GCOE Discussion Paper Series

Global COE Program

Human Behavior and Socioeconomic Dynamics

Discussion Paper No. 277

One male offspring preference: evidence from Vietnam using a split-population model

Tien Manh Vu

August 2012

GCOE Secretariat
Graduate School of Economics
OSAKA UNIVERSITY
1-7 Machikaneyama, Toyonaka, Osaka, 560-0043, Japan

One male offspring preference: evidence from Vietnam using a split-population model

Tien Manh Vu

Osaka School of International Public Policy, Osaka University

1-31 Machikaneyama, Toyonaka, Osaka 560-0043, Japan.

Tel.: +81-(0)90-1717-9773

Fax: +81-(0)6-6850-5656

E-mail: t-vu@osipp.osaka-u.ac.jp

Abstract This paper examines sex preferences for children in Vietnam using the birth cohort from 1990 to 2008. We specify the sex–birth order composition of existing children using a split-population model. The model better fits the data by relaxing the assumption in conventional hazard models that all mothers would eventually have another child. Our results indicate a strong preference for families having precisely one male offspring. In addition, we observe a mixed sex preference at the fourth birth order and find that the involvement of mothers in economic activities does not significantly affect the fertility timing decision until the fourth childbirth. We also find that a 1 percent increase in household income adjusted by household size reduces the probability of having children by 35.3–51.9 percent. The same income increment lowers the level of son preference by 49.8 percent at the second birth order. However, income has only a minor impact on the son preference at higher birth orders. Together, these results suggest that the importance placed on economic reasons for a son preference will gradually weaken as economic development proceeds in Vietnam.

JEL Classification J11, J12, J13

Key words *Son preference, Sex composition, Birth spacing, Birth order, Split-population model*

1 Introduction

During the period 1990–2008 in Vietnam, the combination of improvements in health, a lower fertility rate, and a higher economic growth rate should have weakened the economic foundations of the family preference for a son. In evidence, Vietnamese gross domestic product (GDP) grew at approximately 5 percent annually during this period (GSO 2010). At the same time, the mortality rate of children under 5 years of age in Vietnam fell from 51 per 1,000 live births in 1990 to just 29 in 2005 (WHO 2011). Similarly, the annual population growth rate declined from 1.8 percent in 1990 to just 1.1 percent in 2008, and average life expectancy increased from 65 years in 1990 to 74 years in 2008 (World Bank 2011b, 2011c). Interestingly, however, the participation rate of women in the Vietnamese labor force remained at about 70 percent throughout the same period (World Bank 2011a). These facts encourage further investigation of the case of Vietnam in terms of families' preference for a son.

The purpose of this paper is to examine sex preferences for children in the 1990–2008 birth cohort using data from the 2008 Vietnam Household Living Standard Survey (VHLSS). Unlike previous research in this area, we examine the fertility timing decision up to the fourth birth order conditional on the sex–birth order composition of all preceding offspring using a split-population model. We also analyze the impact of household income on the level of son preference. We find that the split-population model is superior to conventional hazard models when estimating the probability of having another child in each parity progression. This is because it not only fits the data better but also relaxes the assumption of conventional hazard models that all mothers will eventually have another child.

We obtain four major findings as follows. First, during the period 1990–2008, parents in Vietnam ultimately desired precisely one male offspring. For this reason, mothers without a son have the highest probability of having an additional child every month after ninth months since the delivery of the previous child. Second, we also find a mixed sex preference in parity progression to the fourth birth. Third, whether mothers are involved in economic activities has minimal influence on fertility timing up to the third childbirth. Fourth, a 1 percent increase in household income reduces the probability of having additional children by up to 51.9 percent and the level of son preference by up to 49.8 percent during the parity progression to the second childbirth. However, household income has only a minor impact on the level of son preference at all other birth orders. Together, these results suggest that the importance placed on economic reasons for a son preference will gradually weaken as economic development proceeds in Vietnam.

The remainder of the paper is organized as follows. Section 2 details the existing evidence on sex preference in families, especially in Vietnam. Section 3 introduces the data used and Section 4 describes our empirical methodology. Section 5 presents our findings. Section 6 discusses our results and provides our conclusions.

2 Sex preference literature

Substantial research into sex preference—the tendency for families to favor a specific sex for an unborn child—has taken place in both developing and developed countries. For instance, in many developing countries in South Asia, Southeast Asia, and North Africa, there is evidence of a strong son preference (Basu and De Jong 2010). A preference for a mixed sex family also exists. For example, in the Nordic countries, Danish, Norwegian, and Swedish parents have a daughter preference, whereas Finns exhibit a son preference for their third-born child (Andersson et al. 2006). Preferences also appear to change over time. For example, Japanese parents during the period 1920–1939 exhibited a son preference, but this disappeared in subsequent decades, replaced by a mixed sex preference among parents with two children (Kureishi and Wakabayashi 2011). In contrast, although South Korea displayed a strong decline in the fertility rate from 6.0 in 1960 to 1.6 in 1990, the son preference remained strong (Larsen et al. 1998).

Das Gupta et al. (2003) suggest that cultural and economic factors are the two main influences on the son preference. Cultural factors include a patrilineal kinship system, commonly strong in northern India, China, and South Korea, which can increase a family's desire for a son. Moreover, in both China and Korea, ancestor worship—the belief that one should take care of the souls of one's ancestors after their death—motivates the apparent need to have a son. Furthermore, in India and China, possible economic reasons for preferring male offspring include a need to increase the workforce for family farming and to support parents in their old age. High dowry costs in India can also explain a lower preference for female children. Finally, Das Gupta et al. (2003) and Chung and Das Gupta (2007) claim that urbanization, the increased participation of women in the labor force, and social and legal reform movements potentially lead to a decline in the level of son preference.

The consequences of a son preference, although not our focus, include the phenomenon of 'missing women' (Sen 1990) and/or a skewed sex ratio at birth. Parents then engage in prenatal or postnatal discrimination to select the sex they want, accordingly neglecting any children of the other sex (Das Gupta 1987). However, it is not easy to estimate the effect of each factor including sex-selective abortions, female infanticide (as in the works of Sen 1990 and Das Gupta 1987), and Hepatitis B infection of women (as in studies by Oster 2005 and 2010, and Das Gupta 2005, 2006 and 2008) to the outcome of skewed sex ratio at birth. Additionally, Edlund (1999) argues that skewed sex ratios return to normal through the marriage market. Moreover, Chung and Das Gupta (2007) suggest that unbalanced sex ratios in South Korea would transit toward the global average because of the widespread development of nonfarm employment and improvements in gender perceptions. Das Gupta et al. (2009) also provide evidence that the number of 'missing girls' in China and India has begun to decline.

Investigation of the evidence attached to sex preference is common in existing empirical studies, where researchers typically search for behavioral evidence, rather than merely analyzing the outcome of a specific questionnaire on individual preferences. For example, Yamaguchi (1989) calculates the average number and birth order of boys in a family under both population homogeneity and heterogeneity in the probability of having a boy. The sex difference in birth order is only found in population heterogeneity while the sex difference in the number of siblings exits in both types of populations. Haughton and Haughton (1998) construct eight different tests under limited information circumstances for families that have completed childbearing. Notably, this line of research suggests specifying the time interval between a given birth order and the next in a hazard model. Furthermore, Basu and De Jong (2010) identify a common son-targeting fertility behavior in developing countries, with the outcome being that girls generally have more siblings and more often appear in lower birth orders.

