Sex-Selective Abortions and Birth Spacing [*]

Claus C Pörtner

Department of Economics

Albers School of Business and Economics

Seattle University, P.O. Box 222000

Seattle, WA 98122

cportner@seattleu.edu

www.clausportner.com

&

Center for Studies in Demography and Ecology

University of Washington

April 2017

Abstract

This paper addresses the question: how does birth spacing interact with the use of sex-selective abortions? I introduce a statistical method that incorporates how sex-selective abortions affect both the likelihood of a son and spacing between births. Using India's National Family and Health Surveys, I show that use of sex selection leads to longer spacing after a daughter is born. Women with 8 or more years of education, both in urban and rural areas, are the main users of sex-selective abortions. Women with less education do not appear to use sex selection. Predicted lifetime fertility for high-education women declined more than 10% between 1985–1994, when sex selection was legal, and 1995–2006, when sex selection was illegal. Abortions per woman increased almost 20% for urban women and 50% for rural women between the two periods, suggesting that making sex selection illegal has not reversed its use. Finally, sex selection appears to be used to ensure one son rather than multiple sons. JEL: J1, O12, I1 Keywords: India, prenatal sex determination, censoring, competing risk

1   Introduction

The males-to-females ratio at birth increased dramatically over the last three decades in India as access to prenatal sex determination expanded.[1] Our understanding of the use of sex selection is, however, constrained by lack of information; there are no official data on sex selection, and the few surveys that ask about use of sex selection show signs of serious under-reporting (?). With no direct information, researchers have relied on a simple method for establishing factors that affect sex selection: use the sex of children born as the dependent variable, and estimate the effects of variables on the probability of having a son.[2] Based on this simple method, we know that families with no sons are more likely to use sex selection the higher the parity; that use of sex selection increases with socioeconomic status, especially education; and that sex selection is more widespread in cities than in rural areas (???).[3]

I address an important questions that the prior literature has been unable to examine because of the simple method's limitations: how does birth spacing interact with use of sex-selective abortions? I introduce a novel method that directly incorporates that sex-selective abortions affect both the likelihood of a son being born *and* the duration between births. I use the method to show how spacing between births play an important role in our understanding of sex selection decisions.

In the absence of sex-selective abortions, son preference often leads to a shorter duration until the next birth if the previous birth was a daughter (see, for example, ?????). The resulting shorter spacing is thought to be associated with worse health outcomes for the girls (????).

What has not previously been appreciated is that the introduction of sex-selective

---

[1] See ?, ?, ?, ? and ?. India is not alone; both China and South Korea saw significant changes in the sex ratio at birth (??).

[2] In the absence of any interventions, the probability of having a son is approximately 0.512, which is independent of genetic factors (??). With fetus sex random, a statistically significant variable indicates an association with use of sex-selective abortions. Examples of studies that have used this approach are ?, ?, and ?.

[3] There is substantial disagreement on whether sex-selective abortion is used for the first birth (??).

abortions substantially changed the relationship between son preference and birth spacing. The change happens because each abortion significantly increases the duration until the next birth. The increase in duration can be divided into three parts starting from the time of the abortion. First, the uterus needs at least two menstrual cycles to recover before conception, because a less than three months space between abortion and conception substantially increases the likelihood of a spontaneous abortion (**?**). Secondly, the average expected time to conception is about 6 months. Finally, sex determination tests are generally reliable only from 3 months of gestation onwards. The result is that each abortion delays the next birth by approximately a year.[4]

How spacing interacts with sex selection is important for two reasons. First, spacing, by itself, can no longer be used to predict son preference. Because each abortions extends spacing by about a year, we now have a situation where the *longest* spacing is likely observed for families with the strongest son preference. To further complicate matters, we may still observe short spacing after the birth of daughter as a representation of son preference for families who either have less access to sex-selective abortions or are willing to have more children. Spacing is still important and useful in understanding sex selection and son preference, but only if combined with the outcome of that spacing—in other words the likelihood of observing a boy or a girl.

Second, families may reverse their decision to use sex selection in-between births. A reversal could happen if they do not want the space between children to be "too long", or if there are concerns about possible infertility as a result of too many abortions in a

---

[4] There are three well-developed technologies for determining a fetus' sex: Chorionic villus sampling (CVS), amniocentesis, and ultrasound. CVS can be applied after the shortest period of gestation (8 to 12 weeks). It is the most complicated, but also the most reliable, and an abortion can be done in the first trimester. Amniocentesis can be performed after fourteen weeks, but requires three to four weeks before the result is available, so an abortion cannot be performed until more than halfway through the second trimester. Ultrasound has the advantages of being non-invasive and relatively cheap but is associated with higher risk of faulty sex determination if done early. Generally, a fetus' sex can be determined in the third month of gestation if it is a boy and the fourth month if it a girl.

The waiting time to conception does vary by woman, but even if it is very short, say one month, the minimum additional space between births would be 6 months per abortion. For a discussion of time to conception see (**?**)

row without a birth. Hence, for a given parity, the sample of women for whom we do observe a birth may behave differently from the sample of women who have not yet had a birth. This is a problem because births to women with shorter spacing are more likely to be in the survey for a given parity. Hence, if decisions on sex selection change with duration since last birth, the predicted sex ratio for a parity using the simple method will be a biased estimate of the final observed sex ratio for the parity when childbearing is complete.

[TK maybe not focus so much on the simple method but rather understanding son preference]. The main contribution of this paper is the introduction and application of a method that incorporates the decisions about birth spacing, and sex-selective abortions. The empirical model is a competing risk, non-proportional hazard model with two exit states: either a boy or a girl is born. This approach has TK major advantages over the simple method

Secondly, by explicitly incorporating censoring of birth spacing, it addresses any potential bias from ignoring how sex selection decisions may change with duration from the previous birth.

None of these is possible using the simple method.

I apply the method to birth histories from India's National Family and Health Surveys for Hindu women covering the period 1972 to 2006. There are three main findings.

There is, however, still no evidence of sex selection being used on the first birth.

Secondly, sex selection appears to be used for securing one son, rather than a large number of sons. There is only limited use of sex-selective abortions for better-educated women with one or more sons. The exception is rural women with one son and one daughter, presumably to compensate for the higher child mortality in rural areas. These results are in line with the differential stopping behavior observed in many studies before sex selection became available (**???**).

The method can better predict "final" sex ratios than the simple model

2 The Interaction between Son Preference and Birth Spacing

The extension of birth spacing that result from abortions as described above is only one part of the interaction between son preference and birth spacing.

- the decision on sex selection and spacing - the mechanical extension - the possibility of changing use of sex selection

- what other factors might change birth spacing? - increased use of contraceptives - changes in labor force participation

(**?**) for a recent example of spacing and son preference

In the absence of prenatal sex determination, parents' only choice if they wanted a son was to try for another pregnancy as soon as possible after the birth of a girl. This led to shorter spacing after the birth of a girl and what is known as differential stopping behavior, where parents were more likely to cease child bearing after the birth of a son than after a daughter [TK add references here]. As discussed above, the shorter spacing after the birth of a girl is associated with worse health outcomes for both mother and children. Because of the potential health effects, most of the literature on spacing for countries like India naturally focus on whether spacing was too short.

With the introduction of sex-selective abortions, parents can abort fetuses of unwanted sex. With an approximately 12 months additional duration per abortion, this strategy can result in very long durations between births, and "too long" spacing may become a greater concern for parents than "too short" spacing.[5] If the duration from last birth becomes long enough parents may reverse their decision to use prenatal sex determination and carry the next pregnancy to term whether it is a male or female fetus.

At least four different factors may make a reversal of the decision to use sex selection

---

[5] Shorter spacing was likely less of a concern to begin with for better educated mothers—who are also the most likely to use sex selection—since their children are substantially less likely to be negatively affected by short spacing than children of less educated mothers (**?**).

between one birth and the next more likely, everything else equal. First, the more concerned parents are about possible infertility from repeated abortions or the closer they are to the end of their reproductive period, they more likely they are to decide that any child is better than not having an additional child at all. Second, the higher the parents' discount factor is—that is, the more the value present outcomes relative to future outcomes—the more likely they are to reverse their decision to sex selection. Third, economies of scale in childrearing—both in terms of time cost and direct cost—are easier to take advantage of if births are closer (**?**, p 947). Hence, the stronger these economies of scale are, the more likely parents are to reverse their decision to use sex selection. Finally, if women's wages increase with age, parents prefer to have their children early and close together, while the opportunity cost of children is lower (**?**).[6] The more pronounced the increase in wage is with age, the more attractive it will be to reverse the decision to use sex selection. Although, the latter two are most associated with the literature on shorter spacing in middle- and high-income countries, the increasing education and wealth in India means that a growing portion of the population is likely to behave in a similar manner.[7]

If "too long" spacing leads parents to reverse their decision to use prenatal sex determination within a spell this has important implications for the empirical method.

[TK find examples of the son preference spacing papers and look at what they did for "theory"]

[TK this needs to move to later] Finally, increased reliability of access and effectiveness of contraceptive can lead to shorter spacing between births (**??**). With less reliable contraception parents choose a higher level of contraception—resulting in longer spacing—to avoid having too many children by accident. But, as contraception becomes more effective parents can more easily avoid future births, allowing them to reduce the spacing

---

[6] The time needed to care for young children can also impact timing if skills depreciate when out of the labor market (**?**). Whether depreciation leads to longer or shorter spacing depends on the specification of the depreciation function. See also the discussion in **?**, p 315 and references therein.

[7] The possible exception is for the first birth, where the pressure to show fecundity—which shortens the duration to the first birth—is potentially an important factor in India (**???**).

5

between births without having to worry about overshooting their preferred number of children. This idea may also help explain shorter spacing for better educated women than for less educated women, provided that knowledge and ability to use contraception differ across education groups (**??**).

Lowering the cost of prenatal sex determination does, however, bias the overall sex ratio further towards boys, and the stronger the son preference and the more patient the parents, the larger the change in sex ratio. Correspondingly, both the number of abortions per woman and the average number of spells per woman where a prenatal scan is used also increases. Since fertility changes little, the increased use of abortion is associated with longer spacing between births. This increase in spacing is, however, not uniform across parities. In addition to increasing the overall sex ratio, lower cost of prenatal scan leads to use of sex selection at earlier parities. The result is that in some cases there is either no change or a *decrease* in average spacing for a given parity—although with a more than compensating increase in an earlier parity. Hence, it is possible for individual parities to show a *decline* in sex selection at the same time as the overall use goes up. This is an important caution against over-interpreting individual parity results. It also highlights the need to incorporate fertility progression into estimates, something the simple model is incapable of.

