Does Family Planning Policy Matter? Dynamic Evidence from China *

Fei Wang†

This version: December, 2013

Abstract

This paper estimates the dynamic effect of China's family planning policy on fertility using an individual-level panel sample from the China Health and Nutrition Survey. This paper applies a multiple-spell mixed-proportional hazard model where the unobserved individual heterogeneity is non-parametrically estimated, as suggested by Heckman and Singer (1984). Simulations from the model estimates find that the one-child policy, the harshest and ongoing family planning policy of China, reduced the probability of having exactly 2 and 3 births by 31.1% and 35.3%, and correspondingly raised probability of childlessness and having exactly 1 birth by 54.9% and 67.0%. Policy phases prior to the one-child policy have shown similar but smaller effects. However, simulations further show that, had there been no family planning policy, fertility levels would still have decreased greatly over cohorts. Family planning policy only explains about one third of the over-cohort fertility decline. Lastly, better-educated women are less likely to have a large number of births; women whose first birth is a son tend to have smaller families than those whose first child is a daughter.

Keywords: Family planning policy, Fertility, Duration analysis
JEL codes: J13, J18

*I am grateful to the guidance of my advisor, John Strauss, and to the insightful comments from Richard A. Easterlin, Jeffery Nugent, Anant Nyshadham, Geert Ridder, the audience of the USC Development/Applied Micro Seminar, the 2013 North American Summer Meeting of the Econometric Society, and the 2013 IUSSP International Population Conference. Errors are my own. Comments are welcome.

†Department of Economics, Dana and David Dornsife College of Letters, Arts and Sciences, University of Southern California, 3620 S. Vermont Ave. KAP 300, Los Angeles, CA 90089-0253. Email: wangfei@usc.edu.

1 Introduction

After World War II, family planning programs, aiming to lower birth rates, started to prevail in the developing world, triggered by popular beliefs that rapid population growth could obstruct the economic development of developing countries (Bongaarts, Mauldin, and Phillips (1990), Lapham and Mauldin (1985), Szreter (1993)). Soon afterwards, fertility rates substantially dropped in many developing countries, particularly in Asia and Latin America, during the 1970s and 1980s (Bongaarts, Mauldin, and Phillips (1990), Lapham and Mauldin (1985)). The concurrence of family planning programs and demographic transitions has spawned a large number of studies investigating their relations.

Such coexistence also emerged in China, the most populous country and the largest developing economy. China initiated family planning policy in 1963. The policy evolved over time, from mildly enforced version to the harshest one-child policy which is still operative at present. In parallel, China's fertility transition commenced at the beginning of 1970s. Figure 1 illustrates China's total fertility rates, in 1949–2001. During the 1950s and 1960s, fertility rates stayed at high levels, except for a dip caused by a great famine in 1959–1961. After 1970, fertility rates dived from 6 to nearly 2 within just 10 years, and continued to decline after 1980.[1]

[Figure 1 is here.]

This paper explores to what extent China's family planning policy can explain its fertility transition, and attempts to contribute additional evidence to the literature. Most studies examining the relation between family planning policy and fertility in China possess two defects. First, they are static analyses which are incapable to unveil dynamic policy effects on fertility. Second, they inappropriately created policy measures. Specifically, they might set up policy measures based on incomplete policy history, or construct them endogenously. Moreover, their measures generally fail to catch people's heterogeneous policy exposure. This paper adopts a semi-parametric dynamic duration model and constructs policy measures in a better way.

This paper uses a panel sample recording birth history of ever-married women, from the China Health and Nutrition Survey (CHNS), and applies a multiple spell duration model to evaluate policy effects for the first four birth spells. The fertility outcome is a dummy variable indicating whether a woman had a birth in some year. The duration model expresses the probability of having a birth in some year as a non-linear function of observed variables and unobserved individual heterogeneity. The observed variables include family planning policy and other demographic and socio-economic factors that have been con-

[1]The most recent official number of total fertility rate, 1.18 in 2010, failed to convince many scholars who believed the true figure should have been higher. Nevertheless, it has been a consensus that China's current fertility rate is below the replacement level.

sidered to be related to fertility. Their coefficients are estimated parametrically and are assumed to change over birth spells. China's family planning policy have three periods over time, and differ by residential location and ethnicity within each period. This paper creates policy variables by taking advantage of the secular and cross-sectional policy variations which are more complete than previous studies. Moreover, the policy measures include only exogenous variables and characterize women's heterogeneous policy exposure by age. Therefore, these measures to a considerable extent overcome the shortcomings of those constructed by previous studies.

The unobserved individual heterogeneity is assumed to follow a mass-point distribution, and is estimated non-parametrically. Specifically, the number of mass points and their locations and probabilities are all estimated from the data, as suggested by Heckman and Singer (1984).

Based on the estimated model, this paper derives the probability of childlessness, having exactly 1, 2, 3, and 4 or more births. The predicted probabilities match well with actual numbers in the data. By turning on and off the policy period by period, this paper further calculates the probability difference with and without some period of policy. Being exposed to the one-child policy reduces the probability of having exactly 2 and 3 births by 31.1% and 35.3%, respectively, and correspondingly increases the probability of childlessness and having exactly 1 birth by 54.9% and 67.0%. Earlier periods of policy present analogous patterns, but lesser effects. This paper further examines the policy effects by residential location and ethnicity. Generally, the policy has stronger effects for urban women and ethnic majorities than for rural women and minorities. These results are consistent with the policy history.

Moreover, this paper simulates fertility rates by birth cohort of women under different circumstances of policy exposure. The probability of having exactly 2 and 3 births under actual policy exposure are smaller for all cohorts than the case had women been never exposed to any policy. Consequently, the probability of childlessness and having exactly 1 birth is bigger for all cohorts in the former case. However, without any policy, fertility would still have demonstrated a downward trend over cohorts. Family planning policy only explains about one third of the fertility transition over cohorts.

In addition to family planning policy, other individual characteristics have shown noticeable impact. Better-educated women tend to substantially decrease their likelihood of childbearing. If a woman's first birth is a son, she would be much less likely to have a large number of births, which manifests the strong son preference in China.

This paper is arranged as follows. Section 2 briefly introduces the history of China's family planning policy. Section 3 reviews the literature studying relations between China's policy and fertility, and the application of duration models in relevant fields. Section 4 describes the data used by this paper. Section 5 derives the duration model and illustrates the way of constructing policy measures. Section 6 shows results for estimation, policy effects on fertility, and simulations. Section 7 concludes.

2 Brief history of China's family planning policy

China's family planning policy was launched in 1963, and is still active at present. The policy consists of propaganda, family planning service and birth quota.

Propaganda attempts to convince people that small families will enhance their own benefits and accelerate China's development. Family planning service is widely available at local clinics. People may get free contraceptives and access family planning surgeries at low costs. These two components are commonly observed in the family planning programs of other countries.

Birth quota, the distinctive feature of China's policy, sets ceilings for a family's number of births. A married couple will be penalized for having a higher level of births than the ceiling. Birth quota was adjusted over time, and varied by residential location and ethnicity. Table 1, a replicate from Wang (2012), lists how birth quota differed secularly and cross-sectionally.

[Table 1 is here.]

The evolution of birth quota is segmented into three periods. Birth quota was more and more stringent over periods. In each period, birth quota was tighter for urban or Han people, than for rural or non-Han people.[2]

Period 1 started in 1963. A great famine in 1959–1961 immensely lowered fertility rates. Chinese families made up births right after the famine so that fertility rates hit new high points in 1962 and 1963.[3] Concerned about uncontrolled population explosion, Chinese government initiated the first version of family planning policy, where birth quota functioned only for urban Han people, and the enforcement was weak. An urban Han married couple was allowed but discouraged to have three children. Having too many births might result in considerable political and social pressure.

The policy stepped into the second period in 1971. An urban Han family was allowed to have only two children. Meanwhile, birth quota spread to rural Han people, but was less stringent than the urban Han counterpart. Violations of birth quota would not only lead to political and social pressure, but might also cause material sanctions. For instance, food was distributed based on the number of members that a family was supposed to own.

[2]Han people are the ethnic majority of China. China's 2010 census indicated that 91.51% Chinese are Han Chinese.

[3]Figure 1 evidently shows this part of history.

If a family had too many births, they might be threatened by food shortage. Non-Han people at this time point were essentially not controlled by birth quota.

Period 3, the well-known one-child policy, started in 1980, and is still effective nowadays. An urban Han family can have only one child in most cases. A rural Han family may be able to have the second birth, if some conditions are met. For example, if their first birth is a daughter, they can have a second child, which results from the strong son preference in rural China. In this period, noncompliance will be sanctioned by laws. The most prevalent punishment is monetary penalties. Non-Han people are influenced by birth quota in this period, but the implementation is milder than that for Han people. Moreover, among non-Han people, the policy enforcement is weaker for rural people than for urban people.

Propaganda and family planning service have not changed as sharply as birth quota, but they have shown consistent patterns over time and across people. This paper utilizes the secular and cross-sectional variations in Table 1 to construct policy measures.

More detailed history is available in Wang (2012).

3 Literature review

3.1 Worldwide studies on family planning programs and fertility

Observing that family planning programs were followed by fertility decline in many developing countries, numerous studies began to explore their relations.

Mwaikambo et al. (2011) systematically reviewed 63 studies, published between 1995 and 2008, on evaluating family planning interventions. These studies range over African, Asian and Latin American countries, and investigated the effect of family planning programs on subjective outcomes, including knowledge of family planning and fertility preference, and objective outcomes, comprising use of services or contraceptives and fertility behavior.

Most studies that focused on subjective outcomes found significant program effects. 34 out of 38 studies found programs improved knowledge or attitudes; 18 out of 20 studies concluded that programs increased discussions of sexuality or family planning; 6 out of 7 studies argued that programs increased intention to practice family planning, or decreased fertility preferences.

However, the evidence for objective outcomes is more mixed. 4 out of 8 studies concluded programs increased service use; 36 out of 49 studies found programs increased contraceptive use or reduced unmet need; 6 out of 13 studies showed that programs reduced unintended pregnancy or abortion.

3.2 Studies on China's family planning policy and fertility

Similarly, studies on China's family planning policy also concluded differently. Many studies endorsed substantial policy effects on fertility (e.g, Lavely and Freedman (1990), Li, Zhang, and Zhu (2005), Yang and Chen (2004)), some emphasized that family planning policies are important, but one should not overlook the contribution of socio-economic variables (e.g, Poston and Gu (1987), Wang (1988)), and a large number of studies argued that the policy effects were overstated (e.g, Cai (2010), McElroy and Yang (2000), Narayan and Peng (2006), Schultz and Zeng (1995)).

These studies usually have two shortcomings. First, they rely on cross-sectional data and static analyses, and thus are hardly able to discover dynamic policy effects, especially the effects on the timing of childbearing (Hotz, Klerman, and Willis 1997). A few studies applied dynamic models. Ahn (1994) estimated the effect of one-child policy on the possibilities of having the second and third births in three provinces of China. Li and Choe (1997) examined the effect of one-child policy on the likelihood of having the second birth as well as the duration before the second birth. Poston (2002) studied the effect of gender of past births on future births, and they also controlled for a measure for the one-child policy. These papers generally ignored the policy phases prior to the one-child policy. Moreover, when studying multiple birth spells, they simply assumed independence among spells. This paper dynamically estimates the effects of all the three policy phases, and allows for flexible between-spell correlations.

Second, most literature created policy measures improperly. They may only take advantage of part of the secular and cross-sectional policy variations. Some studies utilized secular policy variations, but ignored cross-sectional differences (e.g, Edlund et al. (2008), Narayan and Peng (2006), Yang and Chen (2004)); some caught cross-sectional variations, but neglected secular evolution (e.g, Cai (2010)); more studies took both variations into account, but in incomplete ways (e.g, Banerjee, Meng, and Qian (2010), Islam and Smyth (2010), Li, Zhang, and Zhu (2005), Li and Zhang (2007), Li and Zhang (2008), Qian (2009), Wu and Li (2011)).

Some measures were endogenously constructed, which may bias estimation. For example, Yang and Chen (2004) used year dummies of being married to capture secular policy variations. Similarly, Banerjee, Meng, and Qian (2010), Edlund et al. (2008), Islam and Smyth (2010), Poston (2002) and Qian (2009) defined policy exposure according to whether some child was born under the policy.

Other than endogeneity, these dummy measures also lack heterogeneity. For example, if a 20-year-old woman and a 40-year-old woman both had a birth under the one-child policy, then they would have identical policy measures. However, the younger woman should have larger policy exposure because her age was closer to the peak age of childbearing. Wu and Li (2011) constructed a more heterogeneous measure, which was proportional to the length of time exposed to the policy, but they assumed that the intensity of policy exposure was the same over age.

Wang (2012) improved policy measures in static analysis, to make them more complete, exogenous and heterogeneous. This paper further improves the measurement in dynamic analysis.

3.3 Application of duration analysis

Duration analysis is used to analyze what determines the duration of a state. In social sciences, duration models were long ago adopted by labor economists who mainly focused on the determinants of unemployment duration. For example, Ham and Rea Jr (1987) and Katz and Meyer (1990) both studied the effect of unemployment benefits on unemployment duration.

Duration analysis has also been applied to other fields, particularly in birth behaviors (Arroyo and Zhang 1997). For example, Newman and McCulloch (1984) estimated risks of births at some age with various duration model specifications. Olsen and Wolpin (1983) estimated the effect of child mortality on fertility.

However, not many studies on family planning programs have adopted duration models. In addition to the three papers on China, reviewed in Section 3.2, Angeles, Guilkey, and Mroz (2005) jointly estimated how education, marriage and fertility responded to family planning programs in Indonesia; Hashemi and Salehi-Isfahani (2013) estimated the effect of family planning programs on fertility in Iran. This paper contributes more evidence to the branch of literature.

One of the most important econometric issues is to specify duration dependence and individual heterogeneity for duration models. Ideally, both duration dependence and heterogeneity should be flexibly specified. But Baker and Melino (2000) warned that, models with non-parametrically specified duration dependence *and* heterogeneity may substantially bias estimates. They found that non-parametrically specifying one of the two parts would be enough to generate convincing results. This paper non-parametrically specifies individual heterogeneity, and assumes duration dependence to have quadratic forms which can capture some non-monotonic patterns.

Heckman and Singer (1984) proposed an algorithm for the non-parametric estimation of individual heterogeneity. They let the heterogeneity follow a mass point distribution where the number, location and probability of supporting points all need to be estimated.

Specifically, they started to maximize the likelihood function with 2-point heterogeneity; after the model with M-point heterogeneity was estimated, a new model with $(M+1)$-point heterogeneity would further be estimated, until the overall likelihood could not be increased. They found estimated coefficients following such procedures were close to their true values. This paper adopts their method.

However, Baker and Melino (2000) showed that, such procedures may lead to over-searching problems: too many supporting points for the heterogeneity could largely bias estimates. Instead, they proposed a new criterion which on one hand maximizes the likelihood function but one the other hand penalizes too many supporting points. This paper also applies their criterion, and the estimated results remain the same.