In Vietnam, the son preference has been examined by Haughton and

Haughton (1995, 1996, 1998) and Bélanger (2002b). Using quantitative data on 500 families in a rural village in northern Vietnam and qualitative information on 100 individuals, Bélanger (2002b) suggested that parents prefer to have sons for their social, symbolic, and economic value. In earlier work, Haughton and Haughton (1995) employ a subsurvey in the 1992–1993 wave of the Vietnam Living Standard Survey (VLSS) of 2,636 women aged 15–49 years to argue that although the son preference in Vietnam remains relatively strong, it has only a minor impact on the fertility rate. By investigating the birth spacing between families' second and third births, Haughton and Haughton (1996) apply two separate Weibull regression models under the hypothesis of heterogeneous preferences. Their results indicate that half of the surveyed parents preferred sons, but that the preference was difficult to confirm for the remainder. Most recently, from eight different types of tests, Haughton and Haughton (1998) propose that a simple progression parity model and table are very effective for evaluating the preference for sons if information on sex by birth order is available.

Although these existing studies provide a useful foundation for the analysis of sex preference in Vietnam, they have some key limitations. To start with, most of the data in use employ a certain restrictive sampling method in that counting the aggregate number of boys and girls in specific types of families in a certain birth order in the sample does not guarantee coverage of all necessary conditions for a son preference in other sample data sets as well as in the population. In addition, it is an unrealistic assumption that parents decide both the total number of children and the number of boy(s) before any childbirths have taken place. Parents cannot perfectly control the sex of the child so their decisions will change conditional on the sex of any existing children. Moreover, there would be bias in our estimations if we were to specify overlapping sex composition as an independent variable. For example, Lee (2008) specifies a dummy variable for the children in a two-child family of different sexes. If having a first male child were preferred more than a first female child (a type of sex–birth order preference), the interpretation obtained from the estimates of this single dummy variable would be less persuasive.

In addition, most studies rely on a combination of standard hazard models for women of a fertile age (15–49 years) and logistic models for older women assumed to have completed their period of fertility. This conventional method makes it difficult to interpret results if there is a difference in the birth cohorts of the mothers, the distributional assumptions, or both. Moreover, demographers are generally unlikely to consider interactions between the son preference and household income because collecting information on household income substantially increases the survey cost, especially given that demographic surveys commonly have large sample sizes. Among the few studies that consider this, Gaudin (2011) employs the household asset score as a proxy for wealth and finds a strong negative correlation between wealth and son preference in India. However, as the assets owned by all household members are included in the asset score, it does not adjust for the family size effect.

3 Data

The data we employ are from the 2008 Vietnam Household Living Standard and Consumer Price Index Survey, commonly known as the 2008 VHLSS. The 2008 VHLSS is similar to earlier waves of the survey; it receives technical support from the World Bank and is undertaken nationwide by the General Statistics

Office of Vietnam (GSO). The data set in use is the country sample referred to as Sample 1, which consists of 45,945 Vietnamese households. Sample 1 includes a sampling weight for every household. In these households, our analysis investigates the fertility of parents, one of whom is the household head. Children included in the survey are then the children of the household head, and we obtain this information from a specific question in the survey questionnaire.

We select the data using several specific criteria to prevent problems arising with our sample observations. First, we identify both the son and his spouse as 'children' of the household head in the survey. Second, in Vietnam, a woman does not change her family name when married and she may reside in her spouse's family home. Thus, we select parents for our sample whose children are all single according to their declared marital status. All children must also have the same initial letter of their family name as their father and we restrict the age of the eldest co-residing children to be less than 19[1] years to obtain the birth cohort of interest. Third, we omit families where the age of the mother at childbirth is less than 15 years or more than 49 years to reduce the probability of including stepmothers in the analysis. However, the remaining likelihood of errors from erroneously including stepparents is relatively low. Just 0.24 percent of the fathers and less than 0.01 percent of the mothers in the Sample 1 are divorced.

In addition, we include additional data containing information on all family members working and living away from the household 6–12 months before the commencement of the survey[2]. However, we do not have information on children that may have left the household more than a year before the survey. Finally, we limit mothers to those aged 49 years or less. In total, 14,406 households meet all of our sample selection criteria. Table 1 provides descriptive statistics for the sample households. As shown, the mean number of children per couple is 2.1, which is close to the total fertility rate for Vietnam reported in 2008 by the United Nations Population Fund (UNFPA) (2009a).

[INSERT TABLE 1 HERE]

As shown in Table 2, we perform a mean comparison test of the hypothesis that our selection differs from households having at least one child given geographic distribution and income. We select the proportion of households living in urban areas and the father's wage in the last 12 months prior the survey as the criteria for the comparison. The test results indicate that the differences are small.

[INSERT TABLE 2 HERE]

We recognize two problems with our sample selection. First, selecting only households with an eldest co-residing child younger than 19 years could bias our results. However, our interest lies in sex preferences during the economic transition period 1990–2008, during which time some demographic characteristics of the population changed significantly. Second, some older children may have left the household more than a year before the sampling, which could result in

[1] We define age as at the time of the survey. Households in Sample 1 were surveyed at two different times: May/June 1989 and September 2008. Therefore, the actual childbirth cohort in the data ranges from July 1989 to September 2008, which we refer to in this paper as the 1990–2008 birth cohort. The legal marriage age for women in Vietnam has been 18 years since 1987.

[2] Children who are learning far away or still dependent on their parents are considered by the 2008 VHLSS as household members and information about them is included.

error in defining the birth order[3]. Nevertheless, we preferred this option to the omission of some very important explanatory variables, such as sex composition by birth order, household income adjusted by the square root of household size, and birth spacing in months. These records would be difficult to obtain from other demographic surveys.

4 Empirical models and specifications

In the parity progression to the second, third, and fourth childbirths, we examine the sex–birth order composition of all previous children as a key independent variable affecting the outcome of having a next child. We consider both the probability of having additional children and the timing of fertility, that is, the probability (hazard rate) of having a next child in any month after a given birth sequence.

We include several assumptions on sex preference in our analysis. To start with, there are many possible measures aimed at prenatal and postnatal sex selection, including sperm selection, abortion, infanticide, and neglect, often appearing as an excessive mortality rate of the less preferred sex in early life (Das Gupta 1987). However, we assert that having more children or shortening the spacing of births are among the most common methods for a wide range of people to obtain a child of the preferred sex. Thus, the interaction between an independent variable and the less preferred sex is also an indicator of the level of sex preference across that independent variable.

4.1 Empirical models

4.1.1 Standard hazard model for fertility timing

The study period is the time interval for parity progression. For each parity progression, the period of study starts from the ninth month after the mother delivered the preceding offspring. The study period ends either when the mother delivers the next child or in 2008 if otherwise. In the standard model, if the mother does not have a next child until the survey ends, she is included as a censored observation because she could have had more children after 2008.

Suppose that the time interval for having an additional child follows $F(t)$. The survival function, indicating the probability that a woman does not have more children at time t, is then $S(t) = 1 - F(t) = P(T > t)$. $S(t) = 1$ at $t = 0$.

The hazard function[4], $h(t)$, is the instantaneous rate of failure or the rate of having a next child in a small period of time. It is then the probability of having a next child at a certain point in time:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t+\Delta t > T > t | T > t)}{\Delta t} = \frac{f(t)}{S(t)}. \quad (1)$$

The cumulative hazard function is

$$H(t) = \int_0^t h(u)du = \int_0^t \frac{f(u)}{S(u)} = -\ln(S(t)). \quad (2)$$

However, unlike the conventional literature on sex preference analysis, we do not apply standard hazard models to our empirical work, but instead use a split-population model.

---

[3] We check the seriousness of this problem by testing the sex ratio of children aged 6–18 to be equal to unity in the Appendix.

[4] We simplify all functional forms without displaying covariates. For example, $h(t)$ is used instead of $h(t; X)$, where X denotes independent variables.