Increasing the value of having sons, α, when there is access to sex selection results in slightly lower fertility in most cases, but the effect is neither strong nor uniform. What is clear is that the sex ratio becomes more male biased the stronger the son preference. Correspondingly, there is an increased number of abortions, more spells where a prenatal scan is used, and the spacing between births is longer.

Son preference, however, manifests in two ways in this model. First, if α is larger than 0.5 parents gain more utility from having an additional son than from an additional daughter—provided the had the same number of sons and daughters before. Second, parents will never abort a male fetus, only a female one. This explains why even when

parents have equal utility from an additional son or daughter sex ratios still increases as cost of children increases and cost of prenatal scans decline.

At $\alpha = 0.5$ parents prefer an equal number of boys and girls. The more children you have the less likely it is that you will resort to sex selection to ensure a balance between sons and daughters. First, the chances of having, for example, only boys or only girls, decreases with each additional child.[8] Second, the more children you have the lower the marginal utility of having another, which means that, for a given price of prenatal scan, the less attractive it is to bear the cost of using prenatal scans to ensure a specific sex. As fertility fall to, say, two, however, those parents who had a girl as their first child will now want to use prenatal sex determination to ensure a balanced number of sons and daughters and this is why the sex ratio is so unbalanced as the cost of children increases. Parents who had a son as their first child would also want to use sex selection to ensure both a son *and* a daughter, but since that option is not available here they have their second birth without scan.

Stretching the interpretation of the model, think of $\alpha = 0.5$ as capturing parents treating sons and daughters equally once born. The use of sex selection to achieve balance—but only when the fetus is female—could be a potential explanation for the suggestion that sons and daughters are treated relatively equally and we are still seeing an unequal sex ratio. What is relevant here it that there is no way to identify which of these two different types of son preference affect sex ratios from fertility data. This is a concern because the way we currently measure son preference allows us to capture only the distribution preference, and the different types of son preference likely have different policy implications.

An important question here is whether parents ever change their use of prenatal sex determination between two consecutive births. I capture this by the percent of spells where prenatal scans were initially used for one or more period, but the birth is the re-

---

[8] For 2 children the likelihood of only boys or only girls is 50%, but for 4 children it is 12.5% (assuming equal probability of a boy or a girl).

sult of a pregnancy without a prenatal scan. Whether parents will consider the duration from last birth too long depends mainly on the discount rate. On average, the more impatient parents are—the less they like long spells between births—the more likely they are to change their decision to use sex selection.[9] Lowering the cost of prenatal sex determination generally leads to a smaller percentage of parents who change their use of sex selection. But, the total number of spells with a change may still go up because lower cost of prenatal sex determination also increases the use of sex selection.

Another reason for parents to change their decision to use sex selection is that they get "too close" to the end of the time horizon after extensive use of abortions and prefer any child to the potential of having no child at all. This type of reversal explains why the effects of changing the discount rate and cost of prenatal sex determination are not uniform. Related, extending the time horizon leads to a smaller likelihood of parents reversing their use of sex selection between births when there is strong incentive to use sex selection.[10]

There are three main conclusions from the model.[11] The increased use of sex selection shows up both in the proportion of boys among births and longer spacing between births—and the higher the parity the longer the duration between births. The other factors, such as the price of sex selection, discount factor, and son preference also affect the use of sex selection, but to a much smaller degree. Second, parents do reverse the use of sex selection, both when they use sex selection relatively infrequently and when they use it heavily. Reversal are more likely the more impatient parents are or the shorter time they have left in their reproductive horizon, which means that the longer the duration has been since the last birth the more likely they are to reverse their decision to use sex

---

[9] Using only those situations where there are any abortions, regressing the discount rate against the percent spells where prenatal scan is initially used but not for the birth show a strong positive effect—higher number, more patient—holding all the other parameters constant.

[10] These results are available in the online Appendix.

[11] A caveat is that these results reflect the particular utility function chosen. If parents have a different utility function, like a preference for one son but otherwise no preference across boys and girls the results may look different.

selection. Finally, the model shows that son preference in the number of children is not required for an unequal sex ratio. As long as parents are only willing to abort female fetuses, declining fertility will lead to increased use of sex selection to ensure a balanced number of sons and daughters.

The model highlights two major limitations of the simple empirical method, which a better empirical approach needs to allow for. First, increased use of sex selection leads to longer spacing which, everything else equal, reduces the number of births available for estimation in the simple model and thereby lowers its precision because fewer parents will make it to a given parity by the time of the survey. Furthermore, the more intensively sex selection is used the larger the decline in sample size. Second, by using only observed births the simple method cannot capture when parents change their decision on the use of sex selection; with a reversal the spell can still result in the birth of a girl. This last point is especially important when trying to establish whether son preference has changed over time in a situation where the cost of prenatal sex determination has likely *increased* over time because of the PNDT act. As I show an increase the cost of prenatal sex determination does decrease the sex ratio and lower the number of abortions, but it also *increases* the likelihood that parents change their initial decision to use sex selection within a spell.

3 Estimation Strategy

The theoretical model has three implications for fertility decisions with access to sex-selective abortions: a higher probability that next child is a son, longer average time to next birth, and that the use of prenatal sex determination may change between one observed birth and the next. All three must be incorporated into the estimation strategy to provide unbiased and precise estimates of the use of sex-selective abortions.

The empirical model is a discrete time non-proportional competing risk hazard model

with two exit states: either a boy or a girl is born.[12] A woman's reproductive life is divided into spells which each covers the period between births (from marriage to first birth for the first spell). For each woman, $i = 1, \ldots, n$, the starting point for a spell is time $t = 1$ and the spell continues until time t_i when either a birth occurs or the spell is censored. The time of censoring is assumed independent of the hazard rate, as is standard in the literature.

There are two exit states: birth of a boy, $j = 1$, or birth of a girl, $j = 2$, and J_i is a random variable indicating which event took place. The discrete time hazard rate h_{ijt} is

$$h_{ijt} = \Pr(T_i = t, J_i = j \mid T_i \geq t; \mathbf{Z}_{it}, \mathbf{X}_i), \tag{1}$$

where T_i is a discrete random variable that captures when woman i's birth occurs. To ease presentation the indicator for spell number is suppressed. The vectors of explanatory variable $\mathbf{Z}_{it}$ and $\mathbf{X}_i$ include information about various individual, household, and community characteristics discussed below.

The hazard rate is specified as

$$h_{ijt} = \frac{\exp(D_j(t) + \alpha'_{jt}\mathbf{Z}_{it} + \beta'_j\mathbf{X}_i)}{1 + \sum_{l=1}^{2} \exp(D_j(t) + \alpha'_{lt}\mathbf{Z}_{it} + \beta'_l\mathbf{X}_i)} \quad j = 1, 2 \tag{2}$$

where $D_j(t)$ is the piece-wise linear baseline hazard for outcome j, captured by dummies and the associated coefficients,

$$D_j(t) = \gamma_{j1}D_1 + \gamma_{j2}D_2 + \ldots + \gamma_{jT}D_T, \tag{3}$$

where $D_m = 1$ if $t = m$ and zero otherwise. This approach to modeling the baseline hazard is flexible and does not place overly strong restrictions on the baseline hazard.

Specifying the model as a proportional hazard model, i.e. one where explanatory vari-

---

[12] **?** used a discrete hazard model to examine whether there were under-reporting of births in rural China, although they estimated separate waiting time regressions for boys and girls.

ables simply shift the hazard rates up or down independent of spell length, results in greater efficiency provided the proportionality assumption holds. The problem is that the proportional hazard model does not allow covariates to have different effects over time within a spell and ignoring differences in the shape of the hazard functions between different types of individuals can lead to substantial bias. It is highly unlikely, even in the absence of prenatal sex determination, that the baseline hazards are the same across education levels, areas of residence, or sex composition of previous births. The bias from the proportionality assumption is likely exacerbated by the introduction of prenatal sex determination for two reasons. First, one of the main points of this paper is that use of sex-selective abortions affects birth spacing, and use of sex selection differs across groups. Second, as discussed above, the use of sex selection may vary within a spell depending on the length of the spell.

I therefore use a non-proportional model where the main explanatory variables and the interactions between them are interacted with the baseline hazards. This is captured by the Z set of explanatory variables

$$\mathbf{Z}_{it} = D_j(t) \times (\mathbf{Z}_1 + Z_2 + \mathbf{Z}_1 \times Z_2), \tag{4}$$

where $D_j(t)$ is the piece-wise linear baseline hazard and $\mathbf{Z}_1$ captures sex composition of previous children, if any, and Z_2 captures area of residence. This allows the effects of the main explanatory variables on the probabilities of having a boy, a girl, or no birth to vary over time within a spell. The use of a non-proportional specification, together with a flexible baseline hazard, also mitigates any potential effects of unobserved heterogeneity (**?**).

The remaining explanatory variables, **X**, enter proportionally. To further minimize any potential bias from assuming proportionality, estimations are done separately for differ-ent levels of mothers' education and for different time periods. The exact specifications

and the individual variables are described below.

Equation (**??**) is equivalent to the logistic hazard model and has the same likelihood function as the multinomial logit model (**??**). Hence, if the data are transformed so the unit of analysis is spell unit rather than the individual woman, the model can be estimated using a standard multinomial logit model.[13] In the reorganized data the outcome variable is zero if the woman does not have a child in a given period, one if she gives birth to a son in that period, and two if she gives birth to a daughter in that period.

Direct interpretation of the estimated coefficients for this model is challenging because of the competing risk setup. First, coefficients show the change in hazards relative to the base outcome, no birth, rather than simply the hazard of an event. Second, a positive coefficient does not necessarily imply that an increase in a variable's value increases the probability of the associated event because the probability of another event may increase even more (**?**).