4 Data

4.1 Original cross-sectional data

This paper studies the dynamic effect of China's family planning policy on fertility with a panel sample recording women's birth behaviors over their life cycles. The panel data were expanded from the birth history data of the China Health and Nutrition Survey (CHNS),[4] a pooled cross-sectional sample that has been used by the static analysis of Wang (2012).

The ongoing CHNS is one of the most widely used micro-data sets on China. Conducted by an international team, the CHNS collected information on household and individual economic, demographic, and social variables, particularly the factors about health and nutrition, every several years since 1989, across nine provinces.[5] A large group of interviewees have been followed longitudinally.

The CHNS surveyed ever-married women, below 52,[6] on their birth history,[7] in 1991, 1993, 2000, 2004, 2006, and 2009.[8] The CHNS team combined data of all rounds, kept only the most recent record for each woman, and released the refined pooled cross-sectional data online.[9] Other demographic and socio-economic variables are available in other modules, and can be merged to the birth history data.

[4] More information about the CHNS can be found on its official website: `http://www.cpc.unc.edu/projects/china`.

[5] Before the 2000 round, the survey covered eight provinces: Guangxi, Guizhou, Henan, Hubei, Hunan, Jiangsu, Liaoning, and Shandong. Heilongjiang was included in round 2000 and thereafter.

[6] Surveyed women were under 50 in round 1991. However, around 13% surveyed women were above the age ceilings. They are valid observations and are kept to enlarge the sample.

[7] The birth history includes the date of birth, gender, living arrangement, and the date of death of each child that a woman had ever had by the date of survey.

[8] The 2011 data came out in the summer of 2013, and are not included for analysis in this paper.

[9] The data used in this paper were posted in July 2011.

Birth history was reported only by ever-married women because marriage is traditionally and legally a pre-condition for childbearing in China. In the CHNS sample, the proportion of women who ever had births before marriage is below 5% and shows no rising trend over cohorts.[10]

I dropped women below age 15 from the data. Further, women who ever had births below 15 or above 49 were excluded.[11] The refined sample includes 6,533 women who were born in 1931–1991. The solid curve in Figure 2 shows the average number of births by birth cohort of women.

[Figure 2 is here.]

The number of births rises from about 1 to around 3.5, from cohort 1931 to 1942. Starting from cohort 1942, marked with a vertical dashed line, fertility levels decline over cohorts.

The dashed curve illustrates the number of births derived from one percent sample of China's 1990 census. This census interviewed women aged 15–64 on their total number of births.[12] The number of births dramatically falls over cohorts, from around 5 to below 1.

Compared to the census, the CHNS sample generates a divergent fertility trend for cohorts 1931–1941, implying this part of sample is far from being nationally representative. For cohorts 1942 and after, both data have displayed similar over-cohort trends.[13] Therefore, this paper further restricts the sample to cohorts 1942 and after. The number of women in the trimmed sample is 6,214.

4.2 Panel data and notations

The dynamic analysis of this paper relies on a panel sample which was expanded from the CHNS pooled cross-sectional data introduced above. The panel data record each woman's information at each age in years. A woman's observations span from age 15 to the age of being censored, or age 49, whichever is smaller. In other words, if a woman was 49 or younger in the survey year, her age of being censored would be exactly the age at the survey; if a woman was older than 49 in the survey year, her age of being censored would be 49.

[10] This is different from what Hotz, Klerman, and Willis (1997) presented about non-marital childbearing in the U.S. They pointed out, in the U.S., less than 6% of births were out-of-wedlock in 1963, while this proportion rose to 30% in 1992.

[11] The percentage of such women is only 0.4%.

[12] The census surveyed never-married women on their fertility, and such women take an extremely low proportion. I dropped them when plotting the dashed curve. Total number of births was reported by women born in 1926–1975. Figure 2 plots only for cohorts 1931–1975, for the sake of comparison.

[13] After cohort 1962, the census fertility trend rapidly goes beneath the CHNS trend, mostly because the younger interviewees were far from completing childbearing in 1990.

The childbearing life of a woman is segmented by her birth behaviors. The first birth spell, or spell 1, is from age 15 to age at the first birth. Similarly, spell $j + 1$ indicates the time interval between the j'th birth and the $(j + 1)$'th birth.

Through out the paper, I use i to denote a woman, j to represent a spell, and k to indicate a year within a spell. Within spell j of woman i, $k = 1, 2, ..., K_{ij}$ where K_{ij} is the observed length of duration in years of that woman-spell.

The fertility outcome variable y_{ijk} is a dummy variable indicating whether woman i has new births during the k'th year of spell j. It equals 0 if woman i remains in spell j, after k years stay in that spell. If y_{ijk} equals 1, woman i exits spell j during the k'th year of that spell, and then enters spell $(j + 1)$ in the following year.

4.3 Descriptive statistics

Detailed descriptive statistics of the original pooled cross-sectional sample can be found in Wang (2012). Table 2 describes the number and percentage of women who have a certain number of birth spells. Women with exactly j spells are almost equivalent to those with $(j - 1)$ births, because most women did not have births during their last year in the sample.[14]

[Table 2 is here.]

Only 5.2% women have exactly 1 spell, and 45.7%, 29.7% and 12.3% women have exactly 2, 3 and 4 spells, respectively. Women with exactly 1, 2, 3 and 4 spells comprise 93% of all the sampled women. This paper only analyzes birth behaviors for the first four spells. In other words, the whole birth history of the 93% women is included for analysis, and for the rest 7%, their first four spells are considered, and they are assumed to be censored in the year of having the fourth birth.[15]

Table 3 shows descriptive statistics of selected variables for the first four spells. Variables in each spell are summarized based on the women who experienced that spell.

[Table 3 is here.]

94.2% women ended the first spell with births, and their average waiting time before the first birth is about 10 years. Over spells, the percentage of women having births at the end of spells largely dropped, as well as their duration before births.

Urban women, Han women, more educated women and women living in the more developed coastal provinces are less likely to appear in high order birth spells. Older cohorts of

[14] The percentage of women having births during their last year in the sample is only 1.01%.

[15] This paper assumes that parameters vary across spells. Sample size of spells beyond 4 is too small to estimate a different set of parameters, and therefore this paper only focuses on the first four spells. I also tried to include all spells, but assume spells beyond 4 share the same parameters with spell 4. This generates similar results.

women, cohorts 1942–1960, tend to have more births than younger cohorts. More interestingly, the percentage of the first birth being a son decreases over spells, implying that a woman whose first birth is a daughter tends to have more births later, which reflects son preference in China (Das Gupta et al. (2003), Jensen and Chintan (2003)).

5 Duration model and policy measurement

5.1 Duration model

Duration analysis focuses on what affects the duration of a state. It has been widely used for studies on, for instance, unemployment durations (for example, Ham and Rea Jr (1987), Katz and Meyer (1990)). Many other fields, including birth behaviors, have also adopted this method (Arroyo and Zhang 1997).

Social scientists often reinterpret durations into decisions made at each period. For example, if a woman's duration between her first and second births is 10 years, then we can equivalently state that the woman decides not to have a new birth until the 10th year after her first birth.

Technically, duration analysis may start with specifying a hazard function $\theta(t)$,[16] which illustrates the rate of exiting some state (or some birth spell in this paper) at time point t after having stayed in the state for t. Here t is a continuous measure of time.

By definition,

$$\theta(t) = \lim_{\Delta t \to 0} \frac{Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{1 - F(t)}, \tag{1}$$

where T indicates durations, and $F(t)$ and $f(t)$ are the distribution function and density function of T. Conversely,

$$F(t) = 1 - exp\left(- \int_0^t \theta(s) ds \right). \tag{2}$$

Hazard functions are assumed to differ over spells, specified as $\theta_j(t)$, with $j = 1, ..., J$. Further, woman i's hazard rate in spell j does not only depend on the elapsed time t, but also relies on her observed time-varying characteristics $\mathbf{w}_i^H(t)$, time-invariant variables $\mathbf{z}_i$ and

[16] Natural scientists usually start with specifying the distribution function of duration T, $F(t)$, constructing likelihood functions based on observed durations, and estimating parameters of the distribution function using maximum likelihood methods. As social scientists, particularly economists, prefer to study people's decisions, hazard functions would be a more natural start (Van den Berg 2001). But the two ways are technically equivalent, reflected by the equivalence of equation (1) and (2).

her unobserved heterogeneity v_i which is also called frailty in duration analysis. Therefore, hazard functions can be expressed as $\theta_j(t|\mathbf{w}_i^H(t), \mathbf{z}_i, v_i)$, where $\mathbf{w}_i^H(t) = \{\mathbf{w}(\tau)|0 \leq \tau < t\}$, representing the entire history of $\mathbf{w}_i(\tau)$ from time point 0 to t.

One of the most popular specifications for $\theta_j(t|\mathbf{w}_i^H(t), \mathbf{z}_i, v_i)$ is the Mixed Proportional Hazard (MPH) function (Van den Berg 2001), i.e,

$$\theta_j(t|\mathbf{w}_i^H(t), \mathbf{z}_i, v_i) = \psi_j(t)\phi(\mathbf{w}_i^H(t), \mathbf{z}_i; \boldsymbol{\beta}_j)v_i. \tag{3}$$

$\psi_j(t)$, the baseline hazard, is a function of elapsed time for spell j. $\phi(\mathbf{w}_i^H(t), \mathbf{z}_i; \boldsymbol{\beta}_j)$ is a function of individual characteristics and allows them to have spell-specific coefficients.

Two more assumptions are needed (Wooldridge 2010).

Assumption 1. *Only contemporaneous covariates matter.*

$$\theta_j(t|\boldsymbol{w}_i^H(t), \boldsymbol{z}_i, v_i) = \theta_j(t|\boldsymbol{w}_i(t), \boldsymbol{z}_i, v_i).$$

Assumption 1 is strong. However, it would be much less restrictive if contemporaneous time-varying variables are constructed to accumulate past information, which is the case in this paper. For simplicity, let $[\mathbf{w}_i(t), \mathbf{z}_i] = \mathbf{x}_i(t)$.

Assumption 2. *Time-varying variables are locally constant within one year.*

$$\boldsymbol{x}_i(t) = \boldsymbol{x}_i(k) \text{ if } k - 1 \leq t < k$$

Theoretically t is continuous, but actual data are usually grouped. For example, one time unit in this paper is one year. Assumption 2 is crucial to simplify analyses for grouped data.

Under the two assumptions, hazard functions can be connected to observed fertility outcomes y_{ijk}.

$$
\begin{aligned}
&Pr(y_{ijk}=1|y_{ij1}=0,...,y_{ij(k-1)}=0) \qquad (4)\\
=&Pr(k-1\leq T<k|T\geq k-1)\\
=&\frac{F(k)-F(k-1)}{1-F(k-1)}\\
=&1-exp\left(-\int_{k-1}^{k}\psi_j(s)\phi(\mathbf{x}_i(s);\boldsymbol{\beta}_j)v_i ds\right)\\
=&1-exp\left(-\left(\int_{k-1}^{k}\psi_j(s)ds\right)\phi(\mathbf{x}_i(k);\boldsymbol{\beta}_j)v_i\right)
\end{aligned}
$$

All terms in the first three lines of equation (4) are conditional on $\mathbf{x}_i(t)$ and v_i. The second line translates birth decisions into durations. The third line simply uses the conditional probability formula. Equation (2) is plugged in to generate the fourth line. Assumptions 1 and 2 apply for the last two lines.

Figure 3 illustrates how the probability of having the jth birth changes by the duration after the $(j-1)$th birth ($j=1,2,3,4$). The probability of having the first birth is bell-shaped over duration, reaching the peak in the ninth year. The probabilities of having the second, third and fourth births all rise first and then fall over duration, with peaks in the second year. As the probabilities of childbearing by duration exhibit non-monotonic patterns, $\int_{k-1}^{k}\psi_j(s)ds$ is simply parametrically specified as $exp(a_jk^2+b_jk+c_j)$.

[Figure 3 is here.]

Further, assume $\phi(\mathbf{x}_{ijk};\boldsymbol{\beta}_j) = exp\left(\mathbf{x}_{ijk}\boldsymbol{\beta_j}\right)$, which is a common specification. Then, Equation (4) becomes

$$
\begin{aligned}
&Pr(y_{ijk}=1|y_{ij1}=0,...,y_{ij(k-1)}=0) \qquad (5)\\
=&1-exp\left(-exp(c_j+b_jk+a_jk^2+\mathbf{x}_i(k)\boldsymbol{\beta}_j+ln(v_i))\right).
\end{aligned}
$$

Lastly, $ln(v_i)$ is non-parametrically specified, and is assumed to follow a mass-point distribution, as suggested by Heckman and Singer (1984). Assume its distribution has M supporting points, with locations $q_1,...,q_M$ and probabilities $p_1,...,p_M$. All p's and q's are unknown parameters such that $\sum_{m=1}^{M}p_m=1$ and $\mathbb{E}\left(ln(v_i)\right)=0$. Heckman and Singer (1984) showed that such frailties triumph over parametrically specified ones in uncovering true coefficients of interest. Practically, one starts with a 2-point frailty, and estimates the locations and probabilities of the two supporting points. Then, the frailty owns three supporting points if adding the third point generates larger likelihood; such steps continue until the likelihood function cannot be improved. The procedure of estimating the distribution of frailty is discussed in Appendix A.

An intuitive way of understanding the model is to start with its underlying model, as follows.

$$y^*_{ijk} = c_j + b_j k + a_j k^2 + \mathbf{x}_i(k)\boldsymbol{\beta}_j + ln(v_i) + \varepsilon_{ijk}. \tag{6}$$

y^*_{ijk} is woman i's utility gained from having the j'th birth in the k'th year after her last birth. It is determined by elapsed time in spell j, observed characteristics, frailty, and error term ε_{ijk}. By construction, observed birth decisions y_{ijk} can be linked to y^*_{ijk} in the following way.

$$y_{ijk} = \begin{cases} 1 & \text{if } y^*_{ijk} \geq 0 \\ 0 & \text{if } y^*_{ijk} < 0 \end{cases} \tag{7}$$

Assume ε_{ijk} follows the standard type I extreme value distribution, then equation (5) stands out.

$ln(v_i)$ is a common factor over woman i's spells, so it somewhat captures her between-spell correlations, which is crucial for a multi-spell model. A more general way is to specify frailties for each woman-spell, v_{ij}, and allow v_{ij}'s to be correlated for woman i. This paper did not adopt this specification, because firstly, calculation burden would be heavy, and secondly, the error term in equation (6) has already captured some between-spell correlations in a general way.

