OSIPP Discussion Paper: DP-2012-E-001

The lexicographic preference for a son: evidence from household data in Vietnam

February 23, 2012

Tien Manh Vu†

Ph.D Candidate, Osaka School of International Public Policy, Osaka University

[Key words] Son preference, Sex composition, Birth order, Birth spacing

[JEL Codes] J11, J12, J13

[Abstract]

This paper examines son preferences by specifying sex composition by birth order of existing children as key independent variables. The results indicate a lexicographic preference for a son by mothers aged 50 years and older. Mothers without a son are also under substantial pressure to bear more children and shorten their birth spacing. However, once a family includes a son, parents do not consider sex composition over other decisions on family size and fertility timing. It would appear that the preference for a son is relatively stronger for some birth orders in the northern regions of Vietnam but slightly weaker in the Central Highlands and South Central Coast. In addition, while women are important in the Vietnamese labor force, the level of preference for sons does not differ across income at lower birth orders. We also obtain mixed results for son preferences if we include mothers less than 50 years of age in our analysis.

†Contact: Osaka School of International Public Policy, Osaka University, 1-31 Machikaneyama, Toyonaka, Osaka 560-0043 Japan. Tel/Fax: +81-(0)6-6850-5656. E-mail: t-vu@osipp.osaka-u.ac.jp.

[Acknowledgments] The author gratefully acknowledges helpful comments and suggestions by Hisakazu Matsushige and Tsunehiro Ootsuki, 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 author also thanks Professor Hisakazu Matsushige, Osaka School of International Public Policy, Osaka University for permitting to publish this paper.

The lexicographic preference for a son: evidence from household data in Vietnam[1]

Tien Manh Vu

Osaka School of International Public Policy, Osaka University

[1] An earlier version of this paper is Vu (2011) which employs Vietnamese Households Living Standard Survey 2006.

1 Introduction

Recent improvements in health, a low fertility rate, and a high economic growth rate in Vietnam have challenged the economic roots of the preference for sons. During the last twenty years from 1990 to 2008, Vietnamese gross domestic product (GDP) has grown at approximately 5 percent annually (GSO 2010). At the same time, mortality for children under five years of age in Vietnam has fallen from 51 per 1,000 live births in 1990 to just 29 in 2005 (WHO 2011). Similarly, the annual population growth rate has declined from 1.8 percent in 1990 to only 1.1 percent in 2008, and average life expectancy has increased from 65 years in 1990 to 74 years in 2008 (World Bank 2011b, 2001c). Interestingly, however, the participation rate of females in the labor force in Vietnam has remained at about 70 percent throughout this period (World Bank 2011a).

The purpose of this paper is to reexamine the preference for a son in Vietnam using the Vietnamese Households Living Standard Survey (VHLSS) 2008. Unlike previous studies in this area, we examine the decision on having additional children up to the fourth birth order conditional on the sex composition of all previous children. We also compare the level of son preference by geographic region and household income. In the initial analysis, we employ logistic models to estimate the effect of the sex composition on decisions concerning family size. Then, after controlling for the number of children, we apply hazard models to benchmark the corresponding effect on fertility timing. Lastly, we use the difference in difference (DID) method to examine the level of son preference found.

We obtain four major findings as follows. First, all mothers in Vietnam, especially those over 49 years of age, ultimately desire a son. For this reason, mothers without a son tend to have both more children and shorter birth spacing. However, once we locate a boy (B) in the sex composition of the family's children, parents no longer consider sex composition in their decision on family size. Second, the son preference appears be stronger in the North

East (NE), Red River Delta (RRD), North Central Coast (NCC) and North West (NW) regions of Vietnam, but lower in the South Central Coast (SCC) and Central Highlands (CH). Third, while females in Vietnam make a significant contribution to the labor force, we find no difference in the level of son preference across income at lower birth orders. Finally, if we include mothers less than 50 years of age in our analysis, we obtain mixed preferences.

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

2 Gender preference literature

Research into gender preference, the tendency by families to favor a specific gender 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 mixed genders also exists. For example, in the Nordic countries, Danish, Norwegian and Swedish parents have a daughter preference, while Finns exhibit a son preference for their third-born (Andersson et al. 2006). Japanese parents in the 1920-1939 have a lexicographic son preference; however, son preference disappears in the following decades and is replaced with a mixed preference in parents with two children (Kureishi and Wakabayashi 2011).

The investigation of the evidence attached to gender preference is common in existing empirical studies. Here, scholars 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 of boys in a family with a no-son preference and under both population homogeneity and heterogeneity as to the probability of having a boy. In this situation, and given a son preference, the average number of boys in a family should be higher than the estimated number. Alternatively, 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 one birth and the next in a hazard model.