4.1.2 Split-population model

The standard hazard model for fertility timing implies that all mothers that do not progress to the next birth within the study period will eventually have another child sometime after the study period ends. However, this assumption is unrealistic in that some part of the censored data set will never have additional children during their lifetime. The literature typically refers to these cases as 'cured' or 'immune' (Maller and Zhou 1996) compared with 'die' or 'failure' events. In economics, models dealing with 'immune' cases are referred to as split-population models. In addition, split-population models allow the independent variables to influence both the probability of failure and the time of failure for those that will ultimately arrive at a failure event (Schmidt and Witte 1989).

Lambert (2007) assumes that the survival function for all causes $S(t)$ is a product of the expected survival function, $S^*(t)$, and the relative survival function, $R(t)$:

$$S(t) = S^*(t).R(t). \tag{3}$$

The corresponding hazard function of all causes $h(t)$ becomes

$$h(t) = h^*(t) + \lambda(t) \tag{4}$$

where $\lambda(t)$ is the excess hazard rate dealing with the event of interest. $\lambda(t)$ has an asymptote at zero.

There are two principal ways of estimating the 'immune' fraction. The first approach is to obtain the expected survival rate, $S^*(t)$, and/or the expected hazard rate, $h^*(t)$, from national statistics as estimated from other data sources. However, our empirical model could not proceed in this way because of difficulties in obtaining such information in Vietnam, even data as simple as the expected probability of having a next child at a given age for the mother. The alternative is to fit the expected value using standard cure models (Lambert 2007). For the reasons given, we adopt the second approach.

The literature reports two types of model for the second measure, namely, mixture and nonmixture cure models. The mixture cure fraction model assumes that the relative survival rate is

$$R(t) = \pi + (1-\pi).S_b(t) \tag{5}$$

and

$$S(t) = S^*(t).R(t) = S^*(t).[\pi + (1-\pi).S_b(t)] \tag{6}$$

where π is the fraction of women that will never have a next child ('immune') and $S_b(t)$ is the survival function for those who eventually have another child ('failure'). The corresponding excess hazard rate becomes

$$\lambda(t) = \frac{(1-\pi).f_b(t)}{\pi+(1-\pi).S_b(t)} \tag{7}$$

and

$$h(t) = h^*(t) + \lambda(t) = h^*(t) + \frac{(1-\pi).f_b(t)}{\pi+(1-\pi).S_b(t)}. \tag{8}$$

In the nonmixture cure model, the relative survival rate is assumed to be

$$R(t) = \pi^{F_z(t)} = \exp\big(\ln(\pi).F_z(t)\big) = \pi + (1-\pi)\left(\frac{\pi^{F_z(t)}-\pi}{1-\pi}\right) \tag{9}$$

and

$$S(t) = S^*(t).R(t) = S^*(t).\{\pi + (1-\pi)\left(\frac{\pi^{F_z(t)}-\pi}{1-\pi}\right)\} \tag{10}$$

where $F_z(t)$ is the cumulative distribution function of the progression to the next birth. We select $F_z(t) = 1 - S_z(t)$ and $S_z(t)$ as the standard parametric survival function. The excess hazard rate can then be expressed as

$$\lambda(t) = -\ln(\pi)\,.\,Pr_z(t) \tag{11}$$

and

$$h(t) = h^*(t) + \lambda(t) = h^*(t) - \ln(\pi) \,.\, \mathrm{Pr}_z(t) \tag{12}$$
where $Pr_z(t)$ is the probability distribution function for $F_z(t)$.

The general likelihood function for the two models for the i^{th} household is

$$L_i = (h(t_i))^{d_i} . \frac{S(t_i)}{S(t_{0i})} \tag{13}$$
where d_i is a censoring indicator, indicating $d_i = 0$ if censored and $d_i = 1$ if not. For a sample of *n* households with sampling weights, $wt45$,

$$\mathcal{L} = \prod_{i=1}^{n} (L_i * wt45_i^{-1}). \tag{14}$$
Thus, the log-likelihood function can be expressed as

$$\ln(\mathcal{L}) = \sum_{i=1}^{n} [\ln(L_i) - \ln(wt45_i)]. \tag{15}$$
Each $\ln(L_i)$ can be computed from

$$ln(L_i) = d_i \,.\, \ln\big(h(t_i)\big) + \ln(S(t_i)) - \ln(S(t_{0i})). \tag{16}$$
In our empirical model, in the ninth month after delivering a child in a given birth order, all mothers enter the study time. Therefore, $\ln(S(t_{0i}))$=0 for all mothers.

In the mixture cure model, we replace (8) and (16) with (15) and have

$$\ln(\mathcal{L}) = \sum_{i=1}^{n} [d_i . \ln\left\{h^*(t_i) + \frac{(1-\pi_i).\mathrm{f}_b(\mathrm{t}_i)}{\pi_i + (1-\pi_i).\mathrm{S}_b(t_i)}\right\} + \ln\{S^*(t_i)\} + \ln\{\pi_i + (1 - \pi_i).\, \mathrm{S}_b(t_i)\} - \ln(wt45_i)]. \tag{17}$$
Similarly, in the nonmixture cure model, (15) becomes

$$\ln(\mathcal{L}) = \sum_{i=1}^{n} [d_i . \ln\{h^*(t_i) - \ln(\pi_i) \,.\, \mathrm{Pr}_z(t_i)\} + \ln\{S^*(t_i)\} + ln\,(\pi_i) - \ln(\pi_i).\, \mathrm{S}_z(t_i) - \ln(wt45_i)]. \tag{18}$$

4.2 Specifications

In general, we set $\mathrm{f}(\pi) = g\big(X_i, W_{ij}, \beta, \alpha\big)$. W_{ij} is a set of dummies indicating the sex–birth order composition of all preceding children and X_i are control variables, including parental demographic characteristics, income, and economic geography. We also add to X_i the product of a dummy indicating 'all children are female' and the natural logarithm of household income adjusted by the square root of household size. We include the product dummy in separate estimations.

More specifically, when $j = 1$, W_{i1} is G when the first child is female and the base is B when the first child is male. When $j = 2$, W_{i2} includes GG when both the first and second children are female, GB (BG) when the first child is female (male) and the second child is male (female), and the base is BB when both the first and second children are male. The sex–birth order composition of the children is similarly specified when $j = 3$.

The control variables, X_i, can be specified as follows.

$agemum1, agemum2, agemum3$ which are corresponding maternal ages (in years) at the time of the first, second, and third childbirths.

$Wworking$ is a dummy variable indicating a working mother. In the 2008 VHLSS, 'working' is defined as either working for a wage/salary or participating in household production or services in planting, animal husbandry, forestry, or aquaculture or undertaking business for the household. This variable could reduce inconsistent estimates in hazard models caused by unobservable heterogeneity (Nickell 1979, Heckman and Singer 1984, Honoré 1990).

$Hedu$ ($Wedu$) is schooling (in years) of the father (mother).

$lhhincomepc$ is the logarithm of household income adjusted by the square root of household size, where $lhhincomepc = \log(\frac{\text{household income}}{\sqrt{\text{household size}}})$. We appreciate that fertility and the family's income level can be endogenous (Rosenzweig and Wolpin 2000). For instance, poor households may choose a larger family size because more children increase the probability for parents to be supported in their old age. Therefore, parents may rush to have additional children, regardless of any son preference. The variable $lhhincomepc$ could help to control for this factor as household size includes all of the children of the household head.

$urban$ is a dummy indicating the type of residential area.

$area1 - area8$ are dummies indicating the eight economic regions in Vietnam. We use these to control for differences in development, ethnicity, and culture.

