71,636]	[84,601]	[26,822]	[44,676]	[66,747]

(continued on next page)



Table D.1 (continued).

	Pooled	By education		By urban-rural status		By region	
		Illiterate	Literate	Rural	Urban	South	North
<b>Birth order 3 - GB</b>							
[1985–1994] PMB	0.527 (0.003)***	0.532 (0.004)***	0.520 (0.005)	0.528 (0.003)***	0.527 (0.006)**	0.519 (0.005)	0.532 (0.004)***
DAS	0.52 (0.20)**	0.62 (0.24)**	0.28 (0.36)	0.76 (0.23)***	-0.10 (0.42)	0.51 (0.35)	0.50 (0.25)**
<i>n</i>	[29,363]	[18,409]	[10,940]	[21,349]	[8,014]	[10,962]	[18,401]
[1995–2004] PMB	0.526 (0.003)***	0.519 (0.004)	0.535 (0.005)***	0.527 (0.004)***	0.522 (0.007)	0.526 (0.006)**	0.525 (0.004)***
DAS	0.41 (0.24)*	0.45 (0.29)	0.47 (0.40)	0.38 (0.27)	0.46 (0.51)	0.39 (0.43)	0.42 (0.28)
<i>n</i>	[21,593]	[12,771]	[8,779]	[16,336]	[5,257]	[7,248]	[14,345]
[2005–2014] PMB	0.526 (0.003)***	0.528 (0.004)***	0.524 (0.004)**	0.526 (0.003)***	0.528 (0.007)**	0.518 (0.005)	0.530 (0.004)***
DAS	0.41 (0.22)*	0.56 (0.29)**	0.18 (0.34)	0.46 (0.24)*	0.16 (0.58)	0.39 (0.41)	0.37 (0.26)
<i>n</i>	[28,402]	[15,228]	[13,118]	[23,212]	[5,190]	[9,267]	[19,135]
<b>Birth order 3 - BG</b>							
[1985–1994] PMB	0.516 (0.003)	0.512 (0.004)	0.523 (0.005)**	0.512 (0.003)	0.527 (0.006)**	0.505 (0.005)*	0.523 (0.004)***
DAS	-0.51 (0.20)**	-0.40 (0.24)*	-0.65 (0.36)*	-0.62 (0.23)***	-0.16 (0.42)	-0.41 (0.35)	-0.64 (0.25)**
<i>n</i>	[28,882]	[18,046]	[10,826]	[20,877]	[8,005]	[10,959]	[17,923]
[1995–2004] PMB	0.521 (0.003)**	0.519 (0.004)	0.524 (0.005)**	0.520 (0.004)*	0.522 (0.007)	0.515 (0.006)	0.524 (0.004)**
DAS	-0.68 (0.24)***	-0.86 (0.29)***	-0.43 (0.40)	-0.70 (0.27)***	-0.62 (0.52)	-0.64 (0.43)	-0.73 (0.28)***
<i>n</i>	[21,540]	[12,886]	[8,613]	[16,250]	[5,290]	[7,399]	[14,141]
[2005–2014] PMB	0.515 (0.003)	0.510 (0.004)	0.521 (0.004)*	0.514 (0.003)	0.523 (0.007)	0.502 (0.005)**	0.522 (0.004)**
DAS	-0.09 (0.22)	-0.36 (0.28)	0.34 (0.35)	0.04 (0.24)	-0.59 (0.57)	0.41 (0.41)	-0.45 (0.26)*
<i>n</i>	[27,717]	[14,979]	[12,691]	[22,632]	[5,085]	[9,152]	[18,565]
<b>Birth order 3 - BB</b>							
[1985–1994] PMB	0.505 (0.003)**	0.511 (0.004)	0.495 (0.005)***	0.506 (0.003)**	0.505 (0.006)	0.495 (0.005)***	0.512 (0.004)
DAS	0.30 (0.20)	-0.03 (0.24)	0.82 (0.38)**	0.28 (0.23)	0.33 (0.42)	0.43 (0.34)	0.18 (0.26)
<i>n</i>	[28,949]	[18,832]	[10,106]	[20,952]	[7,997]	[11,415]	[17,534]
[1995–2004] PMB	0.500 (0.004)***	0.503 (0.004)**	0.496 (0.006)***	0.505 (0.004)**	0.487 (0.007)***	0.495 (0.006)***	0.503 (0.004)**
DAS	0.30 (0.25)	0.31 (0.30)	0.23 (0.42)	0.26 (0.28)	0.33 (0.51)	0.19 (0.43)	0.33 (0.30)
<i>n</i>	[20,228]	[12,356]	[7,838]	[15,124]	[5,104]	[7,356]	[12,872]
[2005–2014] PMB	0.491 (0.003)***	0.501 (0.004)***	0.478 (0.005)***	0.491 (0.004)***	0.488 (0.007)***	0.482 (0.005)***	0.495 (0.004)***
DAS	0.40 (0.24)*	0.23 (0.30)	0.41 (0.38)	0.13 (0.26)	1.52 (0.59)**	0.36 (0.43)	0.35 (0.29)
<i>n</i>	[25,105]	[13,845]	[11,213]	[20,292]	[4,813]	[8,947]	[16,158]