In Vietnam, the idea of son preference has been examined by Haughton and Haughton (1995, 1996, 1998) and Bélanger (2002). Using quantitative data on 500 families in a rural village in Northern Vietnam and qualitative information on 100 individuals, Bélanger (2002) suggested that people desire to have sons for social, symbolic and economic value. In earlier work, Haughton and Haughton (1995) employ a sub-survey in the 1992-93 wave of the Vietnamese Living Standard Survey (VLSS) of 2,636 women aged 15-49 years to argue that while the son preference in Vietnam is relatively strong, it has only a minor impact on the fertility rate. Later, investigating the birth spacing between the family's 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 prefer sons, while for the remainder any gender preference is difficult to confirm. Later still, from eight different types of tests, Haughton and Haughton (1998) propose that the simple progression parity model and table are very effective if information on gender by birth order is available.

Although these existing studies provide a useful foundation for the analysis of gender preference in Vietnam, they have some key limitations. To start with, parity progression based on counting the aggregate number of boys and girls in specific types of families in the sample does not guarantee all necessary conditions for a son preference in other data sets. In addition, the common assumption by which parents set the family-size decision in concert with the decision on the number of boys in the family cannot hold. In fact, parents cannot perfectly control the gender of the baby they are expecting. Moreover, using overlapping sex composition as an independent variable would cause bias in the estimations. For example, Lee (2008) specifies GG (i.e. the first and second children are girls), $difsex$ (the first child has a different gender to the second child) and $first\ girl$ (the first child is a girl) in the same estimation. Finally, interpretation is less accurate if there is a preference for a certain gender by birth order. For example, the gender birth orders of GB (first child is a girl and second child is a boy) and BG (first child is a boy and second child is a girl) may exhibit conflicting effects relative to the same dependent variables.

By using instead the full sex composition of all previous children as key independent variables in this analysis, we establish that a son preference exists if the following conditions hold. First, low-birth-order females have a relatively larger number of siblings than low-birth-order males. This is because under a son preference, parents without a son continue to have children until they get a boy. Subsequently, among parents with the same number of children intending to have one more child, families without the expected number of son(s) are more likely to be in some hurry and so shorten the spacing of births. This shorter birth spacing allows mothers to participate in more childbearing trials if they fail to have the required number of sons.

In addition, we include several other assumptions in our analysis. To start with, there are many measures aimed at prenatal sex selection, including sex-selective abortion and sperm selection. However, in our analysis we assert that either having more children or shortening the spacing of births is the most common method for a wide range of people. We also assume that having more children and altering birth spacing to obtain a preferred gender is representative of the degree of preference for one gender over another. Hence, the interaction between an independent variable and the gender parents prefer least is also an indicator of the level of gender preference across that variable.

3 Empirical models and specifications

This section develops the empirical models we use to test the above two conditions while controlling for the number of children. We first employ random utility models to estimate the probability of having more children. We then apply hazard methods to examine the birth spacing attached to a certain birth order to calculate the probability of having another child every month after a certain birth order. Finally, we use the DID method to investigate the level of son preference by income level and region.

3.1 Random utility models of the probability of having additional children

In the first part of the analysis, we examine the probability of women having more than one, more than two, and more than three children in their entire lifetime. In general, parents i consider having additional $l-j$ children if the utility of having more children U_{il} is higher than the utility U_{ij} of the current j number of children.

$$U_{il} = X_i * \beta_l + W_{ij} * \alpha_l + u_{il} \geq U_{ij} \ (l > j) \tag{1}$$

Neither $\mathrm{U_{il}}$ nor $\mathrm{U_{ij}}$ can be calculated. However, it is possible to partly observe the past decisions of families after mothers complete their fertility at 49 years of age. We specify $RUM_i = 1$ if parents have more than i children and $RUM_i = 0$ if parents stop at i children.

$$RUM_i = \begin{cases} 1 \text{ if } U_{il} \geq U_{ij} \\ 0 \text{ if } U_{il} < U_{ij} \end{cases} \tag{2}$$

W_{ij} is a set of $j-1$ dummies indicating the sex composition by birth order of all previous children and X_i are control variables, including parental demographic characteristics, income and economic geography. 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} is GG when both the first and second child 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 child are male. The sex composition of the children is similarly specified when $j = 3$. In terms of the control variables, X_i includes the age (in years) of the mother at the j^{th} birth, schooling (in years) of the father and the mother, the logarithm of household income per capita ($lnhhincomepc$) and a dummy variable indicating a rural area ($urban = 1$) if the household resides in an urban area), along with seven dummies identifying the eight main economic regions in Vietnam as given in the data (see Table 1).

3.2 Right-censored hazard models for fertility timing

In the second part of the analysis, we investigate the time interval or birth spacing in months between the first and second births, the second and third births, and the third and fourth births in three separate estimations.