The model does, however, make it straightforward to calculate the predicted probabilities of having a boy and of having a girl for each t, conditional on a set of explanatory variables and not having had a child before that period. The predicted probability of having a boy in period t for a given set of explanatory variable values, $\mathbf{Z}_k$ and $\mathbf{X}_k$, is

$$P(b_t|\mathbf{X}_k, \mathbf{Z}_{kt}, t) = \frac{\exp(D_j(t) + \alpha'_{1t}\mathbf{Z}_{kt} + \beta'_1\mathbf{X}_k)}{1 + \sum_{l=1}^{2} \exp(D_j(t) + \alpha'_{lt}\mathbf{Z}_{kt} + \beta'_l\mathbf{X}_k)}, \qquad (5)$$

and the predicted probability of having a girl is

$$P(g_t|\mathbf{X}_k, \mathbf{Z}_{kt}, t) = \frac{\exp(D_j(t) + \alpha'_{2t}\mathbf{Z}_{kt} + \beta'_2\mathbf{X}_k)}{1 + \sum_{l=2}^{2} \exp(D_j(t) + \alpha'_{lt}\mathbf{Z}_{kt} + \beta'_l\mathbf{X}_k)}. \qquad (6)$$

---

[13] A potentially issue is that the multinomial model assumes that alternative exit states are stochastically independent, also known as the Independence of Irrelevant Alternatives (IIA) assumption. This assumption rules out any individual-specific unmeasured or unobservable factors that affect both the hazard of having a girl and the hazard of having a boy. To address this issue the estimations include a proxy for fecundity discussed in Section **??**. In addition, the multivariate probit model can be used as an alternative to the multinomial logit because the IIA is not imposed (**?**). The results are essentially identical between these two models and available upon request.

With these two probabilities is it easy to calculate, for each t, the estimated percentage of children born that are boys, $\hat{Y}$,

$$\hat{Y}_t = \frac{P(b_t|\mathbf{X}_k, \mathbf{Z}_{kt}, t)}{P(b_t|\mathbf{X}_k, \mathbf{Z}_{kt}, t) + P(g_t|\mathbf{X}_k, \mathbf{Z}_{kt}, t)} \times 100, \tag{7}$$

and the associated confidence interval for given values of explanatory variables.[14]

For ease of exposition the procedure is presented here in two steps, but the actual calculation of the percent boys is done in one step with the 95 percent confidence interval calculated using the Delta method. Results are presented as the estimated percent boys born by length of birth spacing using graphs.[15] For each graph, the extent to which the percent boys is statistically significantly above the natural sex ratio indicates the use of sex selection.

Another important aspect of the hazard model is the survival curve, which shows the probability of not having had a birth yet by spell duration. The survival curve for t is

$$S_t = \prod_{d=1}^{t} \left(1 - \left(P(b_d|\mathbf{X}_k, \mathbf{Z}_{kd}, d) + P(g_d|\mathbf{X}_k, \mathbf{Z}_{kd}, d)\right)\right), \tag{8}$$

or equivalently

$$S_t = \prod_{d=1}^{t} \left(\frac{1}{1 + \sum_{l=2}^{2} \exp(D_j(t) + \alpha'_{ld}\mathbf{Z}_{kd} + \beta'_l\mathbf{X}_k)}\right). \tag{9}$$

The survival curves are presented for two reasons. First, they show progression to the next birth and how quickly it occurs, if at all. In the absence of sex selection, I expect most parities to show an "inverted s" pattern, where there initially are relatively few births, followed by a substantial number of births over a 1 to 2 year period, and then relatively few births thereafter. As sex selection becomes more widely used the associated longer spacing shows up by making the survival curve straighter, indicating that some of the

---

[14] Within each period $1 - P(b_t) - P(g_t)$ is the probability of not having a birth in period t.
[15] The parameter estimates are available on request.

births that would originally have taken place now take place later because of abortions. The use of sex selection is clearly not the only factor that can change the shape of the survival curves; factors such as the desired number of children and use of contraceptives may also shape the shape. To account for this it is best to compare survival curves for an individual parity between women who are likely to use sex selection and women who are not, for example because they already have one or more sons.

Secondly, the survival curves provide "weighting" for the associated percentage boys born. The steeper the survival curve, the more weight should be assigned to a given spell period because it is based on more births, whereas a period with a flat survival curve should be given little weight because the percentage boys is based on few births. Hence, although they cannot by themselves show sex selection, the survival curves are a crucial complement to the estimated sex ratios by duration.

4   Data

The data come from the three rounds of the National Family Health Survey (NFHS-1, NFHS-2 and NFHS-3), collected in 1992–93, 1998–99, and 2005–2006.[16] The International Institute for Population Sciences in Mumbai collected the data, which have nationwide coverage. The surveys are large: NFHS-1 covered 89,777 ever-married women aged 13–49 from 88,562 households, NFHS-2 covered 90,303 ever-married women aged 15–49 from 92,486 households and NFHS-3 covered 124,385 never-married and ever-married women aged 15–49 from 109,041 households.

I exclude visitors to the household, as well as women married more than once, divorced, or not living with their husband, women with inconsistent information on age of marriage, and those with missing information on education. Women interviewed in NFHS-3 who were never married or where gauna had not been performed were also

---

[16] A delay in the survey for Tripura means that NFHS-2 has a small number of observation collected in 2000.

dropped. The same goes for women who had at least one multiple birth, reported having a birth before age 12, had a birth before marriage, or a duration between births less than 9 months. Women who reported less than 9 months between marriage and first birth remain in the sample unless they are dropped for another reason.[17]

Finally, I restrict the sample to Hindus, who constitute about 80 percent of India's population. If use of sex selection differ between Hindu and other religions, such as Sikhs, assuming that the baseline hazard is the same would lead to bias. The other groups are, however, each so small relative to Hindus that it is not possible to estimate different baseline hazards for each group. Furthermore, the groups are so different in terms of background and son preference that combining them into one group would not make sense.

There are four advantages to using the NFHS. First, surveys enumerators pay careful attention to spacing between births and probe for "missed" births. Second, no other surveys cover as long a period in the same amount of detail. The three NFHS rounds allow me to show the development from before sex-selective abortions were available until 2006. Third, NFHS has birth histories for a large number of women.[18] Finally, even if probing for missing births may not completely eliminate recall error, the overlap in cohorts covered and the large sample size make it possible to establish where recall error remains a problem.

Recall error arise mainly from child mortality, when respondents are reluctant to discuss deceased children.[19] Systematic recall error, where the likelihood of reporting a deceased child depends on the sex of the child, is especially problematic because it biases

---

[17] Women who report less than 9 months between marriage and first birth are retained because between 10 and 20 percent fall into this category. Although it is possible that some of these births are premature the high number of women who report a birth less than 6 months after marriage indicates that conception likely occurred before marriage in most cases.

[18] The Special Fertility and Mortality Survey appears to cover a much large number of households than the three rounds of the NFHS combined, but **?** only use the births that took place in 1997 making their sample sizes by parity smaller than here. Their sample consists of 133,738 births of which 38,177 were first born, 36,241 second born, and 23,756 were third born. The differences in results for first born children are discussed in the online Appendix.

[19] The online Appendix contains a more thorough discussion of recall error and how I address it.

the sex ratios. Probing catches many missed births, but systematic recall error is still a potentially substantial problem. Three factors contribute to the problem here. First, girls have significantly higher mortality risk than boys. Second, son preference may increase the probability that boys are remembered relative to girls. Finally, in NFHS-1 and NFHS-2 enumerators probed only for a missed birth if the initial reported birth interval was four calendar years or more. But, given short durations between births, especially after the birth of a girl, that procedure is unlikely to pick up all missed children.

Observed sex ratios by cohort provide a straightforward way to determine whether recall error is a problem. Because prenatal sex determination techniques did not become widely available until the mid-1980s, a higher than natural sex ratio for cohorts born before that time must be the result of systematic recall error. As shown in the online Appendix, the observed sex ratio by parity becomes more male dominated the further back births took place. In addition, births in the same cohort tend to be more male dominated the more recent the survey (births in the cohort took place longer ago relative to the survey). Hence, there is evidence of recall error and the degree of recall error increases with length of time between survey and cohort.

Using cohort year of birth to analyze recall error and decide which observations to keep is, however, problematic because the year of birth for a given parity is affected by recall error; for example, a second born child listed as first born will be born later than the real first born child. Year of marriage should, however, be unaffected by recall error. Using year of marriage the basic recall error pattern remains with women married longer ago more likely to report that their child was a boy for a given parity. Similarly, comparing women married in the same period across surveys shows that women married longer ago are more likely to report having sons.

That recall error increases the longer ago somebody was married means that duration of marriage is a better predictor of recall error than calendar year of marriage. Figure **??** shows the observed sex ratio for children reported as first born as a function of duration of

marriage combining all three surveys.[20] The solid line is the sex ratio of children reported as first born by the number of years between the survey and marriage, the dashed lines indicate the 95 percent confidence interval and the horizontal line the natural sex ratio (approximately 0.512). To ensure sufficient cell sizes the years are grouped in twos.

Figure 1: Ratio of boys in "first" births

Figure **??** clearly illustrates the systematic recall error problem. The observed sex ratio is increasingly above the expected value the longer ago the parents were married. The increasingly unequal sex ratio with increasing marriage duration suggests that a solution to the recall error problem is to drop women who were married "too far" from the survey year. The main problem is establishing what the best cut-off point should be. The observed sex ratio is consistently significantly higher than the natural sex ratio from around 24 years of marriage, so one possibility is simply to drop all women married more than 24 years at the time of the survey. But, as the Appendix shows, there are differences across the three surveys and between parities. Based on the differences between surveys the cut-off points I use vary by survey round. For NFHS-1 women married 22 years or more were dropped, with the corresponding cut-off points 23 years for NFHS-2 and 26 years for NFHS-3. The final sample consists of 146,096 women, with 332,951 parity one through four births.[21]

4.1 Spell Definition

Spell duration is measured in quarters of a year, that is 3 month periods, hereafter referred to as quarters. The first spell begins at the month of marriage rather than 9 months after because many women report giving birth less than 9 months after they were married. For

---

[20] The graph for second births shows a similar pattern. The graphs for the second births and the individual survey rounds are available upon request.

[21] The online Appendix presents the results for a more restrictive definition and for the women dropped because of recall error concerns.

women who began living with their husbands at too young an age to conceive the starting point should ideally be first ovulation, when she becomes "at risk" for a pregnancy, rather than month of marriage. Unfortunately, information on age of menarche is only available in NFHS-1. Instead, for women who began living with their husband before age 12, I set the the month they turned 12 years of age as staring point for the first spell.

The second and subsequent spells begin 9 months after the previous birth because that is the earliest we should expect to observe a new birth. A few women report births that occurred less than 9 months after the previous birth; those women are dropped.

All spells continue until either a child is born or the spell is censored. Censoring can happen for three reasons: the survey takes place; the woman is sterilized; or the number of births observed becomes too sparse for the method to work. The timing of censoring because of too few births vary slightly by spell but is generally 21 quarters after the beginning of the spell.