The table shows the proportions of male births (PMB) and the differences in average spacing between boys and girls (DAS) for different samples: B denotes “first child is a boy” and GB, BG, and BB denote “first two children are a girl and a boy, a boy and a girl, and two boys”, respectively. Standard errors are reported in parentheses. *p*-values are for two-sided t-tests for testing that the probability of male birth equals  $\pi$  and that the difference in mean spacing equals 0.

\*  $p < 0.1$ .  
 \*\*  $p < 0.05$ .  
 \*\*\*  $p < 0.01$ .

Lastly, we note that  $f_{T_b}(\cdot)$  uses the fact that the sum of two independent gamma-distributed random variables with the same scale parameter, say  $\beta$ , but (possibly) different shape parameters, say  $\gamma_1$  and  $\gamma_2$ , is gamma-distributed with shape parameter  $\gamma_1 + \gamma_2$  and scale parameter  $\beta$ .

C.2. Calibration of  $\gamma_c$  and  $\beta_c$

The calibration of  $\gamma_c$  and  $\beta_c$  is performed by means of maximum likelihood, fitting a gamma distribution to the time between marriage and first birth (minus 9); it is performed for each of our three time

periods separately.<sup>21</sup> The underlying assumptions are that households try to conceive immediately after marriage, such that the time between marriage and first birth (minus 9) is a good proxy for the time it takes to conceive, and that there is no sex-selective abortion before the birth of the first child. The second assumption is largely corroborated by the numbers in Table 4 and widely accepted in the literature (see, e.g., Dahl and Moretti, 2008; Milazzo, 2018; Heath and Tan, 2018).

<sup>21</sup> Here, we only use DHS 4 because it is the only round that differentiates between date of marriage and date of marriage contract.

**Table D.2**  
*t*-statistics for testing  $H_0 : \alpha_1 = 0$  vs.  $H_1 : \alpha_1 > 0$ .

Birth order	Period	Illiterate	Literate	Rural	Urban	South	North
2	1985–1994	1.55	6.15	3.26	4.92	1.74	5.78
	1995–2004	0.64	9.09	3.57	6.82	2.91	6.07
	2005–2014	1.94	12.57	6.60	9.54	5.48	9.37
3	1985–1994	1.16	6.27	3.96	4.30	2.44	4.56
	1995–2004	2.58	7.77	5.02	6.88	3.68	6.24
	2005–2014	2.86	13.67	8.70	8.26	6.95	8.50

The table shows the *t*-statistics for testing  $H_0 : \alpha_1 = 0$  vs.  $H_1 : \alpha_1 > 0$  for samples with “only girls” sibling compositions.

**Table D.3**  
 Shares of women at risk of multiple abortions, shares of women who abort, and shares of women who abort a second time for selected samples with standard errors.

Group	Birth order	Period	Share of women at risk of multiple abortion	Share of women who abort	Share of women who abort a second time
Literate	2	1995–2004	0.000 (0.021)	0.124 (0.014)	0.003 (0.166)
		2005–2014	0.014 (0.015)	0.122 (0.010)	0.113 (0.131)
	3	1995–2004	0.101 (0.038)	0.189 (0.024)	0.536 (0.253)
		2005–2014	0.109 (0.029)	0.237 (0.017)	0.458 (0.147)
Urban	2	1995–2004	0.000 (0.028)	0.128 (0.019)	0.001 (0.217)
		2005–2014	0.000 (0.026)	0.157 (0.016)	0.000 (0.168)
	3	1995–2004	0.087 (0.053)	0.221 (0.032)	0.394 (0.281)
		2005–2014	0.084 (0.052)	0.246 (0.030)	0.339 (0.242)
North	2	1995–2004	0.030 (0.017)	0.078 (0.013)	0.380 (0.267)
		2005–2014	0.043 (0.014)	0.091 (0.010)	0.468 (0.185)
	3	1995–2004	0.064 (0.028)	0.127 (0.020)	0.503 (0.284)
		2005–2014	0.129 (0.024)	0.133 (0.016)	0.977 (0.266)

The table shows the estimated shares of aborted female fetuses, the estimated shares of women who abort,  $\hat{\alpha}_1$ , and the estimated shares of women who abort a second time,  $\hat{\alpha}_2$ , for selected samples (with “only girls” sibling compositions). Standard errors are reported in parentheses.

**Appendix D. Additional tables**

See Tables D.1–D.3.