The sample period is the time interval in which we expect the mother to have one child. The period of our study starts with the birth of the i^{th} child and continues until the mother has the $(i+1)^{th}$ child or until the survey ends in 2008, where i ∈ [1,3]. If the mother does not have the $(i+1)^{th}$ child during the study period and is less than 49 years of age in 2008, we include her as censored data because she could have more children after 2008. Meanwhile, we exclude mothers over 49 years of age by the end of 2008 with a total of i children because they are less likely to have additional children after 2008.

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

The hazard function, $h(t)$, is the instantaneous rate of failure or the rate of having the next child in a small period of time.

$$h(t) = lim_{\Delta t \to 0} \frac{P(t+\Delta t > T > t | T > t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (3).$$

The cumulative hazard function is:

$$H(t) = \int_0^t h(u)du = \int_0^t \frac{f(u)}{S(u)} = -ln(S(t)) \quad (4).$$

We assume the hazard function follows the Weibull hazard distribution $\mathrm{h(t) = p * t^{p-1}}$, in which p is the shape parameter estimated from the data. $S(t) = exp(-t^p)$, $F(t) = 1 - exp(-t^p)$ and $f(t) = p * t^{p-1} * exp(-t^p)$. Overall, the Weibull distribution deals better with monotone hazard rates that either increase or decrease exponentially over time.

Generally, we set $h(t_{i,i+1}) = g(t, X_k, W_{ik}, \beta, \alpha)$. Similar to the specification for the random utility models, X_k are control variables, parental demographic characteristics and income. W_{ik} is a set of dummies indicating the sex composition of children in the family until the i^{th} birth of mother k at time t_0. β and α are the coefficients corresponding to X and W.

3.3 DID for level of son preference

In the third and final part of our analysis, we assert that it would be interesting to identify whether the degree of son preference differs significantly across each control variable. Suppose $\mathrm{W^1}$ is the sex composition of previous children that is least preferred and X_i is a control variable that potentially affects both the probability of having more children and the birth spacing, such that $\mathrm{W^1 = 1}$ if the parents have such a sex composition, otherwise $\mathrm{W^1 = 0}$. By using $\beta_1 * \mathrm{W^1} + \delta * X_i + \gamma * \mathrm{W^1} * X_i$, we can estimate DID, γ. If significant, the interaction term γ suggests that families of the least preferred sex composition behave substantially differently in terms of the control variable, holding all other factors unchanged.

In our analysis, we use the eight economic regions of Vietnam and the logarithm of household income per capita in the set of control variables X_i to test whether the level of gender preference changes according to the X_i factor.

4 Data

The data we employ in our analysis are from the Vietnamese Household Living Standard and Consumer Price Index Survey 2008, commonly known as VHLSS 2008. The 2008 round of the VHLSS is generally similar to earlier waves of the survey, technically supported by the World Bank and undertaken nationwide by the General Statistics Office of Vietnam (GSO). The data set is a representative country sample of 45,945 Vietnamese households. We use the survey to investigate the fertility of parents, one of whom is the household head, in these households. Children included in the survey are children of the household head, and we obtain this information from a specific question in the survey questionnaire.

To guarantee that the relationship between the child and the household head is truly that of a blood relative, we select the data using several specific criteria. Specifically, we only select parents for our sample whose children are all single according to their declared marital status, and if the household children have siblings, all siblings must have the same family name. We employ these criteria to prevent the problems arising with our sample observations. First, the son and his spouse are both identified as ‘children’ of the household head in the survey. Second, in Vietnam, a woman may not change her family name if married and she may continue to live in her spouse's house even if she is widowed. Third, we omit families where the age of the mother at childbirth is less than 15 or over 49 years to exclude stepmothers from the analysis. However, the likelihood of errors from including stepmothers and stepfathers is relatively low, as only 0.24 percent of fathers and less than 0.01 percent of mothers in the sample are divorced. In fact, 98.65 percent of fathers in the sample are married and living with their spouses. The corresponding value is 91.81 percent for mothers, of which another 6.09 percent are widows. Altogether, 29,582 households meet these selection criteria.

‘Insert Table 1’

We choose 49 years of age as the milestone for a woman to complete her period of fertility. Initially, the random utility models we use to estimate the probability of having more children require that all mothers complete their fertilities. Some 7,063 households with 12,533 children satisfy this criterion. In the remaining data, 98.9 percent of the eldest children of the mother are less than 50 years of age, which is far below the life expectancy of the typical Vietnamese. Therefore, we could consider 12,533 children as being the surviving children of mothers over 49 years of age. In the subsequent analysis of fertility timing, we examine the time interval between the m^{th} birth and the $(m+1)^{th}$ birth of the mother, where m is an integer from 1 to 3. We exclude all mothers with $m-1$ children for each analysis. In addition, we omit mothers that are over 49 years old with m children as they would not be able to have more children. However, we retain mothers under 50 years of age with m children as censored data.