$Glhhincomepc, GGlhhincomepc$, and $GGGlhhincomepc$ are the corresponding products of G, GG, GGG and $lhhincomepc$.

In the nonparametric analysis, we use only sex composition by birth order, W_{ij}, and the smoothed hazard estimates to compare the hazard difference in a given parity progression. In the parametric models, we include the full set of control variables.

We assume the cure fraction function, $f(\pi)$, can take a variety of forms, including linear, $\pi = \beta' X + \alpha' W$, logistic, $\ln(\frac{\pi}{1-\pi}) = \beta' X + \alpha' W$, and log(-log), $\ln(-\ln(\pi)) = \beta' X + \alpha' W$. For the empirical analysis, we employ the add-on module for Stata provided by Lambert (2007). To our best knowledge, this module provides most combinations of the various assumptions[5] concerning the cure fraction functions and the distributions of the parametric survival functions (including the Weibull, lognormal, and gamma distributions, a mixture of two Weibull distributions, and a mixture of the Weibull and exponential distributions) in both the mixture and nonmixture split-population models.

4.3 Goodness of fit in the parametric models

We assess model fitness using the Akaike Information Criterion (AIC). In general, $AIC = 2k - 2\ln(L)$ where k is the sum of the number of model covariates and the number of model-specific distributional parameters, and L is the maximized value of the log-likelihood function from the estimated model (Akaike 1974). The lowest AIC value indicates the most preferred model.

For each analysis of parity progression (to the second, third, and fourth childbirths), we use the same control variables and sex–birth order composition set and combine these with the cure fraction function and the parametric survival function distributions supported by the module. We estimate the AIC for each combination to choose the best-fitted models.

5 Results

Among the several models estimated, the nonmixture split-population model with a combination of a gamma distribution for the parametric distribution and a logistic form for the cure fraction yields the lowest AIC for all parity progressions[6]. The generalized gamma in use has an accelerated failure time metric[7]. The interpretation of the coefficients for the logistic form of the cure

---

[5] STRSMIX/STRSNMIX command (Lambert 2007) supports linear, logistic, and log(-log) forms of the cure fraction. We apply this command with a standard cure model where the baseline hazard, $h^*(t_i)$, is fitted by creating a variable called *rate* comprising all zeros.

[6] Other combinations with a higher AIC are not reported.

[7] The parameter generalized gamma survivor function is

$$S(t) = \begin{cases} 1 - I(\gamma, u) \text{ if } \kappa > 0 \\ 1 - \Phi(z) \text{ if } \kappa = 0 \\ I(\gamma, u) \text{ if } \kappa < 0 \end{cases}, \text{ where } \mu_i = X_i\beta,\ \gamma = |\kappa|^{-2},\ z = \frac{\text{sign}(\kappa).[\log(t) - \mu]}{\sigma},\ u = \gamma.e^{|\kappa|z},\ \Phi(z)$$

fraction is similar to that for the odds ratio in the logistic model, where $e^b = \frac{\pi}{1-\pi}$ is the ratio of 'cured' observations to 'uncured' observations. If $e^b \geq 1$, the mother has a lower probability of having a next child, while the reverse applies if $e^b < 1$. Thus, we also can apply the tips in Buis (2010, 2012) for interpreting the interaction term. The use of these tips can avoid both misinterpretation of the interaction terms (Ai and Norton 2003) and some of the complications associated with testing hypotheses (Greene 2010) in nonlinear models.

5.1 Preference for a son

5.1.1 Evidence of a son preference

Both the graphical display of smoothed hazard estimates and the split-population model estimates of the probability of having one more child indicate that parents have a son preference in the 1990–2008 birth cohort. As shown in Figures 1–3, the hazard curves for parents having only female children (G, GG, GGG) lie above all other curves of the remaining sex–birth order compositions for almost every month after a given childbirth order. This indicates that women who have only female children are more likely to have a next child than other women. Before rising to the top sometime before the first 50 months, the G, GG, and GGG curves are steeper than the other curves, thereby indicating an accelerating desire to have a next child during this stage.

[INSERT FIGURE 1 HERE]

[INSERT FIGURE 2 HERE]

[INSERT FIGURE 3 HERE]

Similarly, in all of the estimations, the coefficients for parents having only female children are statistically significant, less than unity, and smallest among the other coefficients for sex–birth order composition. Thus, mothers who have only female children exhibit the greatest probability of having a next child in each given childbirth order. For instance, in analysis (1) in Table 3, mothers having a female first child, G, are 52 percent more likely to have a second child than mothers having a male first child, B.

[INSERT TABLE 3 HERE]

[INSERT TABLE 4 HERE]

[INSERT TABLE 5 HERE]

5.1.2 Preference for one male offspring

The results also indicate that parents prefer to have precisely one male offspring until the third birth order. As shown in analysis (1) in Table 4, the coefficients for GB and BG are statistically insignificant. The likelihood ratio test of whether the model without GB and BG variables can be nested in model (1) supports the

---

is the standard normal cumulative distribution function, and $I(.)$ is the incomplete gamma function. κ and σ are ancillary parameters estimated from the data.

null hypothesis. Thus, mothers who have two male children display the same probability of having a third child as mothers with one male child (GB and BG). We can still find evidence of this preference for one male offspring in the fourth birth order with GGG mothers 68.6 percent more likely than BBB mothers to have a fourth child (see analysis 1 in Table 5).

5.2 Mixed sex preference at parity progression to the fourth birth

In addition to the preference for a single male child, we observe a mixed sex preference at the fourth birth order[8]. This evidence is consistent for both the parametric and nonparametric analyses. In Figure 3, the GGG and BBB curves are located above the other curves. The mixed sex preference can reflect a desire to have children of both sexes, either with or without a balancing number. For instance, the presence of either extreme of having the first three children of the same sex, BBB and GGG, makes it more likely that there will be a fourth child, to obtain a family comprising children of both sexes. One motivation for the remaining mothers may be the desire for a balanced number of the two sexes.

The differences in the sex–birth order compositions are also significant. As shown in Table 5, we reject the likelihood ratio test of the null hypothesis (H0: model with only GGG can be nested in the model with the full set of sex–birth order composition). However, the underlying reasons for a different preference in the sex–birth order composition are complicated and we possibly require further qualitative study to provide some insight.

5.3 Household income and son preference

Our analyses indicate that high-income families tend to have fewer children. As shown in Tables 3–5, a 1 percent increase in household income adjusted by household size explains about a 35.3–51.9 percent decline in the probability of having a next child. In particular, high-income families exhibit a lower level of son preference at the second birth order. Nevertheless, the level of son preference shows no difference across household income for the remaining birth orders. As indicated in analysis (2) in Tables 3–5, the interaction term $Glhhincomepc$ is statistically significant whereas $GGlhhincomepc$ and $GGGlhhincomepc$ are not. $Glhhincomepc$ can explain about 49.4 percent of the decrease in the probability of having a second child. If we combine this with the baseline value ($const$) of 0.000000152, the odds ratio changes from 0.000000152 to 0.000000075 (= 0.000000152 × 0.494) for a 1 percent increase in household income adjusted by the square root of household size. Note that both G and $lhhincomepc$ are statistically significant. Thus, the increment in income itself provides a 27.7 percent reduction in the probability of having a second child as well as lowering the level of son preference by 49.4 percent. This result concurs with Gaudin (2011) and suggests that the level of son preference in Vietnam will decline because of ongoing economic development.