The likelihood function of woman i is

$$\begin{aligned} &\mathcal{L}_i \qquad (8) \\ =&\sum_{m=1}^{M} p_m \left(\prod_{j=1}^{J_i} \prod_{k=1}^{K_{ij}} Pr(y_{ijk}=1|y_{ij1},...,y_{ij(k-1)},q_m)^{y_{ijk}} Pr(y_{ijk}=0|y_{ij1},...,y_{ij(k-1)},q_m)^{1-y_{ijk}} \right), \end{aligned}$$

where

$$\begin{aligned} &Pr(y_{ijk}=1|y_{ij1},...,y_{ij(k-1)},q_m) \qquad (9) \\ =&1 - exp\left(-exp(c_j + b_j k + a_j k^2 + \mathbf{x}_i(k)\boldsymbol{\beta}_j + q_m\right), \end{aligned}$$

and

$$
\begin{aligned}
&Pr(y_{ijk} = 0|y_{ij1}, ..., y_{ij(k-1)}, q_m) \\
=&exp\left(-exp(c_j + b_j k + a_j k^2 + \mathbf{x}_i(k)\boldsymbol{\beta}_j + q_m\right).
\end{aligned} \tag{10}
$$

Then, the likelihood function for all women is $\mathcal{L} = \prod_{i=1}^{I} \mathcal{L}_i$. An implicit assumption is that given covariates and frailties, women's birth decisions are independent over spells, and among each other.

5.2 Policy measurement

Woman i's exposure to the period r policy ($r = 1, 2, 3$) at age a is defined as

$$
FPP_i^r(a) = \sum_{age=15}^{a} p(age) I(age \in [a_i^{rS}, a_i^{rE}]). \tag{11}
$$

Woman i's exposure to policy r started at age a_i^{rS}, and ended at age a_i^{rE}. For example, a woman born in 1960 started to be exposed to the period 2 policy (1971–1979) at age 15, and stopped being exposed to the policy at age 19; a woman born in 1950 started to be exposed to the period 3 policy (1980–present) at age 30, and stopped being exposed at the end of her last year in the sample. Indicator function $I(.)$ tells whether woman i was exposed to policy r at some age.

Exposure indicators are further weighted by function $p(age)$, the probability of childbearing at some age. $p(age)$ is estimated based on birth records of women of cohorts 1942–1991 in the CHNS sample, and is applied to all women.[17] Such a specification assumes that women's exposure to the same policy in the same year is different by age: a woman is more effectively influenced by a policy if she is exposed at her peak age of childbearing, than a woman who is exposed at younger or older age when she is physiologically less likely to have births. This specification has two limitations. First, it is more restrictive than the one allowing age-specific policy effects. Second, it assumes the age-specific effects are consistent with the shape of $p(age)$, which may or may not be true. Nevertheless, this setting to some extent captures heterogeneous policy effects by age, without largely increasing the burden of estimation.

Lastly, woman i's exposure to the period r policy at age a is constructed as the accumu-

[17]Ideally, $p(age)$ should be estimated on women who have never been affected by family planning policies. But such women don't exist in the sample. Wang (2012) estimated $p(age)$ on different groups of women, and showed that results were robust to $p(age)$ of various shapes.

lation of all her past exposure to the policy. It implies that, even though a policy ended, it would still affect women's birth decisions in the following, reflecting the long-run effect of family planning policy.[18] Assumption 1 is less restrictive with such policy measures.

Figure 4 visualizes equation (11). The curve illustrates the probability of childbearing from age 15 to 49, estimated from the birth record of cohorts 1942–1991 women in the CHNS sample. The shaded area gives an example of policy exposure. Assume $a_i^{rS} = 20$ and $a_i^{rE} = 30$. If a woman's current age a is below 20, then her policy exposure is 0; if a is between 20 and 30, then her exposure is the shaded area between age 20 and a; if a is above 30, then her exposure is the whole shaded area.

[Figure 4 is here.]

Figure 5 shows the maximum exposure to each of the three phases by cohort. The maximum exposure of a woman is her exposure at the end of her last year in the sample, accumulating all her exposure to some policy. The cohort maximum exposure is then the average of individual maximum exposure for a cohort. It tells us that cohorts around 1940 were exposed mostly to the period 1 policy, and the period 2 and 3 policies were strongest for cohorts around 1950 and 1960, respectively.

[Figure 5 is here.]

This paper assumes that birth decisions were made one year before the actual birth year, therefore birth behaviors at age a are assumed to be related to policy exposure up to age $(a-1)$. As the policy varied cross-sectionally, suggested by Table 1, policy measures are therefore further interacted with urban dummy and Han dummy in the model.

5.3 Other covariates

Other than policy measures, the model further controls for exposure to the 1959–1961 great famine, not only because the famine had a huge impact on fertility, as shown in Figure 1, but also because the period 1 policy was triggered by extensive birth makeup after the famine. Famine exposure is similarly constructed, by letting a_i^{rS} and a_i^{rE} be the starting and ending age of being exposed to the famine.[19]

The model also controls for the urban dummy and Han dummy alone. They are assumed to be time-invariant. The urban dummy is defined according to the residential location reported in women's most recent interviews.[20] However, urban women might live in

[18] Examples of long-run effects include permanently changed fertility preference, infertility due to sterilization, etc.

[19] The famine officially ended before 1962, followed by the fertility makeup in 1962 which is also displayed in Figure 1. Therefore, a_i^{rE} is assumed to be the age in 1962.

[20] The urban dummy can also be defined on household registration types, usually known as *hukou* types. But nearly 20% women did not report their *hukou* types.

rural areas before, and rural women might come from urban areas. Either case would underestimate the urban-rural gap of the policy effects, given that the policy was designed to be enforced more stringently in urban areas. But even though, the estimated gap is still fairly large.

Women's years of schooling were reported in their most recent surveys, and their schooling history is deduced by assuming that they started schooling at age 6. The model then controls for three dummies for women's highest education levels: finishing primary school, finishing middle school and finishing high school, treating not finishing primary school as the base group.[21] This paper, like many other studies, assumes education is exogenous to fertility decisions. It would be less restrictive if both education and fertility are assumed to be endogenous and are jointly estimated, as in Angeles, Guilkey, and Mroz (2005).

Many other important determinants of fertility are not available, including various prices directly or indirectly related to having and raising children, the history of income and wealth, infant mortality rates, and so on. Instead, this paper controls for a number of variables which aim to proxy the omitted variables. They are birth cohort trend of women and its square,[22] regional dummies,[23] cohort trend $\times$ urban dummy, cohort trend $\times$ Han dummy, cohort trend $\times$ coastal province dummy,[24] age of women and its square, and a dummy indicating whether the first birth is a son.[25]

6 Results

6.1 Estimation results

Table 4 shows the coefficients of selected variables estimated from the mixed proportional hazard duration model introduced in Section 5.1. Coefficients are assumed to differ by birth spell. Robust standard errors are shown in parentheses. Three stars, two stars and one star symbolize statistical significance at levels of 1%, 5% and 10%.

[Table 4 is here.]

[21]Denote years of schooling by S. Finishing primary school, middle school and high school correspond to $6 \leq S < 9$, $9 \leq S < 12$ and $S \geq 12$. The base group means $S < 6$.

[22]Cohort trends are derived by subtracting 1942 from the birth year of women.

[23]The nine provinces in the sample are grouped to four regions: Northeast China, East China, South Central China, and Southwest China. The division of regions are purely geographic. Heilongjiang and Liaoning belong to the Northeast China, Shandong and Jiangsu belong to the East China, Hubei, Henan, Hunan and Guangxi belong to the South Central China, and Guizhou belongs to the Southwest China.

[24]Liaoning, Jiangsu and Shandong are coastal provinces, representing relatively developed regions. Guangxi is by the ocean, too, but it's usually not regarded as a developed province.

[25]As all women entered the sample at age 15, age is perfectly collinear with duration in the first birth spell. Therefore, age and its square are not controlled for in spell 1. Gender of the first birth is also controlled for in spell 2 and beyond.

As the hazard function can be expressed as $ln\left(\theta_j(k|\mathbf{x}_i(k), v_i)\right) = c_j + b_j k + a_j k^2 + \mathbf{x}_i(k)\boldsymbol{\beta}_j + ln(v_i)$, the coefficients in Table 4 can be interpreted as the semi-elasticity of hazard rates to independent variables. Non-reported variables include great famine exposure, cohort trend and its square, regional dummies, cohort trend $\times$ urban dummy, cohort trend $\times$ Han dummy, and cohort trend $\times$ coastal province dummy.

The first nine coefficients express the effect of the three periods of family planning policy on fertility. Policy effects cannot be directly read from the coefficients. Below in Section 6.2, I calculate the partial effect of each period of policy on the probability of having a certain number of births. Table 4 shows Chi-squared statistics for the joint statistical significance of policy variables for each birth spell, as well as corresponding p values. In the first three spells, the policy variables are strongly significant; while in the fourth spell, they are statistically insignificant determinants.

All three education dummies exhibit negative and statistically significant effects on fertility, except that in the last column only finishing high school education shows significant effect. Moreover, the effect of education is generally stronger as education levels rise.

The coefficients of duration and its square are significant in all spells, and support inverse-U shapes for duration dependence. Figure 6, echoing Figure 3, predicts the probability of having the jth birth by the duration after the $(j-1)$th birth ($j = 1, 2, 3, 4$). The probability peaks around the ninth year in spell 1, and around the first year in the rest spells, which are essentially consistent with Figure 3.[26]

[Figure 6 is here.]

Women's age also shows an inverse-U pattern in spell 2; but in spells 3 and 4, the likelihood of childbearing decreases as age rises. Lastly, if a woman's first child is a son, she is much less likely to have more births. The son preference effect persists through all birth spells.

The estimated frailty has three supporting points: $\hat{q}_1 = -5.6879$, $\hat{q}_2 = -1.6396$, and $\hat{q}_3 = 0.21167$, with probabilities $\hat{p}_1 = 0.0175$, $\hat{p}_2 = 0.0585$, and $\hat{p}_3 = 0.924$. Table 5 tests whether the 3-point frailty is different from no frailty or a 2-point frailty. Panel A of Table 5 implies that the 3-point frailty is significantly different from no frailty. Similarly, Panel B shows that the 3-point frailty is significantly different from a 2-point frailty.

[Table 5 is here.]

Baker and Melino (2000) pointed out that, frailty with too many supporting points may bias estimation. Traditionally, the criterion for determining the number of supporting points of frailty is maximizing the likelihood function. They revised the criterion by

[26] Although the predicted probabilities of having the second, third, and fourth births fail to display non-monotonic patterns, they do capture some curvature of the duration dependence.

imposing a term which penalizes too many parameters, and found that estimates are more reliable under the refined criterion. With their criterion, the frailty with 3 points is still the optimal choice.

6.2 Policy effects on the probability of having a certain number of births

This section calculates policy effects on the probability of having a certain number of births. As estimation is based on the first four birth spells, I can derive the probabilities of having *exactly* 0, 1, 2, and 3 births, and the probability of having 4 or more births. Details of computing these probabilities are in Appendix B.

I first calculate the probability of having a certain number of births under actual policy exposure, and then compare it with the actual proportion of women having that certain number of births in the sample, to check how well the estimated model fits the data. Table 6 shows the comparison. In general, the predicted probabilities fit the actual counterpart very well.

[Table 6 is here.]

Figures 7 to 10 display how well the predicted probabilities of having exactly 1, 2, 3, and 4 or more births match the actual proportions over women's birth cohort. Generally, the goodness of fit is satisfactory, but the probabilities of having exactly 1, 2 and 3 births do not well match for the youngest cohorts, and the probability of having 4 or more births is systematically, though slightly, underestimated for the oldest cohorts.

[Figures 7 to 10 are here.]

The probability of childlessness is not shown for checking the goodness of fit, because it can be obtained simply by subtracting other probabilities from 1. If other predicted probabilities match well, so would the probability of childlessness.

As discussed above, I have predicted the probabilities of having exactly 0, 1, 2, 3, and 4 or more births when all independent variables, particularly the policy measures, take their actual values. Then, I turn off the period r policy ($r = 1, 2, 3$), and calculate the probability of having a certain number of births had women not been exposed to policy r, all the other variables, including the other two phases of policy, remaining unchanged. Further, the difference between the former and latter probabilities are defined as the effect of the period r policy on the probability of having a certain number of births.

We would expect that, with exposure to a policy, a woman tends to reduce her probability of having a large number of births, and correspondingly increase the probability of having a small number of births. Therefore, Table 7 first lists policy effects on the probabilities of having exactly 2, 3, and 4 or more births, and then shows what happens to the probabilities of childlessness and having exactly 1 birth. Numbers in squared brackets are percentage changes.

[Table 7 is here.]

Exposure to the period 1, 2, and 3 policy reduces the probability of having exactly 2 births by 0.025, 0.045, and 0.139 (or by 7.5%, 12.6%, and 31.1%), respectively. The effect of the period 3 policy is greater than period 2, and the effect of the period 2 policy is larger than period 1, which is consistent with the policy history.

Similarly, exposure to the three phases decreases the probability of having exactly 3 births by 0.007, 0.020, and 0.067, respectively. The probability change is smaller than the case of having exactly 2 births, but the percentage change is in similar scales.

It is difficult to understand the positive policy effect on the probability of having 4 or more births. Firstly, these effects may be statistically insignificant, as policy variables are jointly insignificant for the fourth birth spell in Table 4. Secondly, they could be the result of anticipation effects. Women raised the probability of having 4 or more births while exposed to policy, probably because they shifted future births to earlier dates after they anticipated that they might have to pay a higher price for their desired family size in the future as policy sanctions tended to be stronger and stronger. As policy became more and more stringent, it would be more and more difficult to shift future births, and thus the anticipation effect would be smaller and smaller. Table 7 indeed shows decreasing effects over policy periods.

Next, exposure to the three phases increases the probability of childlessness, by 2.3%, 5.9%, and 54.9%, respectively, and raises the probability of having exactly 1 birth by 0.9%, 8.7%, and 67.0%, respectively. Again, both sets of effects are stronger over policy periods. Particularly, the period 3 policy has much greater effects than earlier periods.

6.3 Heterogeneous policy effects by residential location and ethnicity

Table 8 shows policy effects on the probability of having a certain number of births for urban Han, rural Han, urban non-Han and rural non-Han women.

[Table 8 is here.]

Exposure to the three periods of policy lowers the probability of having exactly 2 births for all the four groups of women. Within every group, policy effects get larger over periods. Policy effects corresponding to the second and third phases are generally bigger for urban or Han women, than for rural or non-Han women. Such cross-sectional patterns also fit the policy history.

Effects on the probability of having exactly 3 births are also rising over policy periods. But different from the case of having exactly 2 births, policy effects, particularly the the effect of period 3 policy, tend to be relatively greater for rural or non-Han women. It is because rural or non-Han women desire more births than urban or Han women, so that they may respond to policy exposure more actively at higher order of births.

Because rural or non-Han women tend to desire larger families, they are more likely to shift future births to earlier dates, and thus generate bigger anticipation effects. The third panel of Table 8 supports the hypothesis. Moreover, within each group, anticipation effects, if any, decline over policy periods.

In the last two panels, within each group, policy effects become stronger over periods. Particularly, the effect of the period 3 policy is much larger than earlier phases for each group of women. Furthermore, policy effects on the probability of having exactly 1 birth are generally greater for urban or Han women, than for rural or non-Han women.

6.4 Policy effects on fertility transition

This section tries to answer the question posed at the beginning: To what extent can family planning policy explain China's fertility transition?

Fertility transition in Figure 1 is expressed by decreasing total fertility rates over calendar years. This paper illustrates fertility transition with dropping births over cohorts of women, particularly after the 1942 cohort. In other words, this section explores to what extent China's family planning policy can explain fertility decline over cohorts of women.