5 Results

5.1 Lexicographic preference for sons and mixed preferences

The analysis indicates a lexicographic preference for a son of mothers over 49 years of age but an additional mixed preference if we include those mothers less than 50 years of age[2]. Compared with Haughton and Haughton (1995, 1998), in which there is a mixed preference of BBB women alongside a son preference, our estimates include more detail and cover completed cases as follows.

---

2 Additional evidence can be observed from Appendix Figure 1, 2 and 3 using Kaplan-Meier survival estimates for birth spacing. As the curves, denoted with only *G*, are closest to the origin and have the same probability of having the next baby, mothers with only daughters display the shortest birth spacing.

‘Insert Table 2’

The results demonstrate a preference for one son by women that have finished their period of reproduction. As shown in analyses (1), (2) and (3) in Table 2, the sex compositions that do not include any B (G, GG and GGG) are statistically significant and have the greatest impact on the probability of having additional children than any other sex composition. Using Ward tests, we reject the null hypotheses that all sex compositions without B are zero for the cases in columns (2) and (3) in Table 3. Therefore, either having one (GB, BG) or two (BB as the base) sons does not affect the probability of having a third child. Similarly, having three sons (BBB as the base), two sons (BBG, BGB, GBB), or one son (GGB, GBG, BGG) does not influence the probability of having a fourth child. In other words, there is a one-son preference. That is to say, parents look for B in the sex composition of all expected children. Once found, they no longer consider sex composition in their decision on family size.

‘Insert Table 3’

Interestingly, there would appear to be a mixed preference for gender when we include mothers less than 50 years of age. Parents without B in the sex composition of their children have the greatest probability of shortening their birth spacing, as shown in analyses (4), (5) and (6) in Table 2. However, the preference for only one son does not hold, as the Ward tests in (5) and (6) in Table 3 reject the null hypothesis. If we rank the probability of having more children given a particular sex composition, we can distinguish three groups.

$\text{Group 1 (including B } and \text{ G , ending with B)} <$

$Group\ 2\ [(\text{including B and G, ending with G}) \cup (\text{without letter G})] = 0 <$

$Group\ 3\ (without\ B)$.

As shown in Table 2, all statistically significant sex compositions including B end with B, and this has a negative effect on the size decision and birth spacing compared with the base. The sex compositions including B and ending with G are statistically insignificant and equal to zero in a joint Ward test ($H0: GBG = BGG = BBG, Chi2(2) = 1.99, Prob > chi2 = 0.377$). Of these, Group 3 has the greatest probability of having another child, which we indicate as a strong son preference. In half of Group 2, those without a G, there is a daughter preference. However, in the remaining half of Group 2, it is difficult to reach a conclusion, as they could have either a marginal son preference or no gender preference. Group 1 displays a son and/or marginal son preference.

5.2 Level of son preference by region and household income per capita

The level of son preference by region or DID appears only in the hazard models. As indicated in Table 5, the Ward tests that all DID coefficients jointly equal zero is rejected in (4), (5) and (6). As shown in Table 4, the level of son preference at some specific birth orders differs across NE, NW, RRD, NCC, SCC and CH. In addition, NE and RRD exhibit the highest level of son preference, whereas SCC and CH display the lowest level of son preference.

'Insert Table 4 and 5'

As shown in Table 6, the level of son preference by income if ever it is statistically significant simply implies the rich can afford a larger family size. The level of son preference does not differ by income when the birth order is less than three, as the interaction terms in (1), (2) and (4) are statistically insignificant. This is important, because 93.64 percent of the women in the sample who are more than 49 years of age have fewer than four children. If a level of son preference by income exists, it should be evident for all birth orders. Therefore, the DID by income at the fourth childbirth in (3) suggests that higher-income families with three daughters can afford a larger family size. Similarly, the interaction term in (5) suggests that higher-income households would be more ready to have their next child.

In addition, if a son preference derives from economic reasons, such as the higher lifetime value of sons, the poor would have a generally higher level of son preference. In addition, parents would have as many sons as possible. In fact, as shown in Table 6, just one of the interaction terms is negative, but even that is statistically insignificant. Therefore, the one-son preference in Vietnam probably arises for reasons other than economic reason and/or long-term scheme, which is difficult to conclude from the cross section data.

‘Insert Table 6’

6 Conclusion and discussion

After controlling for the number of children, this paper constructs two complementary ways of examining gender preferences in Vietnam. First, random utility models examine whether the sex composition of existing children in the family affects the decision on having more children. Second, hazard models attempt to investigate the relation between sex composition and birth spacing.