5.4 Other influencing factors on birth spacing

Involvement in economic activities by women aged 15–49 years displays a weak impact on the fertility timing decision. In our selected sample, 95.3 percent of mothers are working mothers. As shown in Tables 3 and 4, working ($Wworking$)

---

[8] We acknowledge that it is difficult to distinguish between families having a son preference and others having a mixed sex preference among GGG families.

does not have any significant influence on fertility timing at the second and third birth orders. However, it does at the fourth childbirth. We also find that schooling and the age of the mother have a greater influence on the fertility timing decision than those for fathers. For example, an additional year in school by the mother reduces the probability of having a second, third, and fourth child by 7.8, 16.8, and 14.0 percent, respectively. Similarly, a year older at the time of the previous child delivery of the mother can explain for decreases the probability of a second, third, and fourth child by 30.8, 17.8, and 14.0 percent, respectively. Consequently, older mothers are less likely to engage in childbearing or shorten the birth spacing if they want to have more children. These results are consistent with those in Yamaguchi and Ferguson (1995), Gray et al. (2010), and Basu and De Jong (2010). The age and education of the father exhibit smaller effects. For example, the age of the father has no impact on the decision to have a second child but does explain about 7.6 and 5.6 percent of the respective probability of having a third and fourth child. In all likelihood, this is because men do not have the same strict age-related bounds of biological fertility as women. Lastly, an additional year in school for the father lowers the probability of having a second and third child by 4.8 and 12.2 percent, respectively.

6 Discussion and conclusions

After constructing the full set of sex–birth order composition, this paper examines sex preference, birth spacing, and household income in Vietnam using a split-population model. When we control for the number of children, the full set of sex–birth order composition best specifies the order of sex in the sex composition. At the same time, the split-population model helps both to overcome the unrealistic assumption that all mothers will proceed to have another child and to fit the data better. The models deal with the probability of having a child each month after a given preceding birth order in three separate parity progressions, up to the second, third and fourth childbirths.

The results show a preference for a son until the third childbirth, while we observe an additional preference, mixed sex, at the fourth childbirth. Involvement in economic activities by the mother has minimal impact on fertility timing until the fourth childbirth. The income effect is consistent across the analysis and reduces the probability of having a next child. Higher-income families especially display a lower level of son preference at the second birth order, while there is no difference in son preference across income levels for the remaining birth orders.

Sex-selective abortion in Vietnam is unlikely to affect our findings. In evidence, UN World Abortion Policies (2007, 2011) reports that abortions for women aged 15–44 years in Vietnam decreased from 35.2 per 1,000 females in 2000 to 18.4 in 2007. In addition, the most common reasons given for abortion in Vietnam include a decrease in the preferred family size, the reliance on a single method of contraception (intrauterine device), the poor availability of alternative methods of contraception, the low cost of abortion, and the increasing rate of sexual activity among unmarried women (Goodkind 1994). Furthermore, there is no evidence that sex-selective abortion is widespread (Bélanger 2002a). Reporting from various data sources in Vietnam, UNFPA (2009b) finds that the absolute value of the sex ratio at birth (average number of newborn boys per 100 girls) has been increasing to above 105 since 2000. The report hypothesizes that this increase could be due to sex-selective abortion as a result of the wider availability of sex-determination facilities. However, the report does not provide any evidence, such as the proportion of abortion cases for sex-selective reasons among all causes. In addition, the report does not explain why the sex ratio at birth is below 105 in 2003.

To verify this issue in the selected sample of the 2008 VHLSS, we use a one-sample mean t-test to determine whether the sex ratio, R=B×100/G, equals 105.5 for children under 6 years of age for each birth order. The test results provide weak evidence of an abnormally skewed sex ratio (see the Appendix). There are some cases where R is above 105.5, such as when the first child is aged 3 and 5 years, the second child is aged 2, 3, or 4 years, the third child aged is aged 0 years, and the fourth child is aged 4 years. The variation by age and birth order weakens the existence of an unusually skewed sex ratio in the selected sample. If parents did pursue prenatal sex selection, the actual level of son preference in Vietnam would be even higher than that indicated in our study.

Despite these previously discussed economic changes and improvements in health indicators, the preference for having precisely one male offspring remains in Vietnam. Our result is consistent with that of Larsen et al. (1998), in which South Korea exhibited a son preference despite a low fertility rate. Almost certainly, when the total fertility rate falls to approximately two (UNFPA 2009a), especially when the mortality rate in early life is also low (WHO 2011), parents would not need to have more than one male offspring as a backup. It is also not necessary to have two or more sons for the purposes of family lineage and old age support. Moreover, the 'price' of having children is increasing alongside economic development and urbanization. Therefore, it would be reasonable to have a preference for precisely one male child for both economic and cultural reasons.

In contrast, the high participation rate of Vietnamese mothers in economic activities has a minimal impact on fertility timing until the third childbirth. This lies contrary to our usual expectations. This is likely because, up to 2008, the relative cost of having more children to being absent from the workplace is sufficiently low until the third childbirth. However, we expect this factor to contribute more to making family size smaller when the 'price' of a child and/or nonfarm work transition accelerates in the future.

Finally, our findings concerning a decline in the level of son preference for high-income families suggest that past economic reasons for a son preference, such as the need for laborers for farming activities and support in old age, may be losing substance with the ongoing economic development in Vietnam. In contrast, cultural reasons for son preference, such as the lineage of the family name and ancestor worship, remain despite these economic changes.

Acknowledgments The authors acknowledge the useful suggestions of Hisakazu Matsushige, Tsunehiro Ootsuki, and Charles Yuji Horioka of Osaka University, and participants at the Labor Economics Conference held in Awajishima, Japan, September 4–6, 2011 and the Fall Meeting of the Japanese Economic Association held at Tsukuba University, Japan, October 29–30, 2011. The authors would especially like to thank Shoshana Grossbard and two anonymous referees for valuable comments and suggestions made during the reviewing process. All remaining errors are ours.

References

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters, 80*(1), 123–129.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19*, 716–723.

Andersson, G., Hank, K., Rønsen, M., & Vikat, A. (2006). Gendering family composition: Sex preferences for children and childbearing behavior in the Nordic countries. *Demography,*