I group spells into three time periods based on spell start date: 1972–1984, 1985–1994, and 1995–2006. The first period covers the time before sex-selective abortions became widely available. Abortion was legalized in 1971 and amniocentesis was introduced in India in 1975, but the first newspaper reports on the availability of prenatal sex determination were not until 1982–83 (**???**). The number of clinics quickly increased, and knowledge about sex selection became widespread after a senior government official's wife aborted a fetus that turned out to be male (**?**, p. 598). The second period covers the time from the widespread emergence of sex-selective abortions until the prenatal Diagnostic Techniques (PNDT) act was passed in 1994. The final period is from the PNDT act until the last available survey. The PNDT act made it a criminal offence to reveal the sex of the fetus and was followed by a campaign against the use of sex selection, although enforcement appears to be relatively lax.

By dividing spells into these periods I can examine how fertility decisions and the use of sex selection have changed across the three different regimes. Note that the periods

18

are based on the spells' beginning year, and some spells will therefore cover two periods. A couple may, for example, be married in 1984, but not have their first child until 1986. That couple's first spell will be in the 1972–1984 period, even though most of the spell actually falls in the 1985–1994 period. Some children born from spells that began in the 1972–1984 period may therefore have been conceive when prenatal sex determination techniques were available, which could result in evidence of sex-selective abortions even for this period. Similarly, a spell that began in the 1985–1994 period may have been partly or mostly under the PDNT act. The overall effect is to bias downwards any differences between the periods.

4.2 Explanatory Variables

The explanatory variables are divided into two groups. The first group consists of variables expected to affect the shape of the hazard function: mother's education, sex composition of previous children, and area of residence. Increasing the number of variables interacted with the baseline hazard lowers the risk of bias but requires more data to precisely estimate. I chose these variables because the prior literature shows that they affect fertility and spacing choices and because the prior literature on sex selection indicates that these are correlated with sex selection. The second group of variables are those expected to have a proportional effect on the hazard: age of the mother at the beginning of the spell, length of her first spell (for second spell and above), whether the household owns land, and whether the household belongs to a scheduled tribe or caste.

Increasing education of mothers is strongly associated with lower fertility, with the negative effect of higher opportunity cost on fertility more than outweighing the positive effect of higher income (**?**).[22] Higher education should therefore be associated with

---

[22] Fathers' education has two opposite predicted effects: the associated higher income should increase fertility and therefore lower the pressure to use sex selection, but the higher income also makes the use of sex selection cheaper. In practice, fathers' education had little effect on the hazards and the use of sex-selective abortions and is not included.

higher use of sex selection. I divide women into three education groups: no education, 1 to 7 years of education, and 8 and more years of education. The models are estimated separately for each education level. A potential concern here is reverse causation, where the sex of children affect women's education. Although no direct information is available on when women left school it is possible to estimate school leaving age from the highest completed grade and the usual starting age. As described in the online Appendix, relatively few women could potentially have returned to school after the birth of their first child, and only 56 could possibly have ended up in the wrong education group. Hence, there is little likelihood that reverse causation is a substantial concern here.

As discussed, the sex composition of previous children affects both the timing of fertility and the use of sex-selective abortions. Here I capture sex composition of previous children by dummy variables for the possible combinations for the specific spell, ignoring the ordering of births. As an example, for the third spell three groups are used: Two boys, one girl and one boy, and two girls.

The area of residence is a dummy variable for the household living in an urban area.[23] Area of residence captures both access to prenatal sex determination techniques and the cost of children. The cost of children is higher in urban areas than in rural and access to prenatal sex determination is easier; both are expected to lead to greater use of sex-selective abortions in urban than in rural areas. Because of concerns about selective migration I use where the household was living at the end of each spell.[24]

The sex composition of children, area of residence, and their interactions are interacted with the piece-wise linear baseline hazard dummies. In other words, the baseline hazards are assumed to be different depending on where a woman lives and the sex composition of her previous children. As an example, for the second spell a separate regression is run for each education level and in each regression four different baseline hazards are

---

[23] NFHS uses four categories for area of residence: Large city, small city, town and countryside. To reach a sufficient sample size urban areas are merged into one group.

[24] As online Appendix Table D.1 shows there is little difference across education levels in migration patterns within each period and patterns across periods are also relatively stable.

included (first child a boy in rural area, first child a boy in urban area, first child a girl in rural area, first child a girl in urban area). Although this approach substantially increases the number of regressions and estimated parameters it reduces the potential problem of including other variables as proportional effects. In addition, it forces attention to how much data are really available to estimate the determinants of sex-selective abortions.

The remaining variables are expected to affect hazards proportionally. Although fecundity cannot be observed directly a suitable proxy is the duration from marriage until first birth. Most Indian women do not use contraception before the first birth and there is pressure to show that a newly married woman can conceive (**???**). This is confirmed by the very short spells between marriage and first birth, even among the most educated. Hence, a long spell between marriage and first birth is likely due to low fecundity. The expected effect of a longer first spell is to reduce sex-selective abortions in subsequent spells. For both this variable and the age of the mother at the beginning of the spell the squares are also included. The remaining variables are dummies for household ownership of land and membership of a scheduled caste or tribe.

4.3   Descriptive Statistics

Appendix Table **??** presents descriptive statistics for the spells by education level and when the spell began. There is a substantial number of censored observations. As an example, for highly educated women who had their first child in the 1995–2006 period, almost half did not have their second child by the time of the survey. Hence, although about 13,000 women began the second spell there are only about 7,000 births to these women. Censoring becomes even more important for the third and fourth spells, where around 70 percent of the observations are censored. Generally, censoring increases with parity and time period. This reflects a combination of factors: timing of the surveys relative to the periods of interest, later beginning of childbearing, falling fertility, and the longer spells from sex-selective abortions. The high number of censored observations

21

underscore the importance of controlling for censoring when examining the relationship between fertility and sex selection.

The descriptive statistics also provide a first indication of how the sex ratio at birth changes over time and by spell. For the first spell the sex ratio is very close to the natural for all education groups and all three time periods. As an example, among the highly educated group for the 1995–2006 period, 51.3 percent of the children born were boys.[25] For the second spell, all but the highly educated group in the last two periods have sex ratios in line with the natural sex ratio. Women with 8 or more years of education have 53.1 and 54.3 percent boys in the 1985–1994 and 1995–2006 periods, respectively. This pattern repeats itself for the third spell, except the percentage boys is higher for the high education group (55.3 and 55.9 for the last two time periods). Finally, for the fourth spell the high education group had 60 percent boys in the last period, i.e. after the PNDT act was introduced. Note, however, that for the fourth spell the number of births is substantially smaller and censoring even more important than for the other spells.

India's population has become progressively more urban. For the first period, 32 percent of the women entering the first spell lived in urban areas. This increases to 35 percent for the second period and to 42 percent for the final period. The population is also substantially better educated. Women with no education constituted almost 60 percent in the first period, but less than thirty percent in the last period. Correspondingly, in the first period just over twenty percent had 8 or more years of education, but in the last period it was almost half. Part of the increase in education is correlated with the increase in urbanization, but the proportion of better-educated women has increased substantially in the rural areas as well. Among the high education group almost 70 percent lived in urban areas during the first period but this had fallen to less than 60 percent in the last period.

The increase in urbanization and education is likely to exert downward pressure on fertility and the high censoring rates for the later periods are evidence of this. The average

---

[25] There still appears to be some recall error for the group of women without education for the 1972–1984 period, where 52.3 percent of the children born were boys.

number of children born to women by the time they turn 35 illustrate how strong the decline in fertility has been.[26] Women born in the early 1940s had on average close to 5 children when they reached 35, but women born in the early 1970s had only just over three children. The low number of children is especially remarkable because it combines all education levels and all areas of residence.

5   Results

One of the main questions here is how fertility affects sex selection use. I start by showing the relationship between self-reported desired fertility and sex selection, which also illustrates how to interpret the results. Although self-reported desired fertility is an imperfect proxy for actual fertility behavior, the analysis of desired fertility indicates that there is a relationship between fertility decisions and the use of sex selection and motivates examining the relationship more closely.

Because of the desired fertility measure's potential shortcomings, I next examine the relationship between actual fertility and use of sex selection for better-educated women. There are two steps in this process. First, I show how fertility has declined for better-educated women with one son—a group unlikely to use sex selection, which means that sex selection will not affect the progression rates to next birth for these women. Second, I show how, corresponding to the fertility decline, there is an increase in sex-selective abortions for better-educated women without a son. I next show how women with lower cost of children appear to follow a high fertility, low use of sex selection strategy compared to the low fertility, high use of sex selection of the better-educated women. The results all show clear evidence of son preference in fertility and sex selection use, but exactly what form son preference takes is important, so I next discuss what we can learn about son preference from the sex selection behavior. Finally, I predict lifetime fertility, number of

---

[26] Figure **??** shows the full results.

abortions, and final sex ratios for the better-educated women.

Except for the predictions, results are presented using graphs of estimated percentage boys born by quarter, the associated 95 percent confidence interval, and the survival curve for a "representative woman" using the method detailed in Section **??**. The "representative woman" characteristics are based on means of continuous explanatory variables and the majority category for categorical explanatory variables. Common to all "representative women" is that they do not own land and belong to neither a scheduled caste nor a scheduled tribe. The graphs also show the expected natural rate of boys, approximately 51.2 percent, for comparison. Graphs for all groups and spells, even if not discussed here, are available in the online Appendix.

5.1 Comparing the Hazard and Simple Models

As detailed in the main paper, one of the main advantages of the hazard model approach is that it can be used to predict what completed fertility, number of abortions, and sex ratios will be once women are done with childbearing. This cannot be done with the simple model because it does not predict fertility progression, does not take into account censoring, and cannot capture if parents change their use of sex selection within a spell. We can, however, still compare differences in predicted sex ratios for *individual* parities.[27]

For each parity, the simple model is

$$Y_i = \gamma + \alpha' \mathbf{Z_i} + \beta' \mathbf{X}_i + \epsilon_i, \tag{10}$$

where Y is a dummy variable that takes the value 1 if the child born is a boy and zero if it is a girl, and the vectors of explanatory variables, $\mathbf{Z}_i$ and $\mathbf{X}_i$, are the same as for the proposed method, except that $\mathbf{Z}_i$ obviously does not include the baseline hazard and that there are

---

[27] The simple model uses only observed births and the predicted sex ratio is therefore simply what we observe for the relevant uncensored sample.

no time dependence.[28] I estimate this using a logit model, although a linear probability model leads to similar results. This model is estimated on all observed births for a given parity in the relevant period to ensure that the results are as close to the standard method of examining sex selection (see, for example, **???**).