Using the above analysis, we find a lexicographic preference for one son in Vietnam for mothers that have completed their period of fertility but a mixed preference for sons in the birth spacing analysis. However, both analyses find that mothers without a son would be under significant pressure to have another child as early as possible. We also find that the level of son preference differs significantly in the northern, SCC, and CH regions across some specific birth orders. However, it is insignificant across the level of income at the first and second birth orders.

Sixteen years after Haughton and Haughton (1995) (using VLSS 1992), the last study in this area in Vietnam, evidence of a son preference remains. In all likelihood, a stable economic growth rate, better life expectancy and the lower mortality rate of children under five years old favor a son preference. This is because parents no longer have to bear as many children in order to obtain a surviving child. It would also appear that a strong son preference exists despite the significant role of women in the Vietnamese labor force, with the female participation rate remaining stable at about 70 percent throughout the period 1990-2006. Furthermore, son and mixed preferences are in evidence together in our estimations. These factors, along with the finding that the level of son preference does not differ across the first two birth orders by income, suggest three possible explanations. First, there would be no economic reason supporting the son preference. It is more likely, therefore, that the lineage of the family name, part of Confucian belief, is the cause of a son preference in Vietnam. Second, probably in long-term scheme, parents need a son as a source for old age support because it is more likely to see old age people are living in their son's house than in their daughters' in Vietnam. Third, it could be a combination of both the lineage of the family name and old age support.

Sex-selective abortion in Vietnam is unlikely to affect our findings. In evidence, UN World Abortion Policies (2007, 2011) indicates 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, in Vietnam, the most common reasons given for abortion include a decrease in the preferred family size, the reliance on a single method of contraception (the intrauterine contraceptive 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). Meanwhile, there is no evidence that sex-selective abortion is widespread (Bélanger, 2002), especially given that 86.31 percent of the children in the sample were born in or before 2001.

Even so, the increased incidence of sex-selective abortion would only serve to lower our coefficient estimates. Therefore, the actual level of son preference in Vietnam would be even higher than that estimated in our study.

References

Andersson G, Hank K, Ronsen 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 (2002) Sex selective abortions: Short-term and long-term perspectives. *Reproductive Health Matters* 10 (19):194-197.

Bélanger D (2002) Son preference in a rural village in North Vietnam. *Studies in Family Planning* 33 (4):321-334.

Goodkind D (1994) Abortion in Vietnam: Measurements, puzzles, and concerns. *Studies in Family Planning* 25 (6):342-352.

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 15 January 2012.

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.

Kureishi W, Wakabayashi M (2011) Son preference in Japan. Journal of Population Economics 24 (3):873-893.

Lee J (2008) Sibling size and investment in children's education: An Asian instrument. *Journal of Population Economics* 21 (4):855-875.

World Bank (2011a) Labor participation rate of female.
http://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?display=default. Accessed 15 January 2012.

World Bank (2011b) Life expectancy at birth. http://data.worldbank.org/indicator/SP.DYN.LE00.IN/countries/VN?display=default.

Accessed 15 January 2012.

World Bank (2011c) Population growth rate. http://data.worldbank.org/indicator/SP.POP.GROW?page=3. Accessed 15 January 2012.

World Health Organization (WHO)(2011) Global Health Observatory Data, Country summary statistics of Vietnam. http://apps.who.int/ghodata/?vid=21300&theme=country. Accessed 15 January 2012.

United Nations (UN)(2007) World Abortion Policy, Department of Economic and Social Affairs. http://www.un.org/esa/population/publications/2007_Abortion_Policies_Chart/2007_WallChart.pdf. Accessed 15 January 2012.

United Nations (UN)(2011) World Abortion Policy, Department of Economic and Social Affairs. http://www.un.org/esa/population/publications/2011abortion/2011wallchart.pdf. Accessed 15 January 2012.

Vu MT (2011) *Lexicographic son preference in Vietnam*. Unpublished Master, Osaka School of International Public Policy, Osaka University.

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.

Table 1 Descriptive Statistics

Variables	Obs.	Mean	Std. Dev.
Number of children of the household head	29,582	2.1	0.986
Age of the mother (years)	28,971	42.566	10.204
Mother more than 49 years of age (yes/no)	29,582	0.239	0.426
Age of the mother when having 1^{st} child (years)	28,971	25.383	5.666
Age of the mother when having 2^{nd} child (years)	21,045	28.044	5.135
Age of the mother when having 3^{rd} child (years)	7,603	30.585	4.891
Years of schooling obtained by the mother	28,971	7.360	4.008
Age of the father (years)	26,800	44.684	10.063
Years of schooling obtained by the father	26,800	8.265	3.942
Annual household income per capita (thousands VND)	29,582	12,162.200	13,766.860
G	29,582	0.439	0.496
GG	29,582	0.156	0.363
GB	29,582	0.177	0.381
BG	29,582	0.184	0.387
BB	29,582	0.208	0.406
GGG	29,582	0.039	0.193
BBG	29,582	0.033	0.180
BGG	29,582	0.028	0.165
BGB	29,582	0.031	0.172
GGB	29,582	0.043	0.202
GBG	29,582	0.025	0.156
GBB	29,582	0.027	0.162
Red River Delta	29,582	0.205	0.404
North East	29,582	0.156	0.363
North West	29,582	0.050	0.217
North Central Coast	29,582	0.117	0.321
South Central Coast	29,582	0.096	0.295
Central Highlands	29,582	0.070	0.255
Southeast	29,582	0.128	0.334
Mekong River Delta	29,582	0.178	0.383