43(2), 255–267.
Basu, D., & De Jong, R. (2010). Son targeting fertility behavior: Some consequences and determinants. *Demography, 47*(2), 521–536.
Bélanger, D. (2002a). Sex selective abortions: Short-term and long-term perspectives. *Reproductive Health Matters, 10*(19), 194–197.
Bélanger, D. (2002b). Son preference in a rural village in North Vietnam. *Studies in Family Planning, 33*(4), 321–334.
Buis, M. L. (2010). Stata tip 87: Interpretation of interactions in nonlinear models. *Stata Journal*, *10*(2), 305–308.
Buis, M. L. (2012). Stata tip 107: The baseline is now reported. *Stata Journal*, *12*(1), 165–166.
Chung, W., & Das Gupta, M. (2007). The decline of son preference in South Korea: The roles of development and public policy. *Population and Development Review*, *33*(4), 757–783.
Das Gupta, M. (1987). Selective discrimination against female children in rural Punjab, India. *Population and Development Review*, *13*(1), 77–100.
Das Gupta, M. (2005). Explaining Asia's 'missing women': A new look at the data. *Population and Development Review*, *31*(3), 529–535.
Das Gupta, M. (2006). Cultural versus biological factors in explaining Asia's 'missing women': Response to Oster. *Population and Development Review, 32*(2), 328–332.
Das Gupta, M. (2008). Can biological factors like Hepatitis B explain the bulk of gender imbalance in China? A review of the evidence. *World Bank Research Observer*, *23*(2), 201–217.
Das Gupta, M., Chung, W., & Shuzhuo, L. (2009). Evidence for an incipient decline in numbers of missing girls in China and India. *Population and Development Review*, *35*(2), 401–416.
Das Gupta, M., Zhenghua, J., Bohua, L., Zhenming, X. Chung, W., & Hwa-Ok, B. (2003). Why is son preference so persistent in East and South Asia? A cross-country study of China, India, and the Republic of Korea. *Journal of Development Studies*, *40*(2), 153–187.
Edlund, L. (1999). Son preference, sex ratios, and marriage patterns. *Journal of Political Economy*, *107*(6), 1275–1304.
Gaudin, S. (2011). Son preference in Indian families: Absolute versus relative wealth effects. *Demography*, *48*(1), 343–370.
General Statistics Office of Vietnam (GSO) (2010). Gross domestic product at constant 1994 prices by economic sector. http://www.gso.gov.vn/default_en.aspx?tabid=468&idmid=3&ItemID=12106. Accessed 9 July 2012.
Goodkind, D. (1994). Abortion in Vietnam: Measurements, puzzles, and concerns. *Studies in Family Planning*, *25*(6), 342–352.
Gray, E., Evans, A., Anderson, J., & Kippen, R. (2010). Using split-population models to examine predictors of the probability and timing of parity progression. *European Journal of Population/Revue européenne de Démographie*, *26*(3), 275–295.
Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters*, *107*(2), 291–296.
Haughton, D., & Haughton, J. (1996). Using a mixture model to detect son preference in Vietnam. *Journal of Biosocial Science, 28*(3), 355–365.
Haughton, J., & Haughton, D. (1995). Son preference in Vietnam. *Studies in Family Planning, 26*(6), 325–337.
Haughton, J., & Haughton, D. (1998). Are simple tests of son preference useful? An evaluation using data from Vietnam. *Journal of Population Economics, 11*(4), 495–516.
Heckman, J., & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, *52*(2), 271–320.
Honoré, B. E. (1990). Simple estimation of a duration model with unobserved heterogeneity. *Econometrica*, *58*(2), 453–473.
Kureishi, W., & Wakabayashi, M. (2011). Son preference in Japan. *Journal of Population Economics, 24*(3), 873–893.
Lambert, P. C. (2007). Modeling of the cure fraction in survival studies. *Stata Journal*, *7*(3), 351–375.
Larsen, U., Chung, W., & Das Gupta, M. (1998). Fertility and son preference in Korea. *Population Studies*, *52*(3), 317–325.
Lee, J. (2008). Sibling size and investment in children's education: An Asian instrument. *Journal of Population Economics, 21*(4), 855–875.
Maller, R., & Zhou, X (1996). *Survival analysis with long-term survivors*. New York: Wiley.
Nickell, S. (1979). Estimating the probability of leaving unemployment. *Econometrica*, *47*(5), 1249–1266.

Oster, E. (2005). Hepatitis B and the Case of the Missing Women. *Journal of Political Economy*, *113*(6), 1163–1216.
Oster, E., Chen, G., Yu, X., & Lin, W. (2010). Hepatitis B does not explain male-biased sex ratios in China. *Economics Letters*, *107*(2), 142–144.
Rosenzweig, M. R., & Wolpin, K. I. (2000). Natural 'natural experiments' in economics. *Journal of Economic Literature*, *38*(4), 827–874.
Schmidt, P., & Witte, A. D. (1989). Predicting criminal recidivism using 'split population' survival time models. *Journal of Econometrics*, *40*(1), 141–159.
Sen, A. (1990). More than 100 million women are missing. *New York Review of Books*. http://www.nybooks.com/articles/archives/1990/dec/20/more-than-100-million-women-are-missing/. Accessed 9 July 2012.
United Nations (UN) (2007). World abortion policies 2007, Department of Economic and Social Affairs. http://www.un.org/esa/population/publications/2007_Abortion_Policies_Chart/2007_WallChart.pdf. Accessed 9 July 2012.
United Nations (UN) (2011). World abortion policies 2011, Department of Economic and Social Affairs. http://www.un.org/esa/population/publications/2011abortion/2011wallchart.pdf. Accessed 9 July 2012.
United Nation Population Fund (UNFPA) (2009a). Vietnam population 2008. http://unfpa.org/webdav/site/vietnam/shared/UNFPA_2008%20Viet%20Nam%20Population_ENG_FINAL.pdf. Accessed 9 July 2012.
United Nation Population Fund (UNFPA) (2009b). Recent change in the sex ratio at birth in Viet Nam. http://www.unfpa.org/webdav/site/global/shared/documents/publications/2009/sex_ratio_birth_report.pdf. Accessed 9 July 2012.
World Bank (2011a). Labor participation rate, female. http://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?display=default. Accessed 9 July 2012.
World Bank (2011b). Life expectancy at birth. http://data.worldbank.org/indicator/SP.DYN.LE00.IN/countries/VN?display=default. Accessed 9 July 2012.
World Bank (2011c). Population growth rate. http://data.worldbank.org/indicator/SP.POP.GROW?page=3. Accessed 9 July 2012.
World Health Organization (WHO) (2011). Global health observatory data: Country summary statistics for Vietnam. http://apps.who.int/ghodata/?vid=21300&theme=country. Accessed 9 July 2012.
Yamaguchi, K. (1989). A formal theory for male-preferring stopping rules of childbearing: Sex differences in birth order and in the number of siblings. *Demography*, *26*(3), 451–465.
Yamaguchi, K., & Ferguson, L. R. (1995). The stopping and spacing of childbirths and their birth-history predictors: Rational-choice theory and event-history analysis. *American Sociological Review*, *60*(2), 272–298.

Appendix: Sex ratio by age and birth order

Age	First birth order		One-sample t-test		
	Total	R=B×100/G	R1=G/(B+G)	Pr(T < t)	Conclusion
0	123	123.64	0.447	0.191	Accept H0
1	202	104.04	0.490	0.539	Accept H0
2	252	104.88	0.488	0.519	Accept H0
3	310	123.02	0.448	0.089	Reject H0
4	343	106.63	0.484	0.461	Accept H0
5	438	122.34	0.450	0.061	Reject H0
6–18	12,738	109.13	0.478	0.000	Reject H0
Total	14,406	109.69	0.477	0.000	Reject H0
	Second birth order				
0	402	114.97	0.465	0.195	Accept H0
1	437	107.11	0.483	0.437	Accept H0
2	482	131.73	0.432	0.008	Reject H0
3	518	118.57	0.458	0.092	Reject H0
4	641	118.77	0.457	0.067	Reject H0
5	663	97.32	0.507	0.850	Accept H0
6–18	8,421	104.54	0.489	0.021	Reject H0
Total	11,564	106.91	0.483	0.000	Reject H0
	Third birth order				
0	205	130.34	0.434	0.066	Reject H0
1	203	125.56	0.443	0.109	Accept H0
2	221	114.56	0.466	0.271	Accept H0
3	247	96.03	0.510	0.769	Accept H0
4	253	96.12	0.510	0.770	Accept H0
5	237	117.43	0.460	0.206	Accept H0
6–18	2,195	102.12	0.495	0.312	Accept H0
Total	3,561	105.6	0.486	0.052	Reject H0
	Fourth birth order				
0	56	115.38	0.464	0.371	Accept H0
1	77	113.89	0.468	0.370	Accept H0
2	72	111.76	0.472	0.404	Accept H0
3	83	97.62	0.506	0.637	Accept H0
4	76	145.16	0.408	0.085	Reject H0
5	66	106.25	0.485	0.489	Accept H0
6–18	464	95.78	0.511	0.679	Accept H0
Total	894	104.11	0.490	0.274	Accept H0

Notes: H0: R1 = 0.4867 (R = 105.5), Ha: mean < 0.4867 if age < 6 years. H0: R1 = 0.5 (R = 100), Ha: mean < 0.5 if age > 5 years.