I focus on those cases where there is evidence of sex-selective abortions: the spells from parity one to two and two to three for women with eight or more years of education.[29] Table **??** shows the number of women who enter the second and third spells for the hazard model, the number of births, observed in the case of the simple model and predicted in the case of the hazard model, predicted percent boys, and the associated predicted sex ratio. It is possible for the number of births used for the simple model to be larger than the hazard models' predicted number of births because the simple model is based on all births that occurred in the period—independently of when the prior birth occurred—whereas the hazard model uses only spells that began in the period.

For the 1985–1994 period—which is when sex selection become available—the hazard model consistently predicts a higher sex ratio than the logit model. Furthermore, some of these differences are substantial. For urban women with 2 girls as their first two children, the hazard model predicts a sex ratio of 175 boys per 100 girls, whereas the simple method "only" predicts 155 boys per 100 girls. The difference is smaller in rural areas but still 14 boys per 100 girls higher for the hazard model than for the simple model. A similar, if less pronounced pattern, show up for the second birth, where the hazard model predict 3 and 5 boys more per 100 girls for urban and rural women.

For the latest period, 1995–2006, the hazard model still produces higher predicted sex ratios for second birth and the differences are of a similar magnitude to the 1985–1994 period. The predicted sex ratios for the third birth are, however, higher for the simple model than for the hazard model. This difference is substantially for urban women where

---

[28] To ease presentation, the indicator for parity number is not shown.

[29] The results for the spell from parity three to four are available on request. They are not presented because the number of educated women who end up with three births and the probability of having a fourth birth are small, which makes comparing the simple and hazard models with any precision difficult.

Table 1: Comparison of Hazard and Simple Models for
Women with 8 or More Years of Education

	Urban		Rural	
	Simple	Hazard	Simple	Hazard
1985–1994 - Second Birth				
1 girl				
Number of women[a]		4,869		3,105
Births[b]	3,848	3,926	2,417	2,795
Percentage boys	54.0	54.7	52.6	53.5
Sex ratio (boys per 100 girls)	117.4	120.7	110.9	115.1
1985–1994 - Third Birth				
2 girls				
Number of women[a]		1,654		982
Births[c]	1,034	1,092	682	853
Percentage boys	60.8	63.7	54.7	57.4
Sex ratio (boys per 100 girls)	155.3	175.3	120.7	134.6
1995–2006 - Second Birth				
1 girl				
Number of women[a]		3,774		2,823
Births[b]	3,031	2,704	2,489	2,407
Percentage boys	58.0	58.6	56.0	56.9
Sex ratio (boys per 100 girls)	138.1	141.3	127.1	132.2
1995–2006 - Third Birth				
2 girls				
Number of women[a]		1,000		863
Births[c]	695	600	690	661
Percentage boys	65.5	62.4	62.0	61.7
Sex ratio (boys per 100 girls)	189.6	166.0	163.4	161.0

Note. The simple models are estimated using logit on all births for a given parity that occurred during the relevant time period. The hazard models are estimated using all spells that began in the relevant time period. For the second birth the hazard model covers the period from beginning of spell to 6 years (24 quarters) after the birth of the first child. For the third birth it covers the period from beginning of spell to 7.25 years (29 quarters) after the birth of the second child for the period 1985–1994 and from beginning of spell to 5.75 years (23 quarters) after the birth of the second child for the period 1995–2006.
[a] Number of women in period who began the spell with given sex composition of prior child/children in the relevant period.
[b] For Logit model the number of women who are observed to have a second birth in the period. For hazard model the predicted number of second births that will occur between beginning of the spell and 6 years (24 quarters) after the birth of the first child.
[c] For Logit model the number of women who are observed to have a third birth in the period. For hazard model the predicted number of third births that will occur between beginning of the spell and 7.25 years (29 quarters) after the birth of the second child for the 1985–1994 period and between beginning of the spell and 5.75 years (23 quarters) after the birth of the second child for the 1995–2006 period.

the difference is 23 boys extra per 100 girls.

A possible explanation for these differences is that the simple model fails to capture the move towards lower fertility and the subsequent increase in the use of sex-selective abortions at earlier parities that occurred over the periods. The results also suggest that

the take-up of sex selection occurred rapidly as the new methods became available and that this take-up is not fully captured by the simple model. More broadly, a possible interpretation of the differences is that the simple model will tend to underestimate when sex-selective abortions are initially spreading in use and overestimate when the spread of sex selection is slowing. This is most clearly seen for the third births for urban women where the hazard model show a *decline* is use of sex-selective abortions, consistent with the lower fertility as discussed in the main paper, while the simple model show a substantial increase.

5.2   One Son or Many Sons?

In the discussion of declining fertility I rely on sex selection not being used by families that already has one son. The extent to which this holds is of interest for two reasons. First and foremost, it can help us understand what form son preference takes in India. This is important in itself, but especially as we try to predict what will happen to fertility and sex selection use in the future. Second, if families use sex selection to reach two or more sons, progression rates for families with one son may overestimate the decline in fertility.

Women with 8 or more years of education is the group the most likely to use sex selection if they have no son as shown above. Hence, if sex selection is also used to achieve more than one son, we should be most likely to find evidence for this group of women as well. Figure **??** therefore shows sex ratios and survival curves for the 1995–2006 period for urban and rural women with 8 or more years of education, conditional on having at least one son. The first column shows outcomes for the second spell if the first born is a son, the second column shows third spell when the first two children were a son and a daughter, and the third column shows third spell when the first two children were sons.

For the second spell the predicted sex ratios closely follow the natural sex ratio for both

Urban

(a) Spell 2 - 1 boy (N=3,969)	(b) Spell 3 - 1 boy/1 girl (N=2,357)	(c) Spell 3 - 2 boys (N=1,098)
Prob. boy (%)	Prob. boy (%)	Prob. boy (%)
Prob. no birth yet	Prob. no birth yet	Prob. no birth yet

Rural

(d) Spell 2 - 1 boy (N=3,044)	(e) Spell 3 - 1 boy/1 girl (N=1,819)	(f) Spell 3 - 2 boys (N=790)
Prob. boy (%)	Prob. boy (%)	Prob. boy (%)
Prob. no birth yet	Prob. no birth yet	Prob. no birth yet

Figure 2: Predicted probability of having a boy and probability of no birth by quarter (3 month period) for women with 8 or more years of education for 1995–2006. N is number of women in the relevant group in the underlying samples.

urban and rural women. Furthermore, the survival curves follow the expected pattern if no sex selection is taking place. There is also no evidence of sex selection for the third spell for women with two sons; the sex ratios follow the natural sex ratio for both urban and rural women. Sample sizes are, however, less than ⅓ of the sample sizes for the second spell and the progression rate to a third birth is low at only 30% for urban women and 55% for rural women, resulting in wide confidence intervals. The survival curves are particularly flat for the first year, meaning that the apparent deviations from the natural sex ratio in the beginning of the spell are based on very few births.

Even though there is no evidence of using sex selection to have an additional son when the previous children are all sons, it is possible that families want more sons than daughters and therefore use sex selection for the third spell when they already have one of each. Urban women show little evidence of this being the case. Although the sex ratio is above the natural from quarter 11 until the end of the spell, only about ⅓ of the births occur in this interval. Furthermore, the dip below the natural sex ratio prior to this point

also accounts for approximately ⅓ of the births. Hence, these two parts cancel each other out. With the first part being to the natural sex ratio the end result is that there is little evidence for urban women using sex selection if they already have a son and a daughter.

For rural women, however, the dip and the associated number of births is not large enough to bring the overall sex ratio for the spell back to natural sex ratio. The sex ratio for the entire spell is therefore likely more male dominated than the natural rate, although it is only statistically significantly higher for two individual quarters of the spell. There are two possible explanations for rural women appearing more likely than urban women to use sex selection on the third spell if they already have a son and a daughter. First, rural women could have a stronger son preference than urban women and want two sons rather than only one. Second, mortality is higher in rural than urban areas and this higher mortality, either experienced or anticipated, make rural women use sex selection to secure a second son, essentially as an insurance. Separating these two possible explanations is difficult, but the "heir and a spare" explanation appears more likely than difference in preferences given that urban women use sex selection more than rural women for the second birth in the absence of a son. An alternative explanation is that what we are observing is simply random variation. This would explain why we get see differences in sex ratios between urban and rural women with one son across the third and fourth spells as shown in the online Appendix.[30]

Son preference can manifest itself through other channels than sex selection. Differential investments in education and health across boys and girls, resulting in lower education and higher mortality of girls relatively to boys, could be thought of as evidence of son preference, although **?** argue that both may represent parents' rational response

---

[30] Urban women with a son and two daughters as their first 3 births show a sex ratio that is significantly higher than the natural, but no evidence of sex selection if they have two sons and a daughter as their first children. Rural women show only a marginally higher sex ratio with a son and two daughters and no effect at all with two sons and a daughter. These results are, however, based on very small samples. There are 590 urban women with a son and two daughters, of which less than 40% have a fourth child, 344 urban women with two sons and one daughter, and only 20% of those have a fourth child. For rural women the numbers are 564, with 40% having a fourth child, and 367 with approximately 30% having a fourth birth. This comes to approximately 300 urban births and only slightly more for rural across the two sex compositions.

to market opportunities rather than inherent preferences of parents. Whatever way son preference presents itself, it appears that the dominant preference—as expressed through use of sex selection—is for one son, rather than multiple sons.

5.3   Predicted Lifetime Number of Children and Abortions

One of the main advantages of the method proposed is that we can predict what completed fertility, total number of abortions, and the sex ratios will be for women in the samples at the end of childbearing. This cannot be done with the simple model because it does not predict fertility progression, does not take into account censoring, and cannot capture if parents change their use of sex selection within a spell.

The predictions are important for three reasons. First, they allows us to quantify whether overall use of sex selection is intensifying or easing over time by examining how the total number of abortions women will have over their childbearing, and the resulting final sex ratios, change. This is a more accurate representation than what we get from sex ratios for individual parities. As the theory section shows, it is possible to get no change—or even a decline—in use of sex selection for an individual parity, at the same time as the overall use of sex selection is increasing. Second, completed fertility and sex ratios are what matters when examining the future impacts of sex selection in areas such as the chance of finding a spouse, savings behavior, and the potential problems associated with a surplus of bachelors (**????**). Finally, as I show in the online Appendix and discuss above, relying only on observed sex ratios of recorded births is problematic because these are based on selection and the observed sex ratio may be biased predictors of final sex ratio when people change their use of sex selection during a spell.