Table 2 Decisions on family size and fertility timing

Variables	Logistic			Hazard		
	RUM_1	RUM_2	RUM_3	$t_{(1,2)}$	$t_{(2,3)}$	$t_{(3,4)}$
	(1)	(2)	(3)	(4)	(5)	(6)
G	0.242***			0.214***		
	(0.0699)			(0.0194)		
GG		0.664***			0.695***	
		(0.147)			(0.0381)	
GB		-0.130			-0.146***	
		(0.133)			(0.0418)	
BG		-0.0467			-0.00977	
		(0.126)			(0.0405)	
GGG			1.051***			0.636***
			(0.322)			(0.0848)
BBG			-0.178			-0.116
			(0.260)			(0.0954)
BGG			-0.410			-0.0228
			(0.293)			(0.0988)
BGB			-0.687**			-0.243**
			(0.288)			(0.101)
GGB			-0.251			-0.340***
			(0.309)			(0.0996)
GBG			-0.0267			0.0300
			(0.312)			(0.102)
GBB			-0.735**			-0.310***
			(0.291)			(0.109)
Observations	5,102	2,620	1,029	23,707	18,080	6,527

The control variables such as mother's age at latest child's birth, urban location, education of the mother, logarithm of household income per capita, education of the father and seven dummies indicating the eight main economy regions are included but estimates not shown. Robust standard errors are in parentheses (*** p < 0.01, ** p < 0.05, * p < 0.1).

Table 3 Ward test for H0: All sex compositions including B are equal to zero

	Logistic		Hazard	
	RUM_2	RUM_3	$t_{(2,3)}$	$t_{(3,4)}$
	(2)	(3)	(5)	(6)
Test value	0.35	6.90	9.94	19.21
$Prob > chi2$	0.5564	0.2284	0.0016	0.0018
Conclusion on H0	Accepted	Accepted	Rejected	Rejected

Original estimations are shown in the corresponding columns in Table 2.

Table 4 Level of son preference by region

	Logistic			Hazard		
Variables	RUM_1	RUM_2	RUM_3	$t_{(1,2)}$	$t_{(2,3)}$	$t_{(3,4)}$
	(1)	(2)	(3)	(4)	(5)	(6)
Red River Delta * All girls	0.101	0.966**	-1.217	0.0939	0.919***	1.181***
	(0.212)	(0.429)	(0.954)	(0.0600)	(0.106)	(0.236)
North East * All girls	0.271	0.846*	0.348	0.114*	0.640***	0.810***
	(0.238)	(0.465)	(0.946)	(0.0641)	(0.111)	(0.246)
North West * All girls	0.122	-0.217	0	0.127	0.342**	-0.00461
	(0.467)	(0.857)	(0)	(0.0918)	(0.146)	(0.309)
North Central Coast * All girls	0.433*	0.423	0.292	0.0389	0.422***	0.609***
	(0.241)	(0.447)	(0.990)	(0.0729)	(0.111)	(0.234)
South Central Coast * All girls	-0.265	0.335	0.860	-0.157**	-0.112	0.158
	(0.250)	(0.482)	(1.509)	(0.0709)	(0.123)	(0.267)
Central Highlands * All girls	-0.685**	-0.624	0.00445	-0.0450	0.0194	0.0390
	(0.336)	(0.600)	(1.374)	(0.0783)	(0.118)	(0.236)
Southeast * All girls	-0.205	0.258	-0.810	-0.0520	0.116	-0.0805
	(0.255)	(0.428)	(0.882)	(0.0728)	(0.119)	(0.256)
Observations	5,102	2,620	1,026	23,707	18,080	6,527

The control variables such as mother's age at latest child's birth, urban location, education of the mother, logarithm of household income per capita, education of the father and seven dummies indicating the eight main economy regions are included but estimates not shown. Robust standard errors are in parentheses (*** p < 0.01, ** p < 0.05, * p < 0.1).