From Figure 1, we should further notice that, during the transition in 1970s, women's fertility decisions were only affected, if any, by the period 1 and 2 policy. In 1980, the period 3 policy started, and might account for some fertility decline afterwards. Therefore, this section is also interested in comparing the effects of the first two periods of policy and the period 3 policy.

To address this question, I first predict the probability of having a certain number of births over cohorts had they never been exposed to any policy. Second, turn on the period 1 and 2 policy only, assuming the period 2 policy has been enforced all the time since it started in 1971, and predict the probability of childbearing over cohorts. Third, further let the period 3 policy start, and predict the probability under actual exposure to all policy phases.

Fertility decline over *all* cohorts may not be of interest, because for fairly young cohorts, their low fertility levels largely reflect the fact that they had not yet finished childbearing by the time of survey. Therefore, we need to restrict analysis to sufficiently old cohorts such that they had essentially completed childbearing by the time of survey.

Implied by Figure 4, women's probability of childbearing at age 35 or any age above is below 0.01. Figure 11 further shows the age of being censored by cohort. The 1972 cohort, marked with a dashed vertical line, and older cohorts, were 35 or above in the year of survey. Therefore, only these older cohorts are considered for this section's analysis.

[Figure 11 is here.]

Similarly, we will go through the probability of having 2, 3 and 4 or more births first, and then move to childlessness and having exactly 1 birth.

In Figure 12, the solid line indicates the predicted probability of having exactly 2 births over cohorts had women not been exposed to any policy. The short-dashed line predicts the counterfactual probability of having 2 births had women been only exposed to the period 1 policy and the long-lasting period 2 policy. The long-dashed line illustrates the probability of having 2 births with actual exposure to all police phases.

[Figure 12 is here.]

Without any policy, the probability of having 2 births are between 0.5 and 0.6, and slowly drops to 0.5 over cohorts. The period 1 and long-lasting period 2 policy pulls down the probability to below 0.4. The period 3 policy has no additional effects in lowering the probability.

In Figure 13, the probability of having exactly 3 births had women not been exposed to any policy drops from 0.3 to 0.1 over cohorts. Appearance of the period 1 and long-lasting period 2 policy decreases the probability for older cohorts, but the effect is smaller and smaller over cohorts. In contrast, the period 3 policy has shown large additional effects since it started.

[Figure 13 is here.]

Figure 14 shows probability only for cohorts 1953 to 1972, because Figure 10 under-predicts the probability of having 4 or more births for cohorts before 1953. If the story of anticipation is true, this figure demonstrates that anticipation effects tend to die out over cohorts, and such effects are much smaller if the period 3 policy is working.

[Figure 14 is here.]

Figure 15 shows that, without any policy, the probability of childlessness would gradually rise from nearly 0 to 0.04. Policies level up the probability for all cohorts. The period 1 and long-lasting period 2 policy displays larger effects than the period 3 policy, but their gap is small in magnitude.

[Figure 15 is here.]

Lastly, Figure 16 shows, without any policy, the probability of having exactly 1 birth increases from about 0.1 to nearly 0.4 over cohorts. With the period 1 and long-lasting period 2 policy, the probability rises from 0.1 to nearly 0.5. After the period 3 policy starts, the probability increases from 0.1 to 0.6.

[Figure 16 is here.]

To summarize Figures 12 to 16, first, family planning policy decreases the probability of having exactly 2 and 3 births for all birth cohorts, and correspondingly increases the probability of childlessness and having exactly 1 birth. Second, the period 3 policy in general has larger effects than the period 1 and 2 policy. Third, even though without any policy, the probability of having a certain number of births are still moving towards lower fertility levels over cohorts.

It is also of interest to investigate policy effects on the number of births. The expected number of births for a cohort of women can be expressed as $E(n) = P_1 + 2P_2 + 3P_3 + \sum_{i\geq 4} iP_i$, where P_i $(i = 1, 2, ...)$ indicates the probability of having exactly i births. P_1, P_2 and P_3 have been estimated from the model, but P_i $(i \geq 4)$ is not available. Therefore, I rewrite the equation as $E(n) = P_1 + 2P_2 + 3P_3 + P_{\geq 4}E(n|n \geq 4)$, where $P_{\geq 4}$ is the probability of having 4 or more births which has been estimated, and $E(n|n \geq 4)$ measures the expected number of births had women had 4 or more births. Strictly speaking, $E(n|n \geq 4)$ cannot be derived from the model, but I can approximately estimate it for each cohort, by calculating the average number of births only for the women with 4 or more births in that cohort.[27]

Figure 17 plots the actual (solid line) and predicted (dashed line) number of births over cohorts. Generally, it fits well, except for the cohorts before 1953 and after 1983. Following similar procedure, Figure 18 plots the predicted number of birth in various scenarios of policy exposure. This figure ranges from cohorts 1953 to 1972, the part with excellent goodness of fit and with women who have completed childbearing.

[Figure 17 and 18 are here.]

The solid line represents the predicted fertility without any family planning policy, the short-dashed line means the predicted number of births had women been only exposed to the period 1 and long-lasting period 2 policy, and the long-dashed line predicts the number of births under actual exposure to all policy phases.

Figure 18 tells a similar story. Without family planning policy, the number of births declines by about 0.6 over the 20 years of cohorts; with all phases of family planning policy, it drops by about 0.9. Therefore, family planning policy explains only about one third of the over-cohort fertility decline. Out of the one third, the one-child policy alone explains a little bit more than half. This is a rough calculation, but it does show that fertility would still have declined greatly had there been no family planning policy.

[27]This estimator is biased because women who *actually* had 4 or more births comprise a self-selective sample. For (younger) cohorts without women having 4 or more births, I simply let $E(n|n \geq 4)$ be 4.

6.5 Model without frailty

It is believed that a duration model without individual heterogeneity would bias estimation. Table 9 compares policy effects derived from the model with 3-point frailty and the one without frailty.

[Table 9 is here.]

It turns out that, the two models generate fairly similar policy effects. But this conclusion only holds in the context of this paper, and does not have generalized implications.

7 Conclusion

In many developing countries, family planning programs were accompanied by fertility decline. But evidence for their relations is mixed, including the most populous and largest developing country, China.

Studies that explored the relation between China's family planning policy and fertility usually have two defects. First, most of them are static analyses which can hardly reveal dynamic policy effects on fertility. Second, their policy variables were improperly constructed.

To address the first issue, this paper applies a multiple spell duration model which is built on a mixed proportional hazard function. The function describes non-monotonic duration dependence for fertility behavior, takes into account the effect of family planning policy and various individual characteristics, and includes unobserved individual heterogeneity which is non-parametrically specified, as suggested by Heckman and Singer (1984).

To address the second issue, this paper creates policy variables by utilizing policy's secular and cross-sectional variations. The policy measures mirror all policy periods and reflect policy differences by residential location and ethnicity, so that they are more complete than existing measures. They are built on exogenous variables and characterize women's heterogeneous policy exposure by age, which overcomes the shortcomings of existing measures that might be endogenously created and might lack heterogeneity when illustrating policy exposure.

This paper uses a panel sample of birth history of ever-married women, from the China Health and Nutrition Survey. Estimated results show that, being exposed to the one-child policy reduces the probability of having exactly 2 and 3 births by 31.1% and 35.3%, respectively, and correspondingly increases the probability of childlessness and having exactly 1 birth by 54.9% and 67.0%. Earlier periods of policy present analogous patterns, but lesser effects. Moreover, such patterns of probability shift hold for all birth cohorts of women. In addition, had been there no family planning policy, fertility levels would increase for all cohorts, but the tendency of fertility decline over cohorts would not stop. Family planning policy only explains about one third of the fertility decline over cohorts.

Other individual characteristics have also shown noticeable impact. Better-educated women tend to substantially decrease their likelihood of childbearing. If women's first birth is a son, they are much less likely to have a large number of births, which manifests the strong son preference in China.

As family planning policy did not explain much of the fertility decline over cohorts, further studies need to explore the effects of other variables on fertility transition, particularly the contribution of rising education of women over cohorts.

References

Ahn, Namkee (1994). "Effects of the one-child family policy on second and third births in Hebei, Shaanxi and Shanghai". In: *Journal of Population Economics* 7.1, pp. 63–78.

Angeles, Gustavo, David K Guilkey, and Thomas A Mroz (2005). "The effects of education and family planning programs on fertility in Indonesia". In: *Economic Development and Cultural Change* 54.1, pp. 165–201.

Arroyo, Cristino R and Junsen Zhang (1997). "Dynamic microeconomic models of fertility choice: A survey". In: *Journal of Population Economics* 10.1, pp. 23–65.

Baker, Michael and Angelo Melino (2000). "Duration dependence and nonparametric heterogeneity: A Monte Carlo study". In: *Journal of Econometrics* 96.2, pp. 357–393.

Banerjee, Abhijit, Xin Meng, and Nancy Qian (2010). "The Life Cycle Model and Household Savings: Micro Evidence from Urban China". Working paper.

Bongaarts, John, W Parker Mauldin, and James F Phillips (1990). "The demographic impact of family planning programs". In: *Studies in Family Planning* 21.6, pp. 299–310.

Cai, Yong (2010). "China's Below-Replacement Fertility: Government Policy or Socioeconomic Development?" In: *Population and Development Review* 36.3, pp. 419–440.

Das Gupta, Monica et al. (2003). "Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea". In: *The Journal of Development Studies* 40.2, pp. 153–187.

Edlund, Lena et al. (2008). "More Men, More Crime: Evidence from China's One-Child Policy". Working paper.

Ham, John C and Samuel A Rea Jr (1987). "Unemployment insurance and male unemployment duration in Canada". In: *Journal of Labor Economics*, pp. 325–353.

Hashemi, Ali and Djavad Salehi-Isfahani (2013). "From health service delivery to family planning: the changing impact of health clinics on fertility in rural Iran". In: *Economic Development and Cultural Change* 61.2, pp. 281–309.

Heckman, James and Burton Singer (1984). "A method for minimizing the impact of distributional assumptions in econometric models for duration data". In: *Econometrica: Journal of the Econometric Society*, pp. 271–320.

Hotz, V Joseph, Jacob Alex Klerman, and Robert J Willis (1997). "The economics of fertility in developed countries". In: *Handbook of population and family economics* 1, pp. 275–347.

Islam, Asadul and Russell Smyth (2010). "Children and Parental Health: Evidence from China". Working paper.

Jensen, Robert Todd and Bazaar Chintan (2003). *Equal Treatment, Unequal Outcomes?: Generating Sex Inequality Through Fertility Behaviour*. Bazaar Chintan.

Katz, Lawrence F and Bruce D Meyer (1990). "The impact of the potential duration of unemployment benefits on the duration of unemployment". In: *Journal of public economics* 41.1, pp. 45–72.

Lapham, Robert J and W Parker Mauldin (1985). “Contraceptive prevalence: the influence of organized family planning programs”. In: *Studies in Family Planning* 16.3, pp. 117–137.
Lavely, William and Ronald Freedman (1990). “The Origins of the Chinese Fertility Decline”. In: *Demography* 27.3, pp. 357–367.
Li, Hongbin and Junsen Zhang (2007). “Do High Birth Rates Hamper Economic Growth?” In: *Review of Economics and Statistics* 89.1, pp. 110–117.
— (2008). “Testing the External Effect of Household Behavior: The Case of the Demand for Children”. In: *Journal of Human Resources* 44.4.
Li, Hongbin, Junsen Zhang, and Yi Zhu (2005). “The Effect of the One-Child Policy on Fertility in China: Identification Based on the Differences-in-Differences”. Working paper.
Li, Lei and Minja Kim Choe (1997). “A mixture model for duration data: Analysis of second births in China”. In: *Demography* 34.2, pp. 189–197.
McElroy, Marjorie and Dennis Tao Yang (2000). “Carrots and Sticks: Fertility Effects of China’s Population Policies”. In: *American Economic Review* 90.2, pp. 389–392.
Mwaikambo, Lisa et al. (2011). “What works in family planning interventions: a systematic review”. In: *Studies in family planning* 42.2, pp. 67–82.
Narayan, Paresh Kumar and Xiujian Peng (2006). “An econometric analysis of the determinants of fertility for China, 1952–2000”. In: *Journal of Chinese economic and business studies* 4.2, pp. 165–183.
Newman, John L and Charles E McCulloch (1984). “A hazard rate approach to the timing of births”. In: *Econometrica: journal of the Econometric Society*, pp. 939–961.
Olsen, Randall J and Kenneth I Wolpin (1983). “The impact of exogenous child mortality on fertility: a waiting time regression with dynamic regressors”. In: *Econometrica: Journal of the Econometric Society*, pp. 731–749.
Poston, Dudley L (2002). “Son preference and fertility in China”. In: *Journal of Biosocial Science* 34.3, pp. 333–347.
Poston, Dudley L and Baochang Gu (1987). “Socioeconomic development, family planning, and fertility in China”. In: *Demography* 24.4, pp. 531–551.
Qian, Nancy (2009). “Quantity-Quality and the One Child Policy: The Only-Child Disadvantage in School Enrollment in Rural China”. Working paper.
Schultz, T. Paul and Yi Zeng (1995). “Fertility of Rural China: Effects of Local Family Planning and Health Programs”. In: *Journal of Population Economics* 8.4, pp. 329–350.
Szreter, Simon (1993). “The idea of demographic transition and the study of fertility change: a critical intellectual history”. In: *Population and development review*, pp. 659–701.
Van den Berg, Gerard J (2001). “Duration models: specification, identification and multiple durations”. In: *Handbook of econometrics* 5, pp. 3381–3460.
Wang, Fei (2012). “Family Planing Policy in China: Measurement and Impact on Fertility”. Working paper.

Wang, Feng (1988). “The roles of individuals’ socioeconomic characteristics and the government family planning program in China’s fertility decline”. In: *Population Research and Policy Review* 7.3, pp. 255–276.

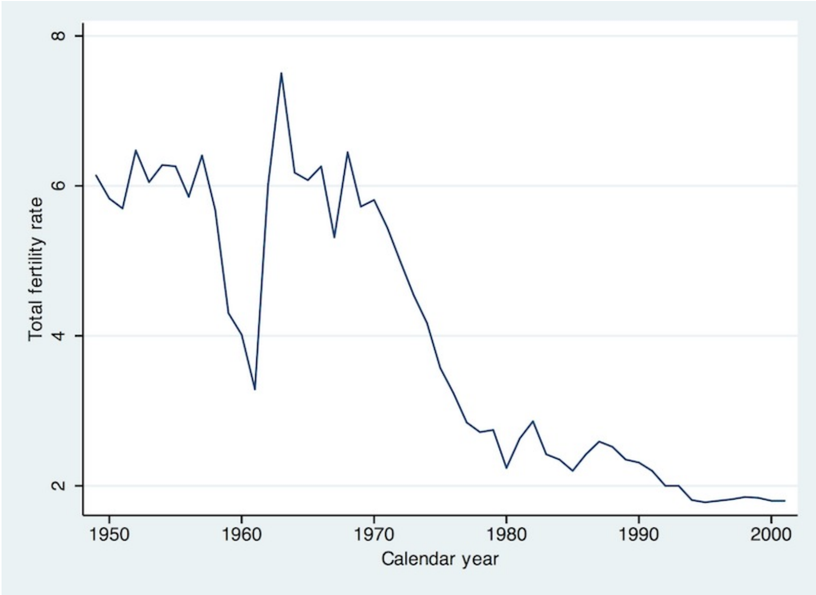
Wooldridge, Jeffrey M (2010). *Econometric analysis of cross section and panel data*. The MIT press.