Table **??** shows predicted birth progression rates, number of abortions, sex ratios, and completed fertility for the three samples of women with 8 or more years of education: the 11,271 women who were married in the the 1972-1984 period, the 19,072 women who were married in the 1985–1994 period, and the 16,620 women who were married in the

30

1995–2006 period.[31] Predictions for each period assumes that behavior remains constant and that all spells a woman goes through are within the same period. Each woman is assumed to follow the fertility and sex selection behavior for each spell up to and including the fourth spell using the period's estimates and her characteristics. All results are for women predicted to have one or more births.[32]

The predictions are done as follows. Starting with the first spell I calculate each woman's probabilities of having a boy, $P(b_t)$, a girl, $P(g_t)$, or no birth, $1 - P(b_t) - P(g_t)$, in each quarter after marriage based on the period results for equations (**??**) and (**??**) using her individual values for $\mathbf{Z}_{it}$ and $\mathbf{X}_i$. She then is assigned one of these outcomes—boy, girl, or no birth—based on random draws from a uniform distribution. The first quarter she has a draw that results in a birth she leaves the spell and the quarter reached is used to calculate her starting age for the second spell. If she has no draw that result in a birth by the end of the spell her childbearing stops and she does not enter the next spell. If she does have birth, the process is repeated for the second through fourth spells. Randomization is necessary because each spell's results, except for the first, depends on the sex composition of previous children and the duration of the first spell. The results presented are averaged outcomes over all urban women and all rural women using 1,000 repetitions.

Birth progression rates are the percentage of women predicted to have a birth by the end of each spell. Fertility is the average number of children born according to the randomization. With the large number of replications this number is the same as using progression rates at the end of the spell to calculate the average number of children.[33]

For abortions I first calculate each quarter's probability of abortions using $\frac{100}{105} P(b_t) -$

---

[31] These are the women used to estimate the results for the spell from marriage to first birth above. For more information see Table **??**.

[32] This means that fertility predictions are larger than standard estimates, which include women with no children. Working in the opposite direction is that number of children is restricted to be at most four.

[33] Given that all numbers are conditional on having at least one child, using progression rates for, say, urban women from the 1985-1994 period leads to $1 + 1 \times 0.782 + 1 \times 0.782 \times 0.499 + 1 \times 0.782 \times 0.499 \times 0.448 = 2.3$ children.

$P(g_t)$, where $\frac{100}{105}P(b_t)$ is the probability of having a girl that should have prevailed in the absence of sex selection (0.952 girls born per boy). This probability is multiplied with the probability of not having had a birth yet at the beginning of the quarter. Abortions for each spell are then found by summing over the entire spell.[34] To account for random variation, the abortion rate is the sum over both positive (more boys than the natural rate) and negative (more girls than the natural rate) values of the abortion calculation within the spell. Finally, the percentage of children born that are boys is based on the randomization of births described above.[35]

The results show a substantial reduction in fertility over time. Predicted completed fertility went from 2.7 for urban women married in the 1972–1984 period to just 2 for urban women married in the 1995–2006 period. Hence, urban fertility is now below replacement, especially considering that these predictions are conditional on having at least one child. Rural women showed a similar decline in completed fertility, going from 3.2 to 2.6. Corresponding to the decline in overall fertility, there were also substantial reductions in the individual parity progression rates. For example, urban women with 3 children went from having a 57% chance of having a 4th child for the 1972–1984 cohort to a 33% chance for the 1995–2006 cohort—and they had a substantially lower probability of even making it to 3 children in the first place.

Associated with the decline in completed fertility is an increase in sex selection. For ur-

---

[34] One way to see how this works is to imagine that we start the spell with 100 women. After the first quarter we are left with $N_2 = (1 - P(b_1) - P(g_1)) \times 100$, where N_2 is the number of women who start the second quarter. After the second quarter we then have $N_3 = (1 - P(b_2) - P(g_2)) \times N_2$, etc.

Say that 65.88 women of the original 100 still have not had a birth yet by the start of quarter 6 and that the predicted probability of having a son in quarter 6 is 0.0931 and the probability of having a girl is 0.0566 (the probability of not having a birth in quarter 6 is 0.8503). For this quarter the abortion probability is then $\frac{100}{105} \times 0.0931 - 0.0566 = 0.0321$. In other words, the probability of having a girl should have been $\frac{100}{105} \times 0.0931 = 0.0887$, but only a probability of 0.0566 was predicted. Hence, in quarter 6 the predicted number of sons born is $0.0931 \times 65.88 = 6.13$, the predicted number of daughters born is $0.0566 \times 65.88 = 3.73$, and $0.0321 \times 65.88 = 2.11$ abortions must have taken place to reach these births.

[35] This number is the predicted sex ratio at birth and differ from actual observed sex ratios when the children are older because of mortality. On one hand, the anticipated or experienced death of a son increases the likelihood of sex selection even if parents already had a son, which would make the later observed sex ratio lower than what is predicted here. On the other hand, mortality risk is higher for girls than for boy, which would make the sex ratio more male dominated.

Table 2: Predicted Fertility Behavior, Sex Ratio, and
Abortions for Women with 8 or More Years of Education

	1972-1984				1985-1994				1995-2006			
	Births (%) / fertility[a]	Abortions per 100		Boys (%)	Births (%) / fertility[a]	Abortions per 100		Boys (%)	Births (%) / fertility[a]	Abortions per 100		Boys (%)
		births[b]	women[c]			births[b]	women[c]			births[b]	women[c]	
Urban												
Spell 2	85.9	2.0	1.8	52.3	78.2	4.7	3.7	53.7	68.0	8.0	5.4	55.3
Spell 3	64.4	5.1	3.3	53.9	49.9	9.8	4.9	56.2	38.6	7.8	3.0	55.3
Spell 4	57.0	2.8	1.6	52.6	44.8	2.6	1.1	52.6	33.7	21.8	7.3	62.3
Overall[d]	2.7	2.0	5.5	52.3	2.3	3.4	8.0	53.1	2.0	4.6	9.4	53.7
Rural												
Spell 2	92.5	−1.4	−1.3	50.5	87.6	3.1	2.7	52.8	83.1	4.4	3.6	53.5
Spell 3	80.9	2.6	2.1	52.5	70.6	6.5	4.6	54.5	59.6	10.6	6.3	56.7
Spell 4	75.7	−3.1	−2.4	49.6	62.0	0.2	0.1	51.3	51.9	6.1	3.2	54.4
Overall[d]	3.2	−0.3	−1.1	51.2	2.9	2.4	6.8	52.8	2.6	4.1	10.5	53.4

Note. Predictions are based on estimates of equations (**??**) and (**??**). The samples consist of all women from the data with 8 or more years of education who married during the indicated period using the estimation results for that periods and the women's characteristics at the start of their marriage. All numbers presented are for women predicted to have one or more births. Birth/progression rates are based on the ratio of women predicted to have a birth by the end of each spell. Abortion rates within each spell are based on the predicted number of abortions each woman would have if she went through the entire spell. To generate conditions such as sex composition and fecundity for second and higher spells each woman is randomly assigned outcomes based on the estimation results and her characteristics as follows. For each quarter, her probabilities of having a boy, a girl, or no birth is calculated. She is assigned one of these outcomes based on a random draw from a uniform distribution. The first quarter where she has a draw that results in a birth she leaves the spell and the quarter reached is used to calculate her starting age for the subsequent spell. If she has no draw that result in a birth within a spell her childbearing stops and she does not enter the next spell. Results are averaged over 1,000 repetitions.
[a] For individual spells the column shows the percent of women who started the spell and ended up with a birth by the end of the spell period covered. For the "Overall" row, the number is the expected number of children born per woman over the first four spells, conditional on having at least one birth.
[b] For individual spells the column shows the number of abortions per 100 births predicted to take place by the end of the spell. For the "Overall" row, the number is the total number of abortion per 100 births predicted to take place over all four spells.
[c] For individual spells the column shows the number of abortions predicted to take place by the end of the spell per 100 women who started the spell . For the "Overall" row, the number is the total number of abortion predicted to take place over all four spells per 100 women who had at least one child.
[d] Covers first through fourth spells.

ban women the predicted number of abortions per 100 women by the end of childbearing have gone up by over 17 percent between the last two cohorts (8.0 to 9.4) and the number of abortions per 100 births have gone up by over 35 percent (3.4 to 4.6). For rural women, abortions per 100 women increased almost 55 percent (6.8 to 10.5), while abortions per 100 births increased over 70 percent (2.4 to 4.1).[36] The combined effects of the declines in fertility and increased use of sex selection are that the percentage boys increases from 53.1 to 53.7 percent for urban women, and from 52.8 to 53.4 percent for rural women.

Of particular note is the high number of abortions among rural women for the 1985–

[36] Unfortunately there are no solid official estimates to compare these numbers to. The total number of abortions per year, not just those based on son preference, vary between the official count of 600,000–700,000 and an estimate of 6.4 million from the 2002 Abortion Assessment Project-India. The 6.4 million abortions per year may even be an underestimate (**?**). NFHS does ask about abortions in all three surveys, although the questions change substantially between surveys. However, the PNDT act prohibits using prenatal diagnostic tests to determine a fetus' sex and revealing a fetus' sex to parents. Hence, respondents have a strong incentive not to disclose abortions motivated by son preference. In addition, abortion are considered a taboo subject and many women are unwilling or uncomfortable reporting their abortions to government sponsored enumerators (**??**).

1994 period. Educated rural women obviously had few constraints on access to prenatal sex determination even when the techniques were relatively new. The higher abortion rate per woman for rural women than urban women likely comes from a combination of higher likelihood of having an additional child and changes in use of sex selection within spell as shown in the graphs above.

The results from theoretical model showed that it is possible for the use of sex selection to fall for an individual parity, even as overall use of sex selection increases. This is the case here, where the use of sex selection fell for the third spell from 9.8 per 100 births to 7.8 for urban women between the last two periods. As already discussed, total use of sex selection went up between the last two periods, so the reduction for the third spell is not an indication of reduced use of sex selection overall, but simply the result of fewer women making it to a third birth and those who did were more likely to already have a son—for the second spell the number of abortions per 100 births for urban women went from 4.7 to 8.0, a 70 percent increase.[37] This underscores why focusing exclusively on individual parities can be misleading.