Table 5 Ward test on DID by region

	Logistic			Hazard		
H0: All DIDs are zero	RUM_1	RUM_2	RUM_3	$t_{(1,2)}$	$t_{(2,3)}$	$t_{(3,4)}$
	(1)	(2)	(3)	(4)	(5)	(6)
Test value	16.16	11.67	5.16	21.52	146.19	54.98
$Prob > chi2$	0.0237	0.1118	0.523	0.0031	0.000	0.000
Conclusion on H0	Accepted	Accepted	Accepted	Rejected	Rejected	Rejected

Original estimations are shown in the corresponding columns in Table 4.

Table 6 Level of son preference by household income per capita

	Logistic			Hazard		
Variables	RUM_1	RUM_2	RUM_1	RUM_2	RUM_1	$t_{(3,4)}$
	(1)	(2)	(3)	(4)	(5)	(6)
G	-0.331			0.203		
	(0.935)			(0.248)		
GG		1.345			-0.362	
		(1.740)			(0.403)	
GGG			-5.857			-0.439
			(3.789)			(0.787)
lnhhincomepc	-0.481***	-0.762***	-0.723***	-0.345***	-0.385***	-0.338***
	(0.0722)	(0.104)	(0.167)	(0.0216)	(0.0306)	(0.0483)
lnhhincomepc *All girls	0.0616	-0.0737	0.775*	0.00120	0.119***	0.123
	(0.101)	(0.189)	(0.423)	(0.0275)	(0.0454)	(0.0897)
Observations	5,102	2,620	1,029	23,707	18,080	6,527

The control variables such as mother's age at latest child's birth, urban location, education of the mother, logarithm of household income per capita, education of the father and seven dummies indicating the eight main economy regions are included but estimates not shown. Robust standard errors are in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

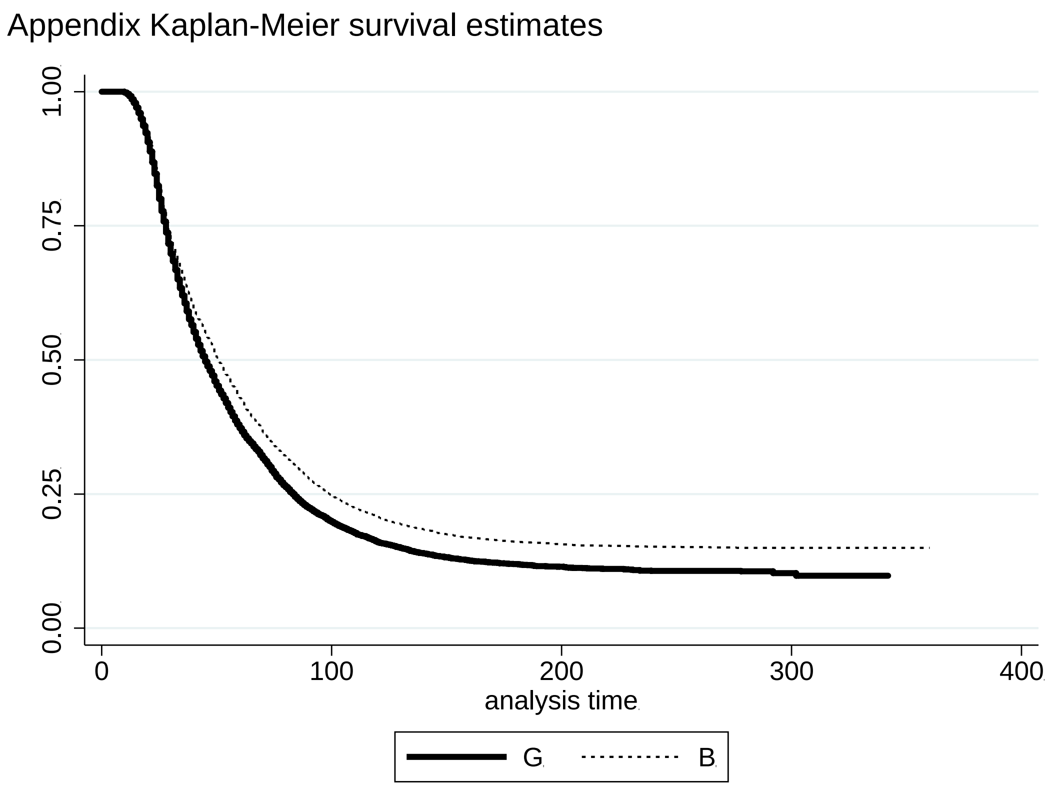

Fig. 1 Birth spacing in months between 1^{st} and 2^{nd} child by sex composition