Wu, Xiaoyu and Lixing Li (2011). “Family Size and Maternal Health: Evidence from the One-Child Policy in China”. In: *Journal of Population Economics*.

Yang, D. T. and D. Chen (2004). “Transformations in China’s Population Policies”. In: *Pacific Economic Review* 9.3, pp. 269–290.

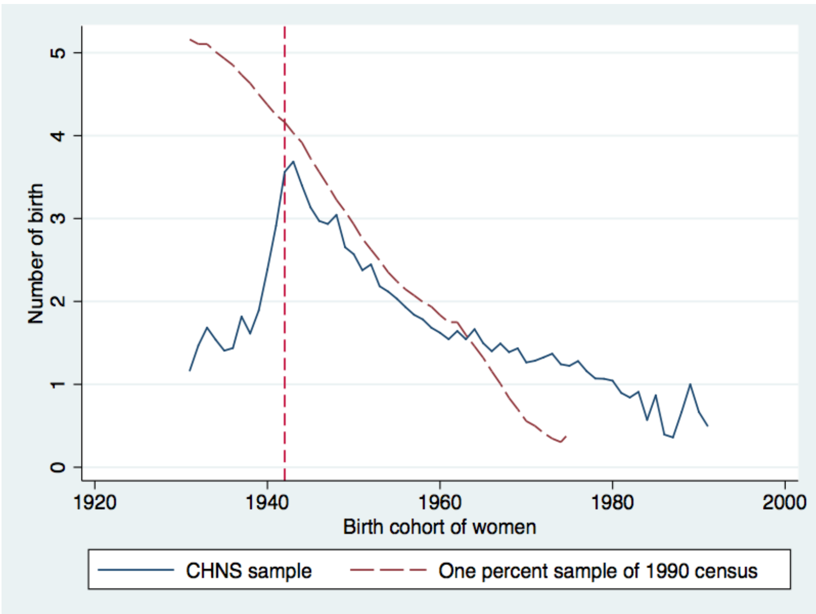
Yang, Faxiang (2004). “Historical Research on Family Planning of Contemporary China”. PhD thesis. Zhejiang University.

Figure 1: Total fertility rates in China, 1949–2001



Notes: Plotted based on the data recorded in Yang (2004, pp. 264–265). Data originally came from various national surveys conducted by China's statistical authorities.

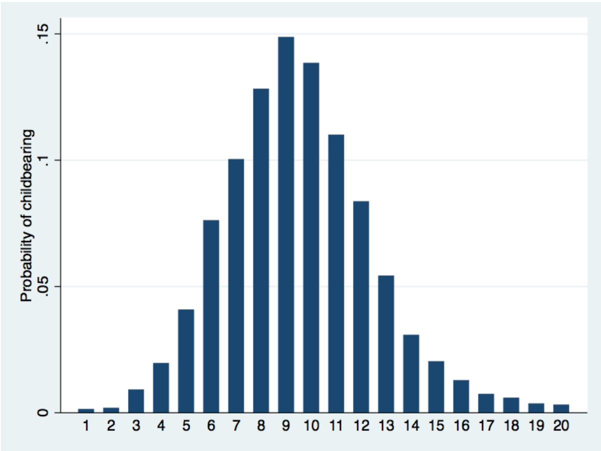
Figure 2: Number of birth, by birth cohort of ever-married women, CHNS versus 1990 census



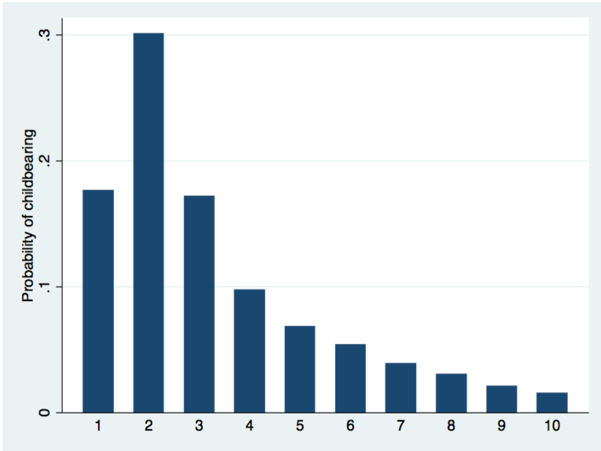
Notes: The solid curve was plotted based on the CHNS sample, with 6533 ever-married women born in 1931–1991. The dashed curve was plotted based on one percent sample of China's 1990 census data, with 2.67 million ever-married women born in 1931–1975. The CHNS sample starts to match the downward fertility trend of the census sample from cohort 1942, marked with a vertical dashed line.

Figure 3: Actual probability of childbearing, by birth spell and duration

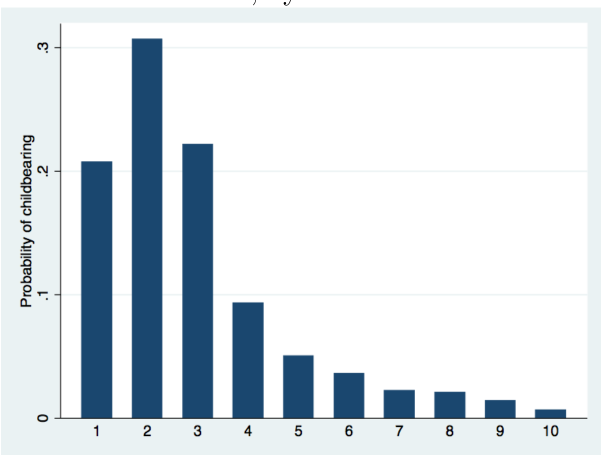
(a) Actual probability of having the first birth, by duration



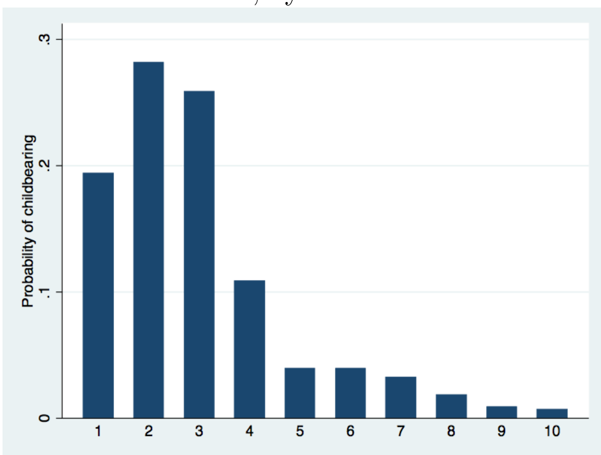
(b) Actual probability of having the second birth, by duration



(c) Actual probability of having the third birth, by duration

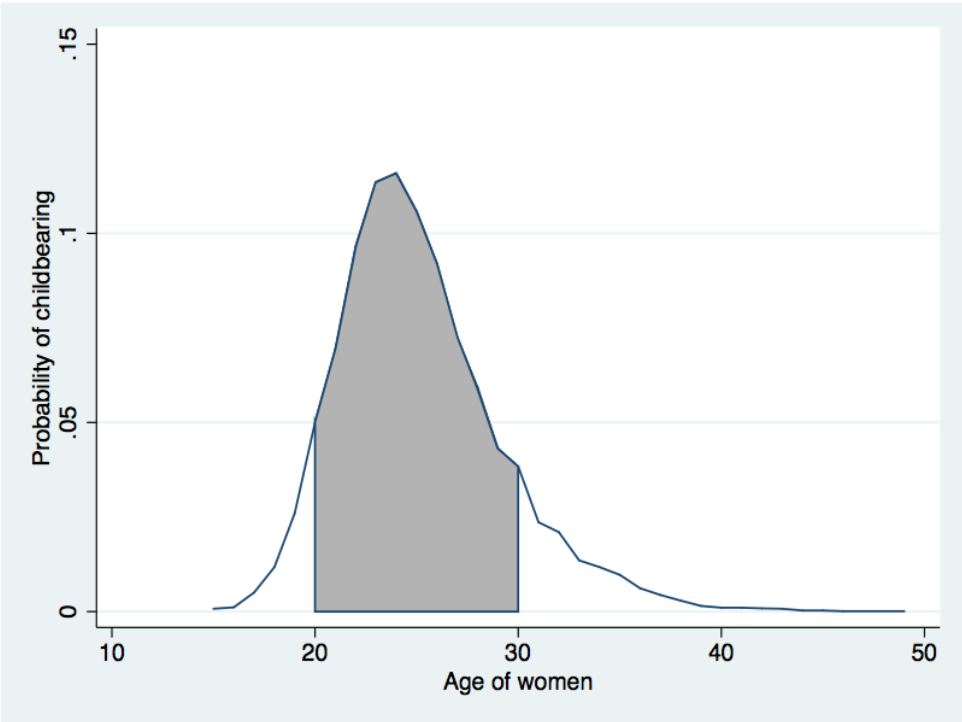


(d) Actual probability of having the fourth birth, by duration



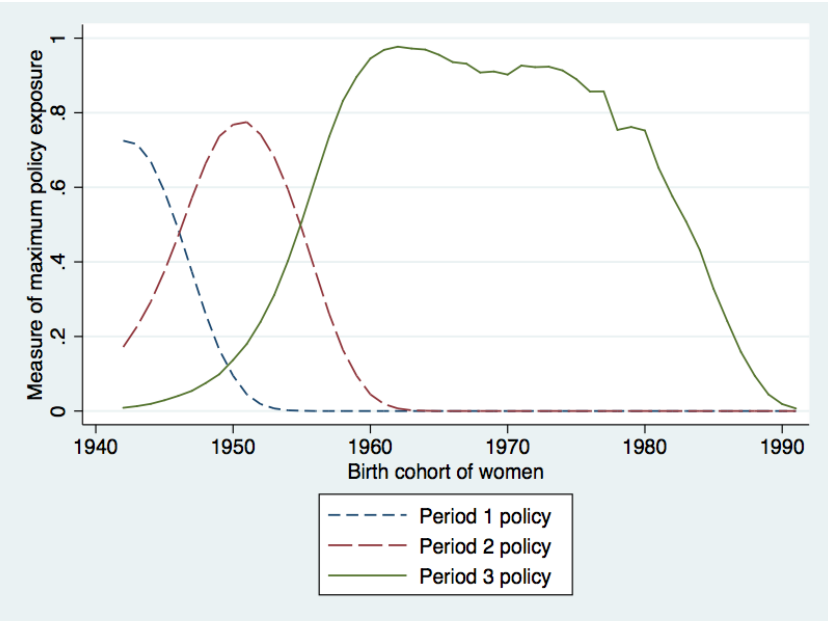
Notes: Figures (a)–(d) are plotted based on all women, women with at least 1 birth, women with at least 2 births and women with at least 3 births. X-axes are duration in years. For figure (a) results after the twentieth year are not displayed; for other three figures, results after the tenth year are omitted.

Figure 4: Probability of childbearing, by age of ever-married women



Notes: A point on the solid curve represents the proportion of women who had births at the corresponding age. The curve was plotted based on the CHNS ever-married women of birth cohorts 1942–1991, and shows positive probabilities for age 15–49. The shaded area, bordering on age 20 and 30, gives an example for a woman's policy exposure if she was 20 and 30 years old when a policy started and ended, respectively. When the woman's age was between 20 and 30, her policy exposure is measured by the shaded area between 20 and her age; when the woman's age was 30 or above, her policy exposure is measured by the entire shaded area.

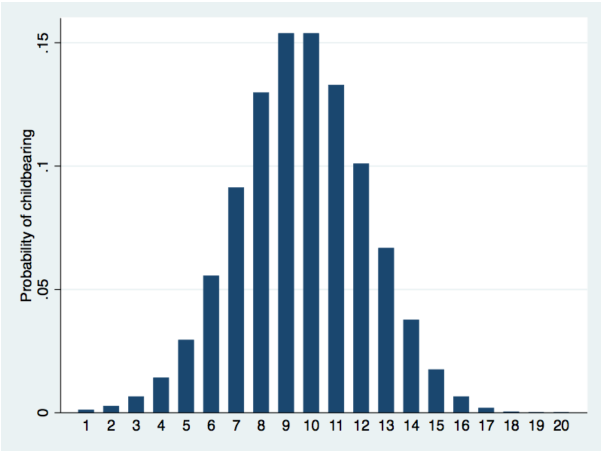
Figure 5: Measures of maximum exposure to each period of policy, by birth cohort of ever-married women



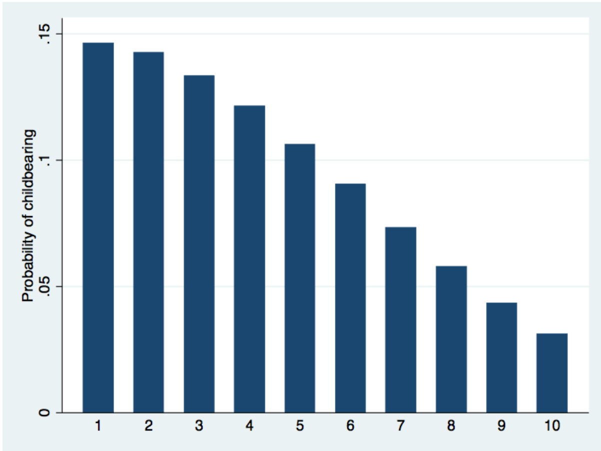
Notes: A woman's maximum exposure to a policy represents the accumulation of *all* her past exposure to the policy. The three curves illustrate the cohort average of maximum exposure to the three periods of policies. Cohorts around 1940, 1950 and 1960 were exposed to the period 1, period 2 and period 3 policies most, respectively.