**References**

Alesina, Alberto, Giuliano, Paola, Nunn, Nathan, 2018. Traditional agricultural practices and the sex ratio today. *PLoS One* 13 (1), e0190510.

Andrews, Donald W.K., Cheng, Xu, 2012. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* 80 (5), 2153–2211.

Andrews, Donald W.K., Guggenberger, Patrik, 2010. Asymptotic size and a problem with subsampling and with the *m* out of *n* bootstrap. *Econometric Theory* 26 (2), 426–468.

Anukriti, Sharma, Bhalotra, Sonia, Tam, Eddy H.F., 2022. On the quantity and quality of girls: Fertility, parental investments and mortality. *Econ. J.* 132 (641), 1–36.

Bhalotra, Sonia R., Cochrane, Tom, 2010. Where have all the young girls gone? Identification of sex selection in India. IZA Discussion Paper.

Bongaarts, John, Guilмото, Christophe Z., 2015. How many more missing women? Excess female mortality and prenatal sex selection, 1970–2050. *Popul. Devel. Rev.* 41 (2), 241–269.

Boyle, Elizabeth Heger, King, Miriam, Sobek, Matthew, 2022. IPUMS-demographic and health surveys: Version 9 [dataset]. IPUMS and ICF.

Cassan, Guilhem, Baland, Jean-Marie, Woitrin, Francois, 2023. Sex-Selective Abortions and Instrumental Births as the two faces of the Stopping Rule. *New Measures and World Evidence*. CEPR Discussion Papers 18014.

Chao, Fengqing, Gerland, Patrick, Cook, Alex R., Alkema, Leontine, 2019. Systematic assessment of the sex ratio at birth for all countries and estimation of national imbalances and regional reference levels. *Proc. Natl. Acad. Sci.* 116 (19), 9303–9311.

Dahl, Gordon B., Moretti, Enrico, 2008. The demand for sons. *Rev. Econ. Stud.* 75 (4), 1085–1120.

Das Gupta, Monica, Ebenstein, Avraham Y., Sharygin, Ethan Jennings, 2010. China’s marriage market and upcoming challenges for elderly men. In: *World Bank Policy Research Working Paper Series*, Vol. 1.

Diamond-Smith, Nadia, Bishai, David, 2015. Evidence of self-correction of child sex ratios in India: A district-level analysis of child sex ratios from 1981 to 2011. *Demography* 52 (2), 641–666.

Dubuc, Sylvie, Coleman, David, 2007. An increase in the sex ratio of births to India-born mothers in England and Wales: Evidence for sex-selective abortion. *Popul. Dev. Rev.* 33 (2), 383–400.

Elliott, Graham, Müller, Ulrich K., Watson, Mark W., 2015. Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica* 83, 771–811.

Goli, Srinivas, Mavisakalyan, Astghik, Rammohan, Anu, Vu, Loan, 2022. Conflicts and son preference: Micro-level evidence from 58 countries. *Econ. Hum. Biol.* 46, 101146.

Grimes, David A., Benson, Janie, Singh, Susheela, Romero, Mariana, Ganatra, Bela, Okonofua, Friday E., Shah, Iqbal H., 2006. Unsafe abortion: The preventable pandemic. *Lancet (London, England)* 368 (9550), 1908–1919.

Guilмото, Christophe Z., 2012. Sex Imbalances at Birth: Current Trends, Consequences and Policy Implications. UNFPA Asia and Pacific Regional Office.

Guilмото, Christophe Z., Chao, Fengqing, Kulkarni, Purushottam M., 2020. On the estimation of female births missing due to prenatal sex selection. *Popul. Stud.* 74 (2), 283–289, PMID: 32489140.

Heath, Rachel, Tan, Xu, 2018. Worth fighting for: Daughters improve their mothers’ autonomy in South Asia. *J. Dev. Econ.* 135, 255–271.

Hwang, Jisoo, Lee, Chulhee, Lee, Esther, 2019. Gender norms and housework time allocation among dual-earner couples. *Labour Econ.* 57, 102–116.

Jacobsen, R., Moller, H., Mouritsen, A., 1999. Natural variation in the human sex ratio. *Human Reproduction* 14 (12), 3120–3125.

Jayachandran, Seema, 2017. Fertility decline and missing women. *Am. Econ. J.: Appl. Econ.* 9 (1), 118–139.