Figure 1 Smoothed hazard estimates for parity progression to the second birth

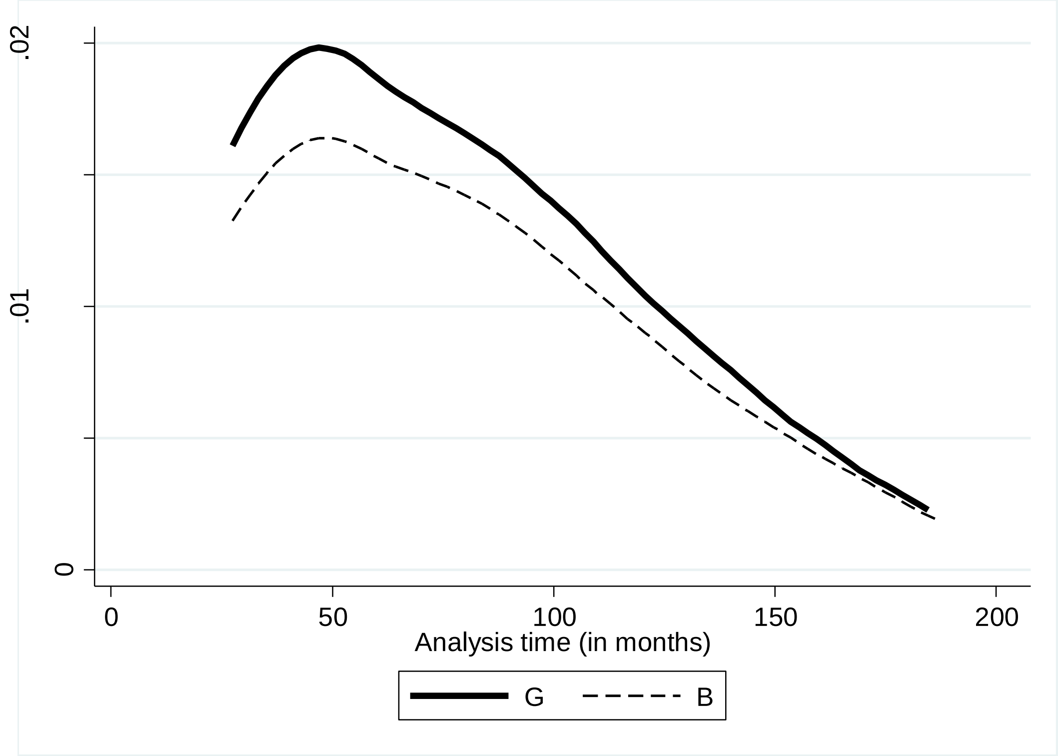

Figure 2 Smoothed hazard estimates for parity progression to the third birth

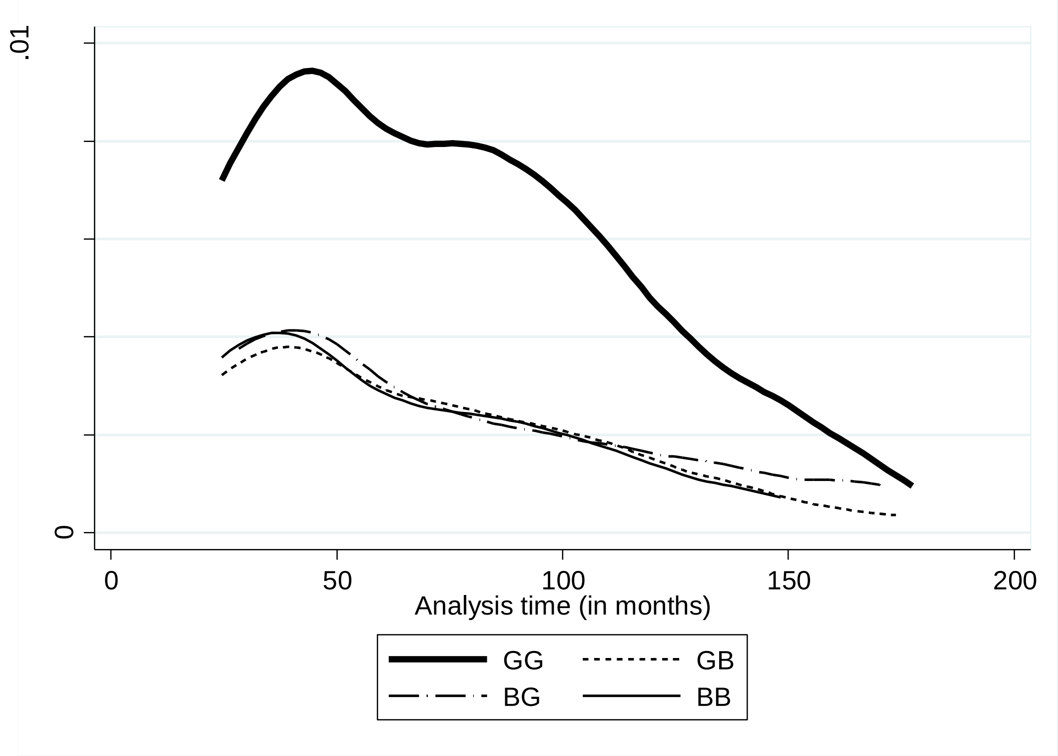

Figure 3 Smoothed hazard estimates for parity progression to the fourth birth

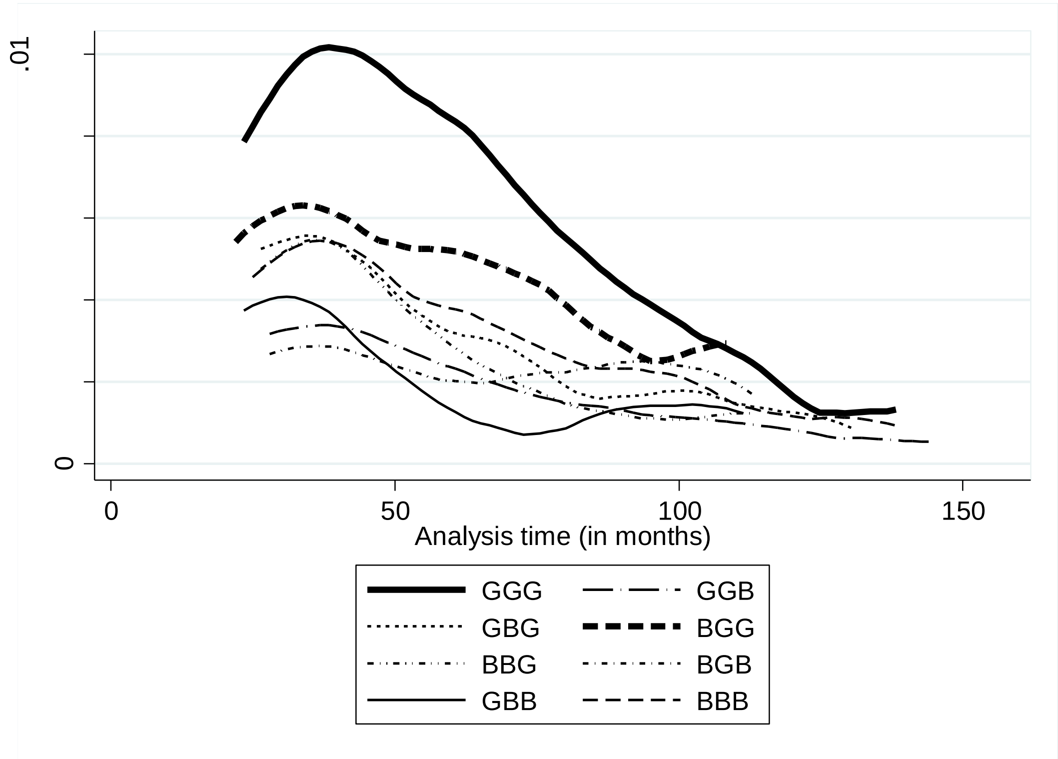

Table 1 Descriptive statistics

Variables	Descriptions	Obs.	Mean	Std. Dev.
nchild	Number of children of the household head	14,406	2.133	0.857
G		14,406	0.477	0.499
B		14,406	0.523	0.499
GG		14,406	0.191	0.393
GB		14,406	0.204	0.403
BG		14,406	0.197	0.398
BB		14,406	0.211	0.408
GGG		14,406	0.045	0.207
BBG		14,406	0.026	0.160
BGG		14,406	0.025	0.157
BGB		14,406	0.026	0.158
GGB		14,406	0.049	0.217
GBG		14,406	0.024	0.153
GBB		14,406	0.026	0.159
BBB		14,406	0.026	0.159
Wage	Age of the mother (in years)	14,406	35.565	5.959
agemum1	Maternal age at first childbirth	14,406	23.366	4.164
agemum2	Maternal age at second childbirth	11,564	26.822	4.438
agemum3	Maternal age at third childbirth	3,561	29.432	4.301
Wedu	Years of schooling for the mother	14,406	7.630	3.926
Wworking	Mother involved in economic activities (yes/no)	14,406	0.953	0.212
Hage	Age of the father (in years)	14,406	38.475	6.289
Hedu	Years of schooling for the father	14,406	8.186	3.896
lhhincomepc	Logarithm of annual household income per capita adjusted by square root of household size	14,406	9.745	0.711
urban	Residing in urban area (yes/no)	14,406	0.231	0.421
area1	Red River Delta (yes/no)	14,406	0.194	0.395
area2	North East (yes/no)	14,406	0.175	0.380
area3	North West (yes/no)	14,406	0.066	0.247
area4	North Central Coast (yes/no)	14,406	0.103	0.304
area5	South Central Coast (yes/no)	14,406	0.089	0.284
area6	Central Highlands (yes/no)	14,406	0.074	0.262
area7	Southeast (yes/no)	14,406	0.122	0.327
area8	Mekong River Delta (yes/no)	14,406	0.178	0.382

Table 2 Two-group mean comparison test

Criteria	Selected sample (1)		Nonselected sample (2)		H0: diff = 0 diff = mean(2) – mean(1)	
	Obs.	Mean	Obs.	Mean	Pr(|T| > |t|)	Conclusion
Urban	14,406	0.2308	11,285	0.2343	0.5114	Accepted H0
Father's wage	5,530	20,814	2,961	20,871	0.8855	Accepted H0

Table 3 Split-population model for parity progression timing of the second birth

	(1)		(2)	
_t	exp(b)	Std. Err.	exp(b)	Std. Err.
pi				
G	0.480***	0.048	0.009***	0.010
agemum1	1.308***	0.030	1.309***	0.030
Wedu	1.078***	0.017	1.079***	0.017
Wworking	0.990	0.176	0.994	0.177
Hage	0.993	0.008	0.993	0.009
Hedu	1.048***	0.016	1.049***	0.016
lhhincomepc	1.519***	0.122	1.277***	0.115
urban	1.634***	0.177	1.631***	0.178
area2	1.204	0.170	1.211	0.171
area3	0.843	0.186	0.833	0.185
area4	0.293***	0.065	0.291***	0.064
area5	0.601***	0.110	0.601***	0.110
area6	0.195***	0.053	0.196***	0.053
area7	1.215	0.183	1.218	0.184
area8	5.215***	0.890	5.240***	0.901