The results for rural women illustrate the two different strategies for achieving one son: either have many children or have few children and use sex selection. For the 1986–1994 period there is no use of sex selection on fourth birth, which fits with 62 percent of women end up with three or more children and a full 62 percent of women with three children have a fourth birth (recall that there is a 94 percent chance of having at least one son if you have four children). The use of sex selection on second and third spells indicates that one group of rural women follows the high fertility route, whereas the other group follows the pattern of urban women and combine restricting fertility with use of sex selection.

Another way to see the spread of access to sex selection over time is to compare urban

---

[37] The amount of sex selection in the fourth spell reinforces this point: abortions per 100 births went up almost tenfold. Only 26 percent of urban women with at least one child make it to the beginning of the fourth spell in the last period, and less than 10 of all women have a fourth birth, but for those who do make it to the fourth spell the pressure to ensure a son is large.

women in the first period with rural women in the last period. At 2.7 fertility is already relatively low for urban women in the 1972–1984 period and there is some evidence of sex selection for these women. As discussed in Section **??**, the higher than natural sex ratio for urban well-educated women this early is likely a combination of early access to prenatal sex determination and that some of the spells end in the 1985–1994 period when there was more widespread access. Rural women in the 1995–2006 period had a fertility only slightly lower than that of urban women in the 1972–1984 periode, but the number of abortions both per birth and per woman were twice as high.

For comparison Appendix Tables **??** and **??** show predicted fertility, number of abortions, and sex ratio for the two other education groups. As expected, fertility for women with no education is substantially higher and the decline over time much smaller than for the high education group. Predicted fertility fell from 3.6 to 3.3 for urban women and from 3.7 to 3.5 for rural women between the 1972–1984 and the 1995–2006 periods.[38] Consistent with the discussion above, there is evidence of sex selection for neither urban nor rural women, with the sex ratio for all predicted births in the last two periods close to normal.[39] Random variation makes it appears that some individual parities had a higher than normal sex ratio, which, if studied in isolation, could be taken as evidence of use of sex selection, and reinforces why supplementing analysis of individual parities with this analysis of the overall pattern is important.

Fertility for women with between 1 and 7 years of education was slightly lower than the no education group in the 1972–1984 period, but the decline over time was larger. For the 1995–2006 period predicted completed fertility was 2.8 for urban women and 3.2 for rural women. There is only limited evidence of sex selection for urban women in this education group. The exception is for rural women in the last period where the overall sex ratio is 52.1% boys, although given the high fertility for this group and the small sample

---

[38] A caveat is that my estimates did not cover parities higher than 4, which could bias downward the change in predicted fertility.

[39] The higher than normal sex ratios for the 1972–1984 period are most likely remnants of recall error not removed by the sample restriction.

size for this education group it is unclear whether this simply reflects random variation.

The predictions discussed here provide further support for the idea that there are two alternative strategies to ensure the birth of a son, either have many children or use sex-selective abortions. The best educated group mostly follow the low fertility, high use of sex selection strategy, whereas the two lower education groups still have relatively high fertility and little to no use of sex selection. It is, however, clear that fertility is declining over time for all groups. With lower fertility in both urban and rural areas, we are likely to see further increases in the use of sex-selective abortions among those who already use sex selection, and among the groups who have not previously use sex selection we will see beginning use of sex-selective abortions.

6   Conclusion

If women have preferences over both the number of children they have *and* the sex composition of these children, they face a trade-off between the cost—both monetary and psychological—of sex selection and the cost of children. This paper introduces a novel approach to understanding the relationship between fertility and use of sex selection, and how sex selection interact with birth spacing. The proposed method allows for joint estimation of fertility and sex-selective abortions using a non-proportional, competing risk hazard model.

Three results stand out. First, lower fertility is an important factor in the decision to use sex-selective abortions. In both urban and rural areas, only women with the lowest predicted fertility—those with 8 or more years of education—use sex selection, and as fertility falls they use it for earlier parities. Women with less education instead follow a high fertility strategy to ensure they have a son and do not appear to use sex selection. Previous research was unable to explain why sex selection in India only occurred among higher education women because it failed to tie the use of sex selection and the fertility

decision together.

Second, the legal steps taken to combat sex-selective abortion have not been able to reverse the practice. The use of sex selection has been increasing over time and is higher now than before the PNDT Act was passed. The predicted number of sex-selective abortions per 100 women during their childbearing years is now around 10 for women with 8 or more years of education in both urban and rural areas. This is especially interesting given that the cost of prenatal sex determination has likely increased in response to the PNDT Act. This also emphasizes why accounting for the possibility that parents may change their decision to use sex selection within a spell is important. As the theory shows an increase in the cost of prenatal sex determination increases the likelihood of parents changing their decision to use sex selection.[40]

Third, there is little evidence that women with one or more sons use sex selection, and their probability of another birth declines substantially once they have a son. The main exception is for rural women with one son and one daughter, presumably as an insurance against mortality. Hence, parents appear to have a preference for one son rather than multiple sons.

These results imply that the way we have been measuring son preference may not be useful for understanding how people make decisions on use of sex selection. Recent research suggests that son preference in India, when measured as ideally having more boys than girls, is decreasing over time and with higher education (**??**).[41] Despite this, use of sex selection has increased and for exactly the group argued to show declining preference for sons. The culprit may be that parents want *one* son and will use sex selection intensively to get that son, but refrain from using sex selection to have more sons, and our

---

[40] A feature of the new approach is that it allows for changes in the decision to use sex selection within a spell. This is most clearly seen for better-educated urban women whose only child is a girl, for whom the use of sex selection declines as the second spell becomes longer.

[41] This measure of son preference is commonly used in the literature. See, for example, **?**, **?**, and **?**. In NFHS, and Demographic and Health Surveys in general, son preference is measured as part of the general fertility questions. The first question is how many children a woman would have if she could choose exactly how many. This if followed by the sex preference question: "How many of these children would you like to be boys, how many would you like to be girls and for how many would the sex not matter?"

standard measure of son preference cannot captured that. A better approach may be to directly ask if there is a minimum number of each sex that respondents want, possibly in combination with ideal number questions.

This is of more than theoretical interest. If son preference is expressed as a very strong desire for having at least one son then reductions in fertility will lead to a further increase in the use of sex selection across all groups that hold that preference. What is more, as shown in the theory section, it is possible to get increasingly unequal sex ratio when fertility falls even if parents do not want more sons than daughters as long as they only use abortion on female fetuses. A deeper understanding of exactly what type of son preference is responsible for the increase use of sex selection is clearly an important question for future data collection and research.

An often-repeated, but poorly substantiated, argument for why there is sex selection in India is the dowry system. This explanation fits poorly with observed behavior. If true, dowries increase the cost of having girls relative to boys because you have to pay to ensure your girls are married. This implies that poorer households should be the most likely to use sex selection, and that the use of sex selection should increase the more girls they have.[42] Similarly, for all households there would be a financial incentive to achieve a higher number of sons than daughters. As shown here, however, sex selection is predominantly used by better educated and therefore, on average, wealthier households and only to secure one son rather than multiple sons. Finally, fertility's rapid decline after the introduction of prenatal sex determination, combined with low cost of sex selection, suggests that parents' perceived costs of boys and girls are similar.[43] Hence, banning dowries, as India has done, is unlikely to affect the use of sex selection.

The results presented here lead to a number of important questions for future research. One is whether marriage market sex ratios affect the use of sex selection. Parents may care

---

[42] The exception would be if dowries increase more than proportionally with education level and income, but there does not appear to be evidence of this.

[43] See the discussion in Section **??** for more detail.

not only about the number and type of children they have but also whether they will have grandchildren. If they care about grandchildren, then a boy could still be preferred over a girl, but a married girl would be preferred over an unmarried boy (**?**). The implication is that parents respond to changes in the expected sex ratio of their children's marriage market, even given their preference for a son. This hypothesis can be tested using measures of the observed distributions of boys and girls and the method proposed here.

A second question for future research, which is especially important when designing policies aimed at reducing the use of sex selection, is who the agents are in the decisions to use sex selection. Here, I have treated the household as a single unit, but it is possible, and indeed likely, that husband and wife have different preferences over both the factors that determine use of sex selection and sex selection itself. Some of the popular press have assumed that the decision to use sex selection rests with the husband. This, however, does not fit well with the general pattern that women, on average, prefer to have fewer children than men do and that declining fertility appear to be the main driver of sex selection. A better understanding of who holds what preferences in the family would be a good first step to disentangling who makes decisions about sex selection.

A third question is whether there are interactions between sex selection and child outcomes, such a survival and education. It has been argued that access to sex selection may benefit girls because those girls who are born are more likely to be wanted by their parents. For mortality it is, however, possible that improvements are the result of a purely mechanical effect. As I have shown, increased use of sex selection lengthens the birth spacing after a girl and that by itself can improve her survival chances, even if she is no more wanted than before. The method suggested here can predict the likelihood of a woman using prenatal sex determination for a given birth based on her characteristics. This predicted probability can then be used in estimations of the determinants of child health to directly test whether increasing use of sex selection is beneficial for the girls who are born. It is unlikely, however, that a substantial mortality effect exists since sex

selection is mainly used by well-educated women, who tend to have lower mortality, but there might be effects on education investments.

In conclusion, the results provide strong clues to how sex selection use will change in the future. Because lower fertility is responsible for the increase in sex-selective abortions, it is likely that we will see further increases in the practice as more families want fewer children, either because of urbanization or because of increases in female education. With already low fertility for better-educated urban women, a substantial future increase in sex-selective abortions per woman is unlikely, but a higher proportion of women will belong to this group in the future. For rural, better-educated women fertility is still falling. To the extent that it falls to the same level as for urban women, I expect a corresponding increase in sex selection. Finally, and most importantly, we are beginning to see evidence of lower fertility for women with lower levels of education. If women with less education hold the same preference for one son as the better-educated, once the less-educated women's fertility begins to drop to three—and even two—we are likely to see a substantially increased use of sex selection in India.