Figure 6: Predicted probability of childbearing, by birth spell and duration

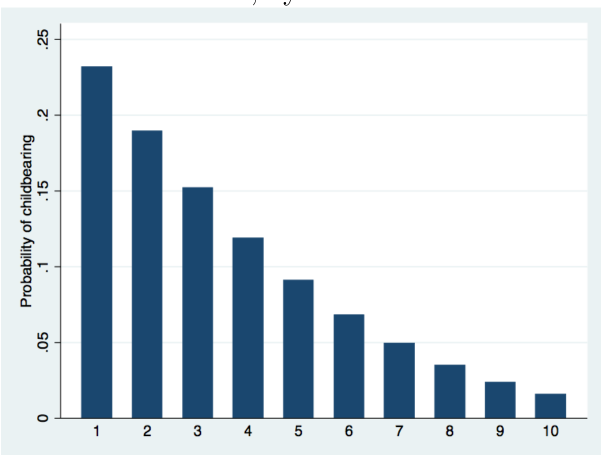
(a) Predicted probability of having the first birth, by duration



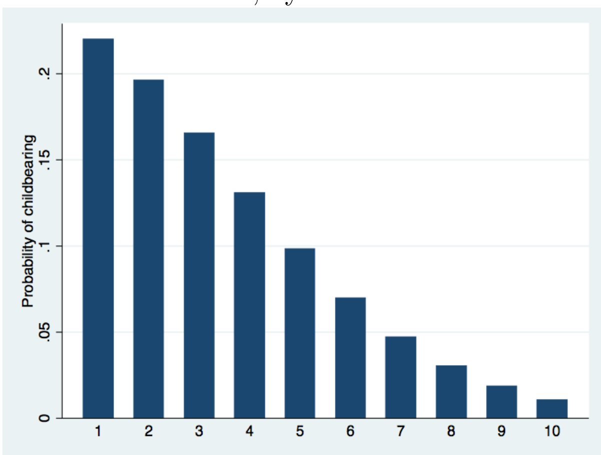
(b) Predicted probability of having the second birth, by duration



(c) Predicted probability of having the third birth, by duration

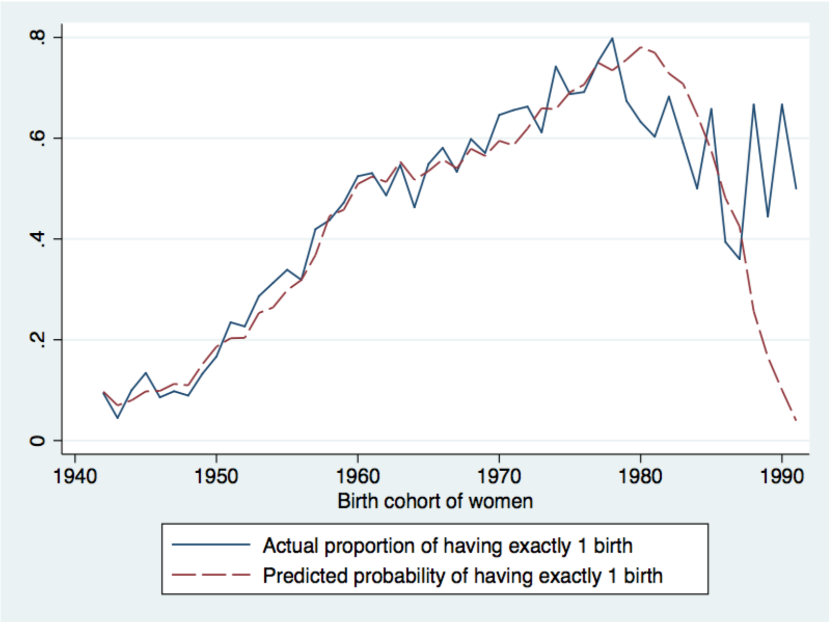


(d) Predicted probability of having the fourth birth, by duration



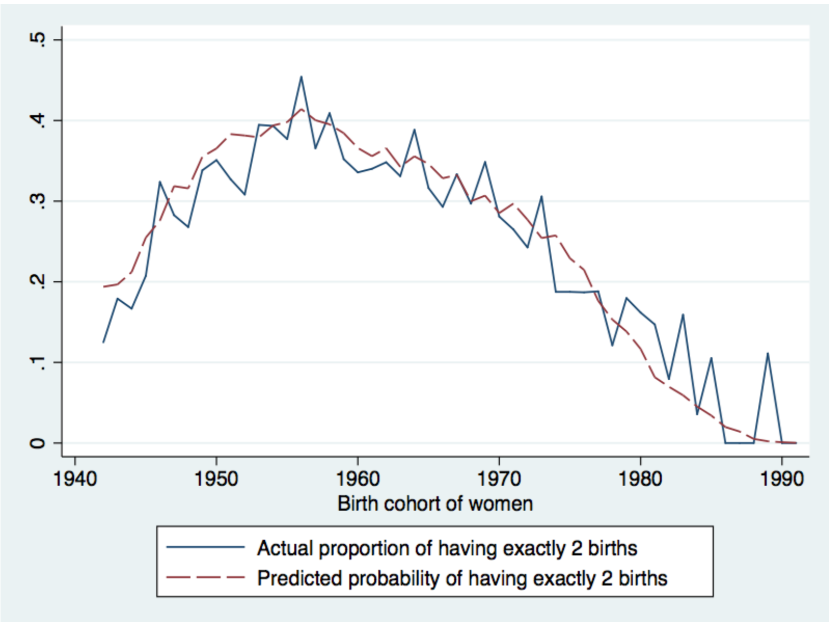
Notes: Figures (a)–(d) are predicted based on all women, women with at least 1 birth, women with at least 2 births and women with at least 3 births. X-axes are duration in years. For figure (a) results after the twentieth year are not displayed; for other three figures, results after the tenth year are omitted.

Figure 7: Actual proportion and predicted probability of having exactly 1 birth, by birth cohort of ever-married women



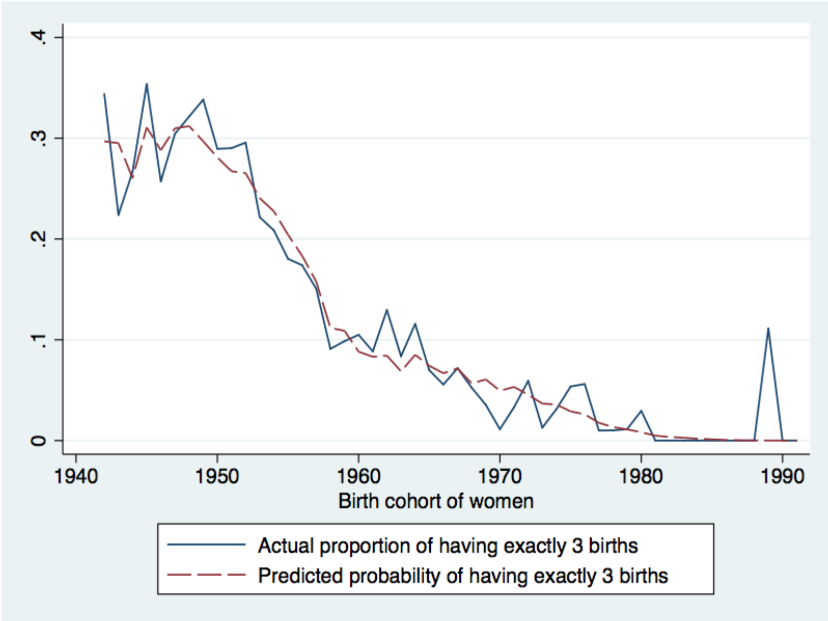
Notes: The solid curve indicates the actual proportion of having exactly 1 birth, by the survey year, for each cohort of women. The dashed curve represents the average probability of having exactly 1 birth, by the survey year, for each cohort of women.

Figure 8: Actual proportion and predicted probability of having exactly 2 births, by birth cohort of ever-married women



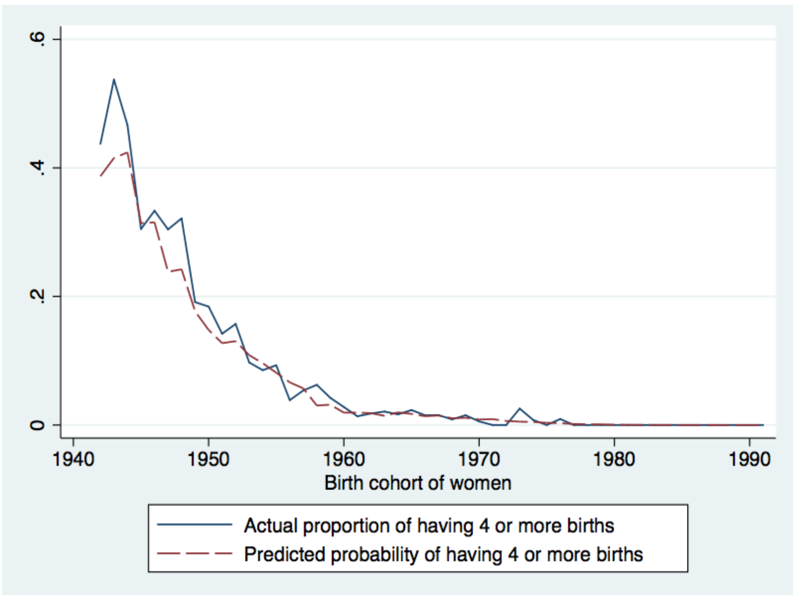
Notes: The solid curve indicates the actual proportion of having exactly 2 births, by the survey year, for each cohort of women. The dashed curve represents the average probability of having exactly 2 births, by the survey year, for each cohort of women.

Figure 9: Actual proportion and predicted probability of having exactly 3 births, by birth cohort of ever-married women



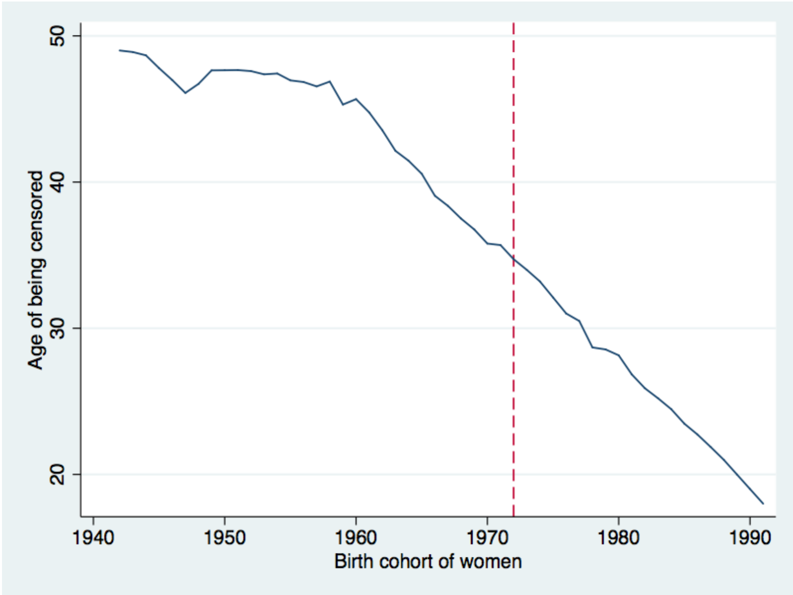
Notes: The solid curve indicates the actual proportion of having exactly 3 births, by the survey year, for each cohort of women. The dashed curve represents the average probability of having exactly 3 births, by the survey year, for each cohort of women.

Figure 10: Actual proportion and predicted probability of having 4 or more births, by birth cohort of ever-married women



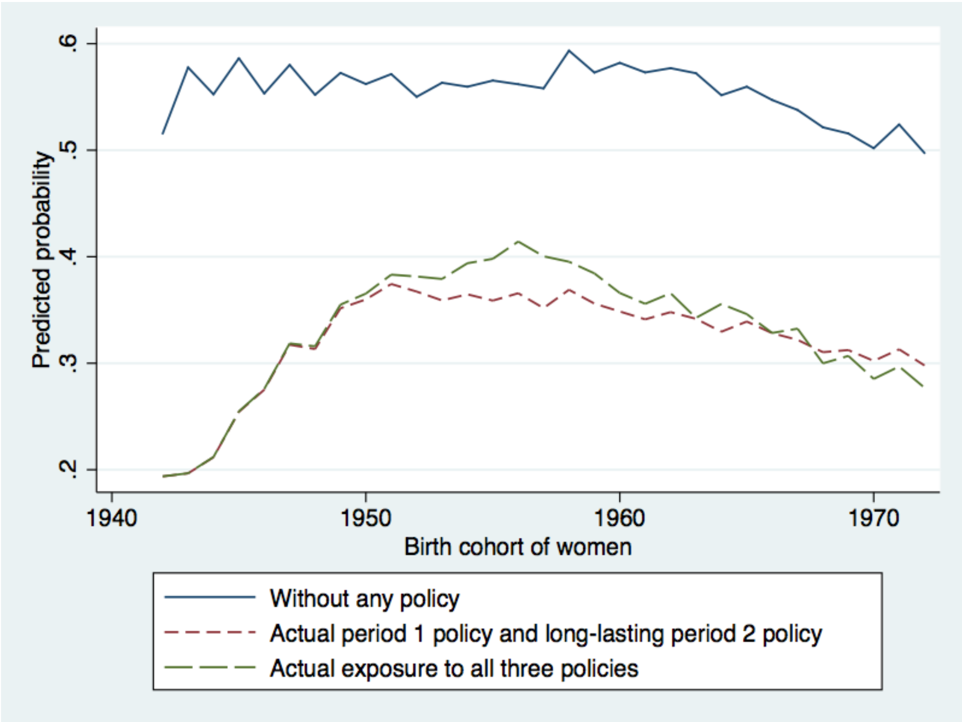
Notes: The solid curve indicates the actual proportion of having 4 or more births, by the survey year, for each cohort of women. The dashed curve represents the average probability of having 4 or more births, by the survey year, for each cohort of women.

Figure 11: Age of being censored, by birth cohort of ever-married women



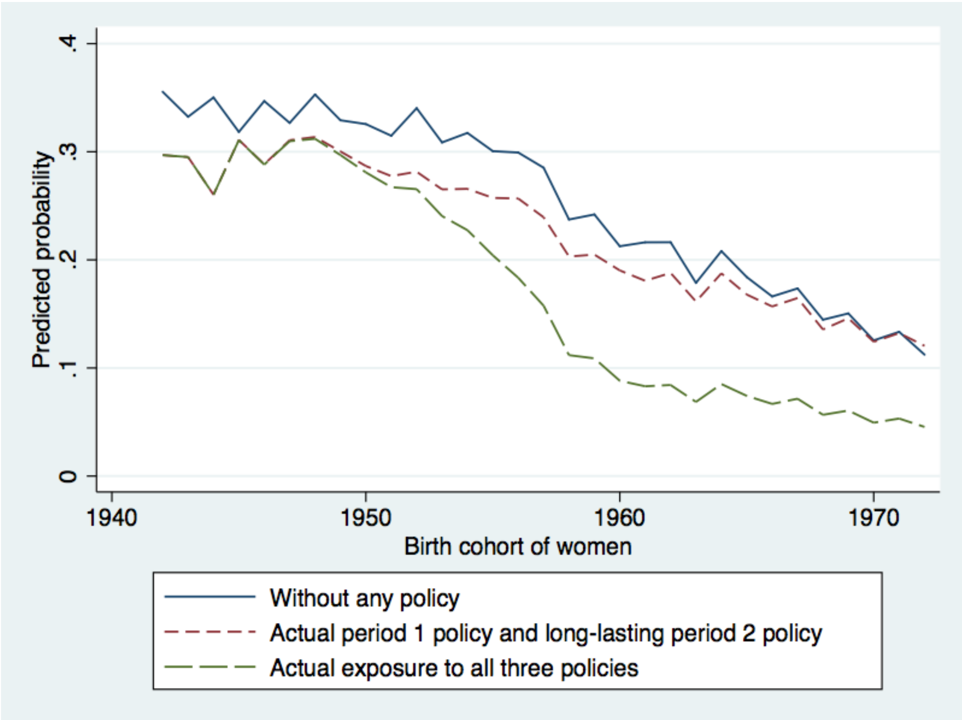
Notes: The solid curve shows the cohort average of age at which a woman exits the sample. Cohort 1972, marked with a dashed line, or older cohorts were censored at age 35 or above. These cohorts are assumed to have physiologically completed childbearing, and are thus included for counterfactual simulations, according to the fact that the percentage of women who had births at 35 or any age above is lower than 0.01, as Figure 4 implies.