Glhhincomepc			1.494***	0.172
const	0.000***	0.000	0.000***	0.000
ln_sigma	0.234***	0.057	0.237***	0.057
kappa	–1.764***	0.184	–1.771***	0.185
mu	4.366***	0.040	4.370***	0.041
Number of obs.	14,406		14,406	
Log likelihood	–57,674.4		–57,667.793	
AIC	115,386.7		115,375.6	

Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table 4 Split-population model for parity progression timing of the third birth

	(1)		(2)	
_t	exp(b)	Std. Err.	exp(b)	Std. Err.
pi				
GG	0.151***	0.020	0.100**	0.110
GB	1.036	0.089	1.036	0.090
BG	0.907	0.079	0.907	0.079
agemum2	1.178***	0.013	1.178***	0.014
Wedu	1.168***	0.017	1.169***	0.017
Wworking	1.199	0.206	1.200	0.206
Hage	0.924***	0.007	0.924***	0.007
Hedu	1.122***	0.014	1.122***	0.015
lhhincomepc	1.353***	0.076	1.341***	0.082
urban	1.614***	0.149	1.614***	0.150
area2	1.640***	0.175	1.641***	0.176
area3	1.145	0.171	1.143	0.171
area4	0.279***	0.036	0.278***	0.036
area5	0.459***	0.059	0.458***	0.059
area6	0.210***	0.033	0.208***	0.033
area7	0.481***	0.063	0.480***	0.063
area8	1.590***	0.179	1.589***	0.180
GGlhhincomepc			1.043	0.114
const	0.002***	0.002	0.002***	0.002
ln_sigma	0.097	0.060	0.101*	0.061
kappa	–2.195***	0.248	–2.212***	0.254
mu	4.002***	0.039	4.003	0.040
Number of obs.	11,564		11,564	
Log likelihood	–21,427.2		–21,427.1	
AIC	42,896.3		42,898.1	
Likelihood ratio test (H0: model without GB and BG can be nested in the original model)	LR chi2(2) = 2.54	Prob>chi2 = 0.281		

Robust standard errors in parentheses (*** p < 0.01, ** p < 0.05, * p < 0.1).

Table 5 Split-population model for parity progression timing of the fourth birth

	(1)		(2)	
_t	exp(b)	Std. Err.	exp(b)	Std. Err.
pi				
GGG	0.314***	0.062	0.586	1.016
BBG	1.435*	0.302	1.434*	0.302
BGG	0.693*	0.143	0.694*	0.143
BGB	1.761***	0.380	1.760***	0.379
GGB	1.619**	0.306	1.618**	0.306
GBG	1.217	0.256	1.217	0.256
GBB	2.064***	0.453	2.062***	0.452
agemum3	1.140***	0.017	1.140***	0.017
Wedu	1.153***	0.024	1.152***	0.024
Wworking	1.672*	0.502	1.673*	0.501
Hage	0.944***	0.011	0.944***	0.011
Hedu	1.030	0.019	1.030	0.019
lhhincomepc	1.367***	0.122	1.387***	0.136
urban	1.512**	0.275	0.937**	0.169
area2	1.631**	0.329	1.510**	0.275
area3	0.983	0.230	1.625	0.328
area4	0.529***	0.095	0.981***	0.229
area5	0.833	0.174	0.529	0.094
area6	0.574***	0.114	0.831***	0.174
area7	0.764	0.161	0.573	0.114
area8	2.174***	0.477	0.762***	0.160
GGGlhhincomepc			2.173	0.477
const	0.004***	0.004	0.003***	0.004
ln_sigma	–0.228***	0.088	–0.231***	0.087
kappa	–1.577***	0.364	–1.566***	0.361
mu	3.711***	0.050	3.711***	0.050
Number of obs.	3,561		3,561	
Log likelihood	–5,440.1		–5,440.1	
AIC	10,930.3		10,932.1	
Likelihood ratio test (H0: model with only GGG and other control variables can be nested in the original model)	LR chi2(6) = 38.05	Prob>chi2 = 0.000		

Robust standard errors are in parentheses (*** p < 0.01, ** p < 0.05, * p < 0.1).