A Appendix

Table A.1: Descriptive Statistics by Education Level and Beginning of Spell

		No Education			1–7 Years of Education			8+ Years of Education		
		1972–1984	1985–1994	1995–2006	1972–1984	1985–1994	1995–2006	1972–1984	1985–1994	1995–2006
First Spell	Boy born	0.467	0.438	0.363	0.476	0.447	0.368	0.483	0.453	0.377
		(0.499)	(0.496)	(0.481)	(0.499)	(0.497)	(0.482)	(0.500)	(0.498)	(0.485)
	Girl born	0.427	0.410	0.352	0.439	0.429	0.356	0.454	0.420	0.353
		(0.495)	(0.492)	(0.478)	(0.496)	(0.495)	(0.479)	(0.498)	(0.494)	(0.478)
	Censored	0.106	0.152	0.284	0.085	0.124	0.276	0.063	0.127	0.270
		(0.308)	(0.359)	(0.451)	(0.279)	(0.330)	(0.447)	(0.243)	(0.333)	(0.444)
	Urban	0.181	0.176	0.200	0.366	0.334	0.322	0.690	0.604	0.561
		(0.385)	(0.381)	(0.400)	(0.482)	(0.472)	(0.467)	(0.463)	(0.489)	(0.496)
	Age	15.854	16.301	16.871	16.983	17.470	17.844	19.516	20.036	20.766
		(2.482)	(2.643)	(2.858)	(2.733)	(3.041)	(3.221)	(3.331)	(3.614)	(3.916)
	Owns land	0.589	0.578	0.540	0.500	0.497	0.482	0.321	0.387	0.414
		(0.492)	(0.494)	(0.498)	(0.500)	(0.500)	(0.500)	(0.467)	(0.487)	(0.493)
	Sched. caste/tribe	0.357	0.395	0.441	0.170	0.227	0.318	0.071	0.119	0.175
		(0.479)	(0.489)	(0.497)	(0.376)	(0.419)	(0.466)	(0.257)	(0.324)	(0.380)
	Number of quarters	330,315	257,757	75,841	108,463	108,863	54,704	102,238	154,582	116,272
	Number of women	30,004	27,669	9,684	11,162	13,104	7,510	11,271	19,072	16,620
Second Spell	Boy born	0.490	0.444	0.359	0.477	0.437	0.338	0.457	0.399	0.285
		(0.500)	(0.497)	(0.480)	(0.500)	(0.496)	(0.473)	(0.498)	(0.490)	(0.451)
	Girl born	0.451	0.413	0.327	0.462	0.413	0.319	0.426	0.351	0.239
		(0.498)	(0.492)	(0.469)	(0.499)	(0.492)	(0.466)	(0.495)	(0.477)	(0.427)
	Censored	0.059	0.143	0.314	0.061	0.150	0.344	0.117	0.251	0.476
		(0.236)	(0.350)	(0.464)	(0.239)	(0.357)	(0.475)	(0.321)	(0.433)	(0.499)
	One boy	0.527	0.515	0.508	0.519	0.513	0.505	0.513	0.520	0.515
		(0.499)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
	One girl	0.473	0.485	0.492	0.481	0.487	0.495	0.487	0.480	0.485
		(0.499)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
	Urban	0.181	0.180	0.212	0.370	0.348	0.345	0.697	0.628	0.578
		(0.385)	(0.384)	(0.409)	(0.483)	(0.476)	(0.475)	(0.459)	(0.483)	(0.494)
	Age	17.774	18.357	18.870	18.620	19.200	19.581	20.988	21.589	22.252
		(2.743)	(3.054)	(3.284)	(2.864)	(3.206)	(3.373)	(3.371)	(3.621)	(3.901)
	First spell length	25.630	29.145	29.133	21.924	24.513	24.787	20.608	22.154	22.217
		(20.345)	(25.044)	(27.216)	(18.142)	(22.027)	(23.204)	(15.371)	(18.651)	(19.176)
	Owns land	0.590	0.572	0.528	0.497	0.492	0.468	0.313	0.369	0.402
		(0.492)	(0.495)	(0.499)	(0.500)	(0.500)	(0.499)	(0.464)	(0.483)	(0.490)
	Sched. caste/tribe	0.353	0.390	0.430	0.164	0.218	0.304	0.068	0.110	0.169
		(0.478)	(0.488)	(0.495)	(0.371)	(0.413)	(0.460)	(0.251)	(0.313)	(0.374)
	Number of quarters	176,227	231,546	68,455	67,432	103,656	49,236	77,416	164,752	119,864
	Number of women	21,171	28,251	9,428	8,209	12,449	6,542	8,204	16,601	13,610

Note. Means without parentheses and standard deviation in parentheses. Interactions between variables, baseline hazard dummies and squares not shown. Quarters refer to number of 3 month periods observed.

Table A.1: (Continued) Descriptive Statistics by Education Level and Beginning of Spell

		No Education			1-7 Years of Education			8+ Years of Education		
		1972–1984	1985–1994	1995–2006	1972–1984	1985–1994	1995–2006	1972–1984	1985–1994	1995–2006
Third Spell	Boy born	0.478	0.425	0.319	0.455	0.388	0.261	0.353	0.264	0.163
		(0.500)	(0.494)	(0.466)	(0.498)	(0.487)	(0.439)	(0.478)	(0.441)	(0.369)
	Girl born	0.442	0.393	0.306	0.423	0.357	0.239	0.313	0.214	0.128
		(0.497)	(0.488)	(0.461)	(0.494)	(0.479)	(0.426)	(0.464)	(0.410)	(0.334)
	Censored	0.080	0.182	0.376	0.121	0.255	0.500	0.334	0.521	0.709
		(0.271)	(0.386)	(0.484)	(0.326)	(0.436)	(0.500)	(0.472)	(0.500)	(0.454)
	Two boys	0.276	0.257	0.243	0.248	0.245	0.232	0.266	0.244	0.238
		(0.447)	(0.437)	(0.429)	(0.432)	(0.430)	(0.422)	(0.442)	(0.429)	(0.426)
	One boy, one girl	0.490	0.502	0.506	0.508	0.501	0.515	0.489	0.510	0.527
		(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.499)
	Two girls	0.234	0.241	0.252	0.243	0.254	0.253	0.245	0.247	0.235
		(0.424)	(0.428)	(0.434)	(0.429)	(0.435)	(0.435)	(0.430)	(0.431)	(0.424)
	Urban	0.179	0.182	0.208	0.378	0.354	0.343	0.693	0.646	0.568
		(0.383)	(0.386)	(0.406)	(0.485)	(0.478)	(0.475)	(0.461)	(0.478)	(0.495)
	Age	19.978	20.693	21.278	20.794	21.453	21.998	22.963	23.825	24.548
		(2.895)	(3.196)	(3.551)	(2.937)	(3.239)	(3.618)	(3.405)	(3.748)	(4.067)
	First spell length	24.547	27.575	27.099	20.971	23.901	23.532	19.991	21.246	21.236
		(18.832)	(22.779)	(24.364)	(16.875)	(20.378)	(21.155)	(14.646)	(16.835)	(17.363)
	Owns land	0.602	0.575	0.538	0.503	0.503	0.490	0.320	0.369	0.421
		(0.490)	(0.494)	(0.499)	(0.500)	(0.500)	(0.500)	(0.467)	(0.483)	(0.494)
	Sched. caste/tribe	0.344	0.392	0.432	0.159	0.218	0.298	0.067	0.107	0.169
		(0.475)	(0.488)	(0.495)	(0.366)	(0.413)	(0.457)	(0.251)	(0.310)	(0.375)
	Number of quarters	109,026	206,988	70,391	43,434	87,537	41,400	49,799	126,733	79,068
	Number of women	13,196	25,135	9,355	4,960	9,775	5,034	4,316	10,688	7,927
Fourth Spell	Boy born	0.463	0.384	0.291	0.403	0.347	0.233	0.310	0.227	0.170
		(0.499)	(0.486)	(0.454)	(0.491)	(0.476)	(0.423)	(0.462)	(0.419)	(0.376)
	Girl born	0.408	0.363	0.268	0.391	0.303	0.198	0.298	0.197	0.113
		(0.491)	(0.481)	(0.443)	(0.488)	(0.460)	(0.399)	(0.458)	(0.397)	(0.316)
	Censored	0.130	0.254	0.441	0.206	0.350	0.569	0.392	0.576	0.717
		(0.336)	(0.435)	(0.497)	(0.405)	(0.477)	(0.495)	(0.488)	(0.494)	(0.450)
	Three boys	0.134	0.124	0.111	0.107	0.105	0.102	0.106	0.101	0.082
		(0.341)	(0.329)	(0.315)	(0.309)	(0.307)	(0.303)	(0.308)	(0.302)	(0.275)
	Two boys, one girl	0.374	0.357	0.346	0.348	0.323	0.308	0.358	0.304	0.289
		(0.484)	(0.479)	(0.476)	(0.476)	(0.468)	(0.462)	(0.480)	(0.460)	(0.453)
	One boys, two girls	0.362	0.389	0.400	0.400	0.413	0.423	0.393	0.439	0.469
		(0.481)	(0.488)	(0.490)	(0.490)	(0.492)	(0.494)	(0.489)	(0.496)	(0.499)
	Three girls	0.129	0.130	0.142	0.145	0.159	0.167	0.143	0.156	0.159
		(0.335)	(0.337)	(0.350)	(0.353)	(0.365)	(0.373)	(0.350)	(0.363)	(0.366)
	Urban	0.171	0.181	0.197	0.363	0.346	0.325	0.670	0.601	0.511
		(0.377)	(0.385)	(0.398)	(0.481)	(0.476)	(0.469)	(0.470)	(0.490)	(0.500)
	Age	21.926	22.885	23.608	22.611	23.519	24.317	24.257	25.472	26.061
		(3.019)	(3.332)	(3.711)	(2.894)	(3.364)	(3.759)	(3.171)	(3.702)	(4.143)
	First spell length	23.281	26.609	26.460	20.495	23.069	24.215	19.106	20.826	21.423
		(17.481)	(21.556)	(23.272)	(15.546)	(19.078)	(21.244)	(13.991)	(16.277)	(16.585)
	Owns land	0.610	0.587	0.557	0.523	0.530	0.509	0.339	0.414	0.466
		(0.488)	(0.492)	(0.497)	(0.500)	(0.499)	(0.500)	(0.474)	(0.493)	(0.499)
	Sched. caste/tribe	0.338	0.403	0.442	0.154	0.224	0.297	0.085	0.113	0.193
		(0.473)	(0.491)	(0.497)	(0.361)	(0.417)	(0.457)	(0.280)	(0.316)	(0.395)
	Number of quarters	54,446	151,285	56,756	18,852	49,108	23,209	14,287	43,339	23,586
	Number of women	6,832	18,436	7,637	2,201	5,564	2,768	1,347	3,933	2,459

Note. Means without parentheses and standard deviation in parentheses. Interactions between variables, baseline hazard dummies and squares not shown. Quarters refer to number of 3 month periods observed.