Figure 12: Predicted probability of having exactly 2 births under various scenarios of policy exposure, by birth cohort of ever-married women



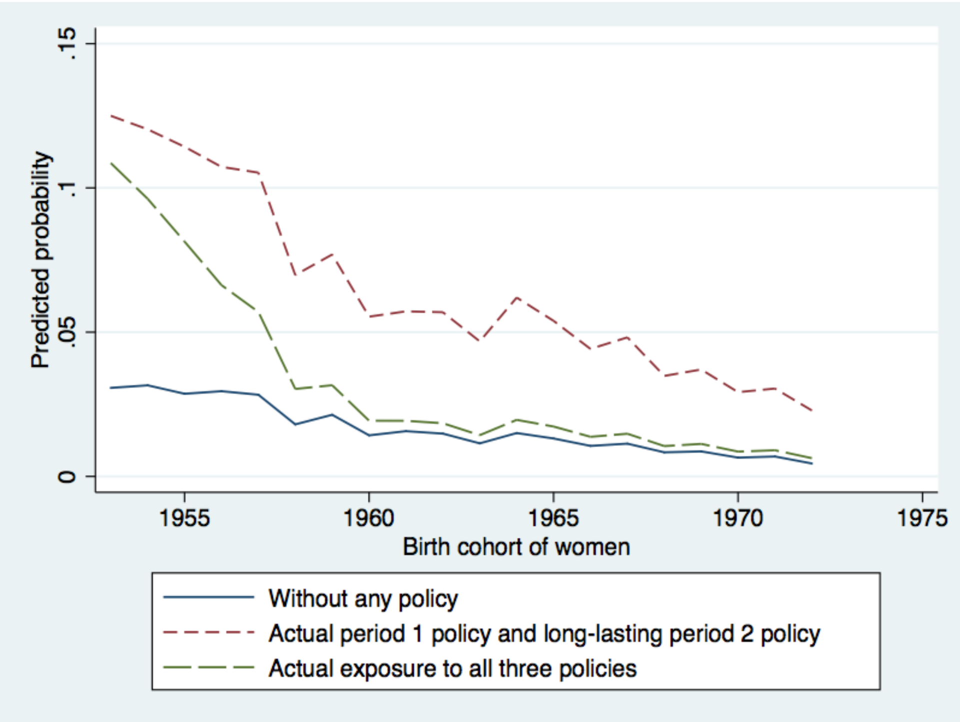
Notes: The solid curve indicates the predicted probability of having exactly 2 births had women *not* been exposed to any policy. The short-dashed curve represents the predicted probability of having exactly 2 births had women been exposed to period 1 and 2 policies. The period 1 policy is the actual policy, while the period 2 policy is assumed to be effective all the time after it started. The long-dashed curve illustrates the predicted probability of having exactly 2 births under actual exposure to all three periods of policies.

Figure 13: Predicted probability of having exactly 3 births under various scenarios of policy exposure, by birth cohort of ever-married women



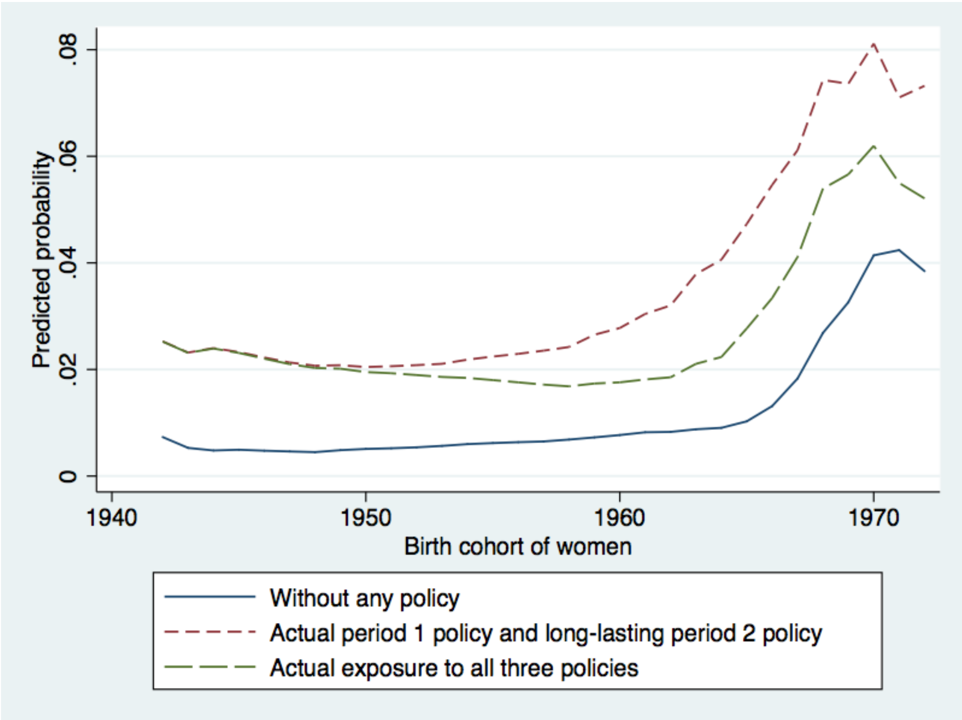
Notes: The solid curve indicates the predicted probability of having exactly 3 births had women *not* been exposed to any policy. The short-dashed curve represents the predicted probability of having exactly 3 births had women been exposed to period 1 and 2 policies. The period 1 policy is the actual policy, while the period 2 policy is assumed to be effective all the time after it started. The long-dashed curve illustrates the predicted probability of having exactly 3 births under actual exposure to all three periods of policies.

Figure 14: Predicted probability of having 4 or more births under various scenarios of policy exposure, by birth cohort of ever-married women



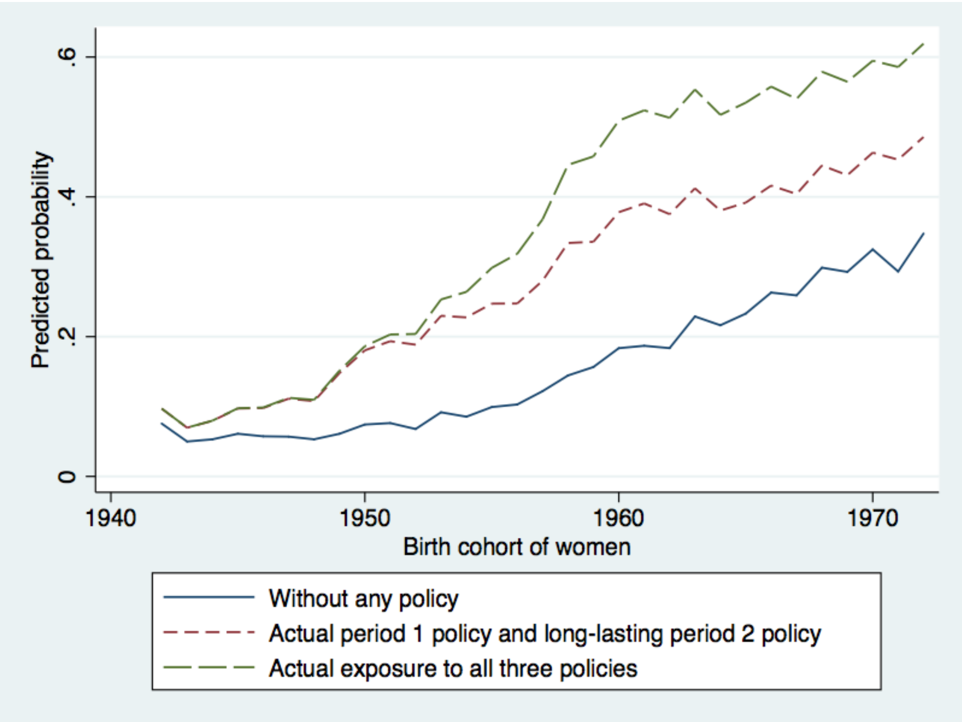
Notes: The solid curve indicates the predicted probability of having 4 or more births had women *not* been exposed to any policy. The short-dashed curve represents the predicted probability of having 4 or more births had women been exposed to period 1 and 2 policies. The period 1 policy is the actual policy, while the period 2 policy is assumed to be effective all the time after it started. The long-dashed curve illustrates the predicted probability of having 4 or more births under actual exposure to all three periods of policies.

Figure 15: Predicted probability of childlessness under various scenarios of policy exposure, by birth cohort of ever-married women



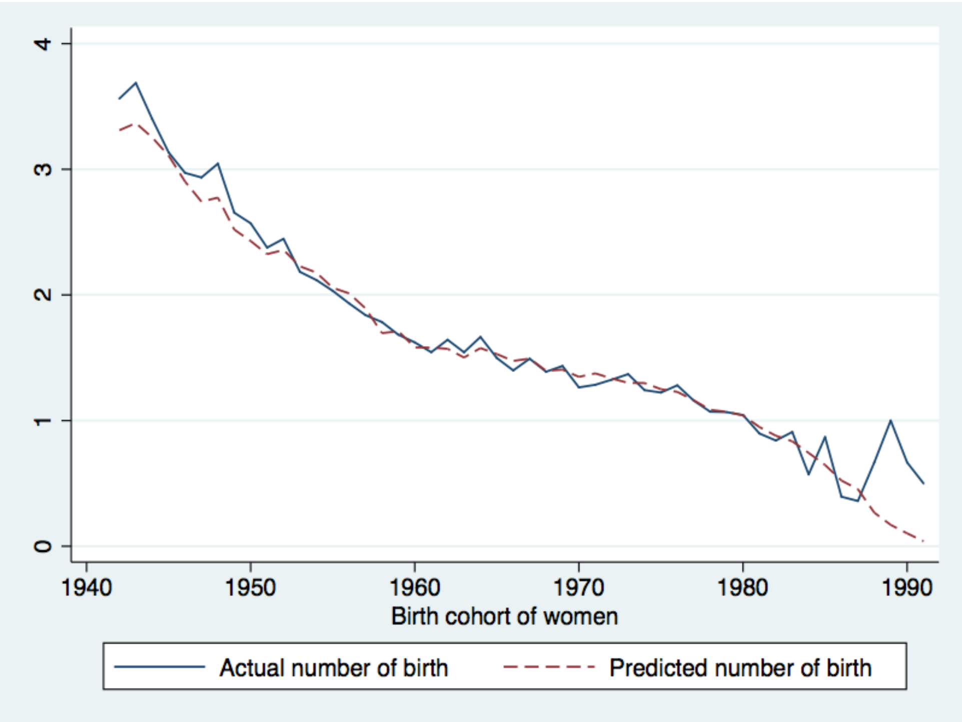
Notes: The solid curve indicates the predicted probability of childlessness had women *not* been exposed to any policy. The short-dashed curve represents the predicted probability of childlessness had women been exposed to period 1 and 2 policies. The period 1 policy is the actual policy, while the period 2 policy is assumed to be effective all the time after it started. The long-dashed curve illustrates the predicted probability of childlessness under actual exposure to all three periods of policies.

Figure 16: Predicted probability of having exactly 1 birth under various scenarios of policy exposure, by birth cohort of ever-married women



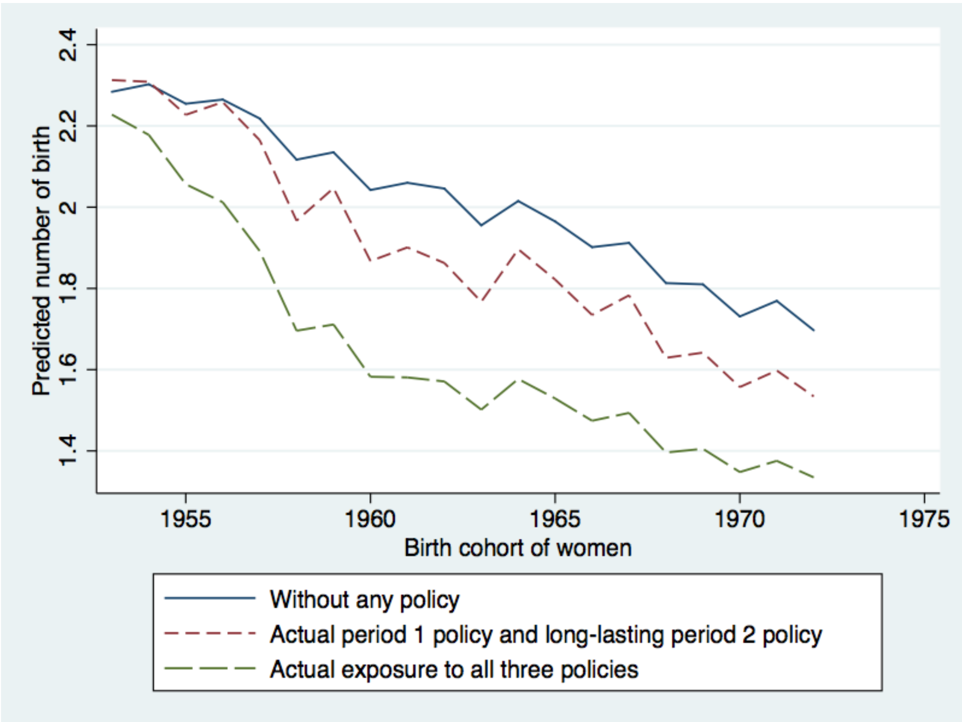
Notes: The solid curve indicates the predicted probability of having exactly 1 birth had women *not* been exposed to any policy. The short-dashed curve represents the predicted probability of having exactly 1 birth had women been exposed to period 1 and 2 policies. The period 1 policy is the actual policy, while the period 2 policy is assumed to be effective all the time after it started. The long-dashed curve illustrates the predicted probability of having exactly 1 birth under actual exposure to all three periods of policies.

Figure 17: Actual and predicted number of birth, by birth cohort of ever-married women



Notes: Predicted number of birth is obtained from $\hat{P}_1 + 2\hat{P}_2 + 3\hat{P}_3 + \hat{n}_{\geq 4}\hat{P}_{\geq 4}$, where $\hat{P}$'s are predicted probabilities of having a certain number of births, and $\hat{n}_{\geq 4}$ is the average number of birth for each cohort conditioning on the women who had 4 or more births.

Figure 18: Predicted number of birth under various scenarios of policy exposure, by birth cohort of ever-married women



Notes: The solid curve indicates the predicted number of birth had women *not* been exposed to any policy. The short-dashed curve represents the predicted number of birth had women been exposed to period 1 and 2 policies. The period 1 policy is the actual policy, while the period 2 policy is assumed to be effective all the time after it started. The long-dashed curve illustrates the predicted number of birth under actual exposure to all three policies.