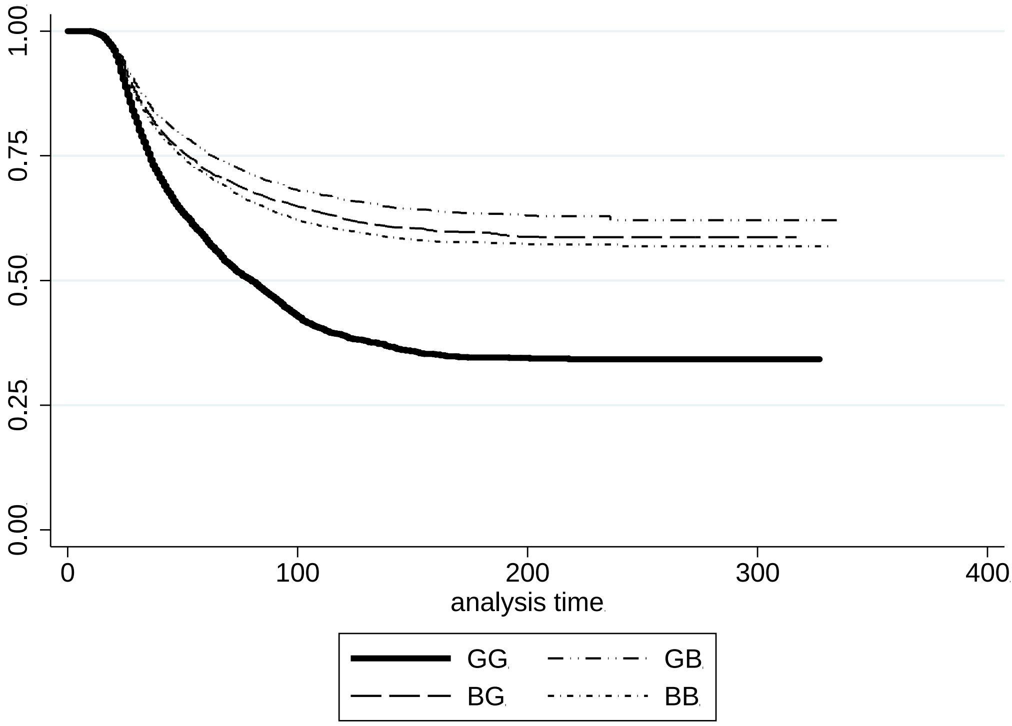

Fig. 2 Birth spacing in months between 2^{nd} and 3^{rd} child by sex composition

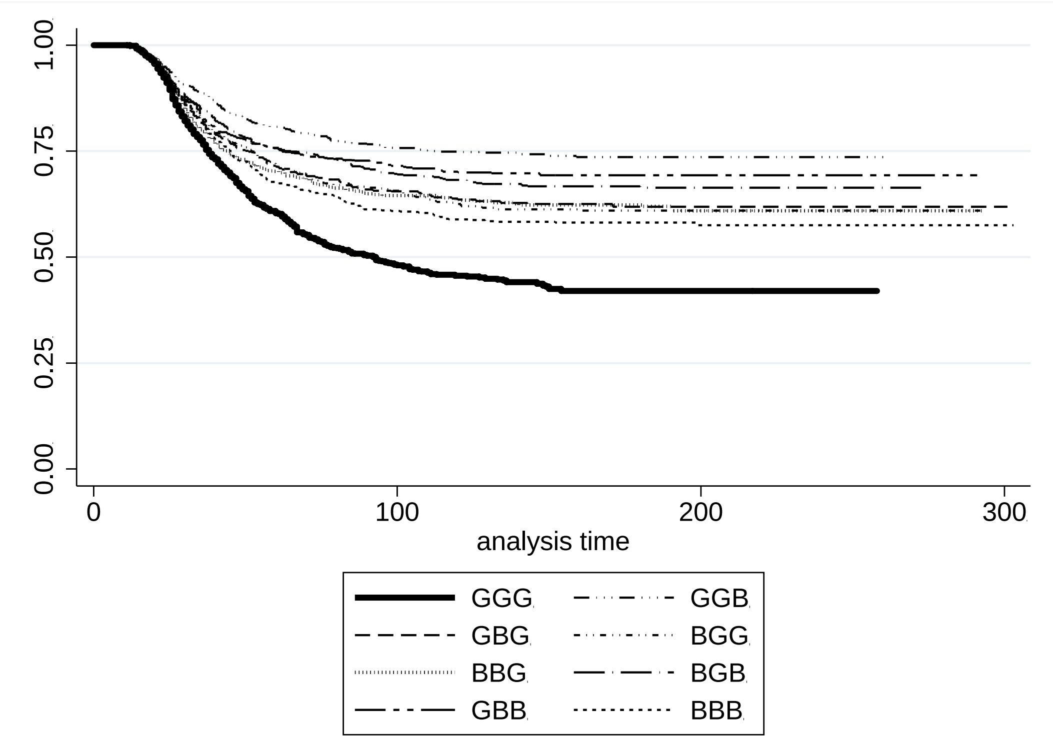

Fig. 3 Birth spacing in months between 3^{rd} and 4^{th} child by sex composition