- Jayachandran, Seema, Kuziemko, Ilyana, 2011. Why do mothers breastfeed girls less than boys? Evidence and implications for child health in India. *Q. J. Econ.* 126 (3), 1485–1538.
- Jha, Prabhat, Kesler, Maya A., Kumar, Rajesh, Ram, Faujdar, Ram, Usha, Aleksandrowicz, Lukasz, Bassani, Diego G., Chandra, Shailaja, Banthia, Jayant K., 2011. Trends in selective abortions of girls in India: Analysis of nationally representative birth histories from 1990 to 2005 and census data from 1991 to 2011. *Lancet* 377 (9781), 1921–1928.
- Ketz, Philipp, 2019. On asymptotic size distortions in the random coefficients logit model. *J. Econometrics* 212 (2), 413–432.
- Ketz, Philipp, 2022. Allowing for weak identification when testing GARCH-X type models. *arXiv:2210.11398*.
- Klasen, Stephan, Wink, Claudia, 2003. “Missing women”: Revisiting the debate. *Feminist Econ.* 9 (2–3), 263–299.
- Mathews, T.J., Hamilton, Brady E., 2005. Trend analysis of the sex ratio at birth in the United States. *Natl. Vital Stat. Rep.* 53 (20), 1–17.
- Matsuo, Koji, Ushioda, Norichika, Udoff, Laurence C., 2009. Parental aging synergistically decreases offspring sex ratio. *J. Obstetrics Gynaecol. Res.* 35 (1), 164–168.
- Mavisakalyan, Astghik, Minasyan, Anna, 2023. The role of conflict in sex discrimination: The case of missing girls. *Econ. Dev. Cult. Change* 71 (2), 443–484.
- Meh, C., Sharma, A., Ram, U., Fadel, S., Correa, N., Snelgrove, J.W., Shah, P., Begum, R., Shah, M., Hana, T., 2022. Trends in maternal mortality in India over two decades in nationally representative surveys. *BJOG: Int. J. Obstetrics Gynaecol.* 129 (4), 550–561.
- Milazzo, Annamaria, 2018. Why are adult women missing? Son preference and maternal survival in India. *J. Dev. Econ.* 134, 467–484.
- Moschopoulos, Peter G., 1985. The distribution of the sum of independent gamma random variables. *Ann. Inst. Statist. Math.* 37 (3), 541–544.
- Nicolich, Mark J., Huebner, Wendy W., Schnatter, A. Robert, 2000. Influence of parental and biological factors on the male birth fraction in the United States: An analysis of birth certificate data from 1964 through 1988. *Fertil. Steril.* 73 (3), 487–492.
- Orzack, Steven H., Stubblefield, J. William, Akmaev, Viatcheslav R., Colls, Pere, Munné, Santiago, Scholl, Thomas, Steinsaltz, David, Zuckerman, James E., 2015. The human sex ratio from conception to birth. *Proc. Natl. Acad. Sci.* 112 (16), E2102–E2111.
- Oster, Emily, 2022. How long does it take to get pregnant?. Retrieved from “<https://www.parentdata.org/p/how-long-does-it-take-to-get-pregnant>” on December 13, 2022.
- Pörtner, Claus C., 2022. Birth spacing and fertility in the presence of son preference and sex-selective abortions: India’s experience over four decades. *Demography* 59 (1), 61–88.
- Qian, Nancy, 2008. Missing women and the price of tea in China: The effect of sex-specific earnings on sex imbalance. *Q. J. Econ.* 123 (3), 1251–1285.
- Rastogi, Garima, Sharma, Anisha, 2022. Unwanted daughters: The unintended consequences of a ban on sex-selective abortions on the educational attainment of women. *J. Popul. Econ.* 1–44.
- Saikia, Nandita, Meh, Catherine, Ram, Usha, Bora, Jayanta Kumar, Mishra, Bhaskar, Chandra, Shailaja, Jha, Prabhat, 2021. Trends in missing females at birth in India from 1981 to 2016: Analyses of 2.1 million birth histories in nationally representative surveys. *Lancet Glob. Health* 9 (6), e813–e821.
- Schoumaker, Bruno, 2014. Quality and Consistency of DHS Fertility Estimates, 1990 to 2012. ICF International Rockville, Maryland.
- Sen, Amartya, 1990. More than 100 million women are missing. In: *The New York Review of Books*.
- Singh, Susheela, Remez, Lisa, Sedgh, Gilda, Kwok, Lorraine, Onda, Tsuyoshi, 2018a. Abortion worldwide 2017: Uneven progress and unequal access.
- Singh, Susheela, Shekhar, Chander, Acharya, Rajib, Moore, Ann M., Stillman, Melissa, Pradhan, Manas R., Frost, Jennifer J., Sahoo, Harihar, Alagarajan, Manoj, Husain, Rubina, 2018b. The incidence of abortion and unintended pregnancy in India, 2015. *Lancet Global Health* 6 (1), e111–e120.
- World Health Organization, 2022. Abortion Care Guideline. World Health Organization.
- Yokoe, Ryo, Rowe, Rachel, Choudhury, Saswati Sanyal, Rani, Anjali, Zahir, Farzana, Nair, Manisha, 2019. Unsafe abortion and abortion-related death among 1.8 million women in India. *BMJ Global Health* 4 (3), e001491.