Table 1: Number of children allowed per married couple, by policy period, residential location, and ethnicity[a]

	Period 1 policy[b]	Period 2 policy[b]	Period 3 policy[b]
Urban Han	Mild policy allowed three children, though not encouraged [c]	Strong policy allowed two children, though not encouraged	One-child policy allows only one child
Rural Han	No restriction	Milder policy than the upper cell	One-child policy conditionally allows two children
Urban non-Han	No restriction[d]	No restriction[d]	Specific policy conditionally allows two children
Rural non-Han	No restriction	No restriction[d]	Specific policy conditionally allows three or even more children

Notes:
[a] The table was summarized based on the policy history that Wang (2012) introduced in Section 2.
[b] Period 1 to 3 policies were effective in 1963–1970, 1971–1979, and 1980–present, respectively.
[c] The urban Han people living in the five autonomous regions were not covered by the period 1 policy.
[d] Part of the non-Han people might be affected by family planning policy, particularly when they were living outside the five autonomous regions, and/or identified the ethnicity of their children as Han.

Table 2: Number and percentage of ever-married women with a certain number of birth spells

	Number of women	%
1 birth spell	321	5.2
2 birth spells	2842	45.7
3 birth spells	1846	29.7
4 birth spells	764	12.3
5 birth spells	282	4.5
6 or more birth spells	159	2.6
Total	6214	100

Notes: Women with 1 spell were either childless, or had exactly 1 birth and the birth occurred in her last year in the sample; women with 2 spells either had exactly 1 birth, or had exactly 2 births and the second birth occurred in her last year; so on and so forth. As the percentage of women who had births in their last year in the sample is only 1.01%, women with exactly j spells are roughly equivalent to those having exactly $j-1$ births ($j = 1, 2, ...$).

Table 3: Descriptive statistics of selected variables, by birth spells

	Spell 1	Spell 2	Spell 3	Spell 4
End up with newborns (%)	94.2	51.0	39.0	36.0
Average duration	9.6	3.5	3.1	3.1
Variance of duration	9.8	7.0	5.7	4.8
Urban (%)	34.6	34.4	24.3	19.5
Han (%)	87.9	87.9	85.5	82.0
Years of schooling	8.2	8.2	6.7	5.4
Coastal province (%)	34.0	34.1	26.5	20.6
Born in 1942-50 (%)	13.4	14.1	24.2	42.8
Born in 1951-60 (%)	28.2	29.5	36.6	38.2
Born in 1961-70 (%)	34.0	34.5	28.9	15.5
Born in 1971-80 (%)	19.0	18.6	9.5	3.4
Born in 1981-91 (%)	5.4	3.4	0.8	0.1
First birth is son (%)		51.7	46.0	43.8
Number of women	6214	5884	3048	1203

Notes: Statistics for spell j are calculated based on the women who had *at least* j spells ($j = 1, 2, 3, 4$). Mean and variance of duration for a spell are calculated based on the women who ended up the spell with newborns. *Urban* is the residential location. *Years of schooling* indicates completed schooling. *Coastal province* includes Jiangsu, Liaoning, and Shandong. *First birth is son (%)* is not available for spell 1 because most women with exactly 1 spell ended up with no birth.

Table 4: Effects on logarithm of hazard rate of childbearing

	Logarithm of hazard rate of having the			
	first birth	second birth	third birth	fourth birth
Period-1 policy	-4.484	-0.493	1.518	4.566
	(0.595)***	(0.635)	(1.007)	(1.992)**
Urban × Period-1 policy	0.233	0.525	-0.717	-1.287
	(0.582)	(0.408)	(0.555)	(1.488)
Han × Period-1 policy	0.490	0.272	0.949	-0.570
	(0.658)	(0.476)	(0.613)	(1.310)
Period-2 policy	-3.395	0.056	1.335	2.722
	(0.444)***	(0.502)	(0.890)	(1.652)*
Urban × Period-2 policy	0.420	-0.891	-0.801	-0.940
	(0.222)*	(0.226)***	(0.320)**	(0.824)
Han × Period-2 policy	-0.227	-0.599	-0.409	-0.425
	(0.358)	(0.293)**	(0.363)	(0.749)
Period-3 policy	-3.026	-0.842	-0.331	1.831
	(0.433)***	(0.489)*	(0.923)	(2.110)
Urban × Period-3 policy	0.618	-0.841	-0.322	-0.800
	(0.251)**	(0.195)***	(0.410)	(1.106)
Han × Period-3 policy	0.546	-0.608	-0.143	-0.344
	(0.322)*	(0.256)**	(0.413)	(1.352)
Urban	-0.354	-0.276	-0.048	-0.078
	(0.115)***	(0.172)	(0.335)	(1.046)
Han	0.152	-0.054	-0.165	0.270
	(0.155)	(0.206)	(0.382)	(1.016)
Primary school	-0.112	-0.224	-0.202	-0.071
	(0.054)**	(0.053)***	(0.074)***	(0.123)
Middle school	-0.329	-0.335	-0.439	-0.028
	(0.047)***	(0.051)***	(0.077)***	(0.137)
High school or above	-0.760	-0.964	-0.345	-1.404
	(0.056)***	(0.076)***	(0.136)**	(0.478)***
Duration	0.983	0.292	0.103	0.286
	(0.030)***	(0.028)***	(0.060)*	(0.114)**
Squared duration	-0.022	-0.022	-0.016	-0.036
	(0.002)***	(0.003)***	(0.006)***	(0.013)***
Age		0.438	-0.092	-0.903
		(0.125)***	(0.226)	(0.384)**
Squared age		-0.009	-0.001	0.010
		(0.002)***	(0.003)	(0.005)*
First birth is son		-0.415	-0.375	-0.241
		(0.038)***	(0.057)***	(0.096)**
Constant	-7.128	-6.740	0.538	11.544
	(0.213)***	(1.805)***	(3.399)	(5.872)**
Chi-squared statistic for 0 policy effect (df = 9)	170.5	119.2	83.9	12.4
P value for 0 policy effect	0.0000	0.0000	0.0000	0.1932
N	157735	157735	157735	157735

Notes: Robust standard errors are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Coefficients can be interpreted as the semi-elasticity of hazard rates to variables. Non-reported variables include exposure to the great famine, regional dummies, cohort trend, squared cohort trend, cohort trend × urban dummy, cohort trend × Han dummy and cohort trend × coastal province dummy. Estimated frailty follows a three-point distribution, with supporting points -5.6879, -1.6396, and 0.21167, and corresponding probabilities 0.0175, 0.0585, and 0.924. Significance of these points is shown in Table 5.

Table 5: Test whether 3-mass-point frailty reduces to no frailty or 2-mass-point frailty

A. Test whether 3-mass-point frailty reduces to no frialty	
Log likelihood of model with 3-point frailty	−28960.401
Log likelihood of model without frailty	−29037.347
Log likelihood ratio statistic	153.892
P value (Chi-squared distribution, $df = 4$)	0.0000
B. Test whether 3-mass-point frailty reduces to 2-mass-point frailty	
Log likelihood of model with 3-point frailty	−28960.401
Log likelihood of model with 2-point frailty	−28969.355
Log likelihood ratio statistic	17.908
P value (Chi-squared distribution, $df = 2$)	0.0001

Table 6: Actual proportion and predicted probability of having a certain number of birth, by characteristics of ever-married women

	Having exactly 1 birth		Having exactly 2 births		Having exactly 3 births		Having 4 or more births		
	actual	predicted	actual	predicted	actual	predicted	actual	predicted	Number of women
Full sample	0.462	0.456	0.299	0.308	0.123	0.123	0.071	0.064	6214
Urban Han	0.618	0.608	0.227	0.247	0.079	0.079	0.023	0.023	1974
Rural Han	0.393	0.389	0.338	0.341	0.139	0.141	0.088	0.078	3487
Urban non-Han	0.418	0.487	0.339	0.284	0.107	0.116	0.085	0.043	177
Rural non-Han	0.358	0.328	0.295	0.329	0.186	0.163	0.130	0.127	576
Middle school or above	0.579	0.570	0.272	0.281	0.068	0.071	0.024	0.022	3713
Below middle school	0.288	0.286	0.339	0.348	0.206	0.199	0.142	0.126	2501
Coastal province	0.574	0.558	0.266	0.278	0.085	0.090	0.033	0.027	2114
Inner province	0.404	0.403	0.316	0.323	0.143	0.140	0.091	0.083	4100
First birth is son	0.544	0.490	0.290	0.308	0.110	0.109	0.057	0.051	3044
First birth is daughter	0.427	0.411	0.341	0.325	0.147	0.144	0.085	0.079	2847

Notes: Actual means the actual proportion of ever-married women that had exactly 1, 2, 3, or 4 or more births in each subsample; *predicted* shows the average probability of having exactly 1, 2, 3, or 4 or more births in each subsample. Actual proportion or predicted probability of childlessness can be obtained simply by subtracting the actual proportions or predicted probabilities for each subsample above from 1.

Table 7: Effects of exposure to each period of policy on the probability of having a certain number of birth

	Period 1 policy	Period 2 policy	Period 3 policy
having exactly 2 births	-0.025	-0.045	-0.139
	[-7.5%]	[-12.6%]	[-31.1%]
having exactly 3 births	-0.007	-0.020	-0.067
	[-5.1%]	[-14.0%]	[-35.3%]
having 4 or more births	0.027	0.025	0.005
	[71.5%]	[65.9%]	[8.5%]
childlessness	0.001	0.003	0.018
	[2.3%]	[5.9%]	[54.9%]
having exactly 1 birth	0.004	0.036	0.183
	[0.9%]	[8.7%]	[67.0%]

Notes: Numbers *not* in squared brackets represent the difference between the probability of having a certain number of birth under a woman's actual exposure to some policy and the probability assuming she had not been exposed to that policy, given all the other variables take their actual values, including other two policies. Numbers in squared brackets are corresponding percentage changes. For example, -0.045 (and -12.6% below), corresponding to *2 births* and *Period 2 policy*, means, exposure to period 2 policy reduced women's probability of having exactly 2 births by 0.045 (and by 12.6%).

Table 8: Effects of exposure to each period of policy on the probability of having a certain number of birth, by residential location and ethnicity

		Period 1 policy	Period 2 policy	Period 3 policy
having exactly 2 births	Urban Han	-0.018	-0.054	-0.203
		[-6.9%]	[-18.1%]	[-45.2%]
	Rural Han	-0.031	-0.038	-0.114
		[-8.3%]	[-10.1%]	[-25.1%]
	Urban non-Han	-0.002	-0.025	-0.142
		[-0.7%]	[-8.2%]	[-33.4%]
	Rural non-Han	-0.019	-0.056	-0.066
		[-5.5%]	[-14.5%]	[-16.8%]
having exactly 3 births	Urban Han	0.004	-0.028	-0.062
		[5.6%]	[-26.2%]	[-43.9%]
	Rural Han	-0.007	-0.015	-0.070
		[-4.8%]	[-9.9%]	[-33.3%]
	Urban non-Han	-0.007	-0.025	-0.086
		[-6.0%]	[-17.6%]	[-42.6%]
	Rural non-Han	-0.040	-0.018	-0.056
		[-19.8%]	[-9.8%]	[-25.6%]
having 4 or more births	Urban Han	0.008	0.004	-0.002
		[56.6%]	[22.9%]	[-6.4%]
	Rural Han	0.034	0.031	0.008
		[76.3%]	[66.8%]	[10.8%]
	Urban non-Han	0.006	0.017	0.002
		[15.2%]	[61.9%]	[4.3%]
	Rural non-Han	0.054	0.065	0.013
		[72.6%]	[105.8%]	[11.2%]
childlessness	Urban Han	0.001	0.003	0.015
		[2.7%]	[6.9%]	[52.2%]
	Rural Han	0.001	0.003	0.017
		[2.0%]	[5.5%]	[51.8%]
	Urban non-Han	0.001	0.002	0.033
		[0.9%]	[3.2%]	[88.6%]
	Rural non-Han	0.002	0.003	0.022
		[3.3%]	[6.4%]	[72.1%]
having exactly 1 birth	Urban Han	0.005	0.075	0.252
		[0.8%]	[14.1%]	[70.5%]
	Rural Han	0.003	0.020	0.159
		[0.9%]	[5.3%]	[69.3%]
	Urban non-Han	0.003	0.031	0.193
		[0.6%]	[6.9%]	[65.9%]
	Rural non-Han	0.004	0.005	0.087
		[1.3%]	[1.5%]	[36.4%]

Notes: Numbers *not* in squared brackets represent the difference between the probability of having a certain number of birth under a woman's actual exposure to some policy and the probability assuming she had not been exposed to that policy, given all the other variables take their actual values, including other two policies. Numbers in squared brackets are corresponding percentage changes. For example, -0.038 (and -10.1% below), corresponding to *2 births*, *Rural Han* and *Period 2 policy*, means, exposure to period 2 policy reduced rural Han women's probability of having exactly 2 births by 0.038 (and by 10.1%).

Table 9: Effects of exposure to each period of policy on the probability of having a certain number of birth, no frailty versus 3-point frailty

		Period 1 policy	Period 2 policy	Period 3 policy
having exactly 2 births	no frailty	-0.028	-0.048	-0.140
		[-8.2%]	[-13.5%]	[-31.2%]
	3 mass points	-0.025	-0.045	-0.139
		[-7.5%]	[-12.6%]	[-31.1%]
having exactly 3 births	no frailty	-0.007	-0.019	-0.067
		[-5.4%]	[-13.2%]	[-34.9%]
	3 mass points	-0.007	-0.020	-0.067
		[-5.1%]	[-14.0%]	[-35.3%]
having 4 or more births	no frailty	0.028	0.027	0.005
		[73.5%]	[69.4%]	[9.0%]
	3 mass points	0.027	0.025	0.005
		[71.5%]	[65.9%]	[8.5%]
childlessness	no frailty	0.004	0.007	0.027
		[7.1%]	[14.8%]	[106.3%]
	3 mass points	0.001	0.003	0.018
		[2.3%]	[5.9%]	[54.9%]
having exactly 1 birth	no frailty	0.004	0.034	0.174
		[0.8%]	[8.2%]	[63.2%]
	3 mass points	0.004	0.036	0.183
		[0.9%]	[8.7%]	[67.0%]

Notes: Numbers *not* in squared brackets represent the difference between the probability of having a certain number of birth under a woman's actual exposure to some policy and the probability assuming she had not been exposed to that policy, given all the other variables take their actual values, including other two policies. Numbers in squared brackets are corresponding percentage changes.

A Procedure for estimating mass-point distributed frailty

One can start with a proportional hazard model without frailty, and then use estimated coefficients as initial values for estimating the model with a two-mass-point distributed frailty.

One can arbitrarily choose initial values for the locations and probabilities of the two mass points, such that the expectation of frailty is 0. In this paper, the initial locations are -1 and $exp(0.4)$, with probabilities $\frac{exp(0.4)}{1+exp(0.4)}$ and $\frac{1}{1+exp(0.4)}$, respectively.

After the model with M-mass-point ($M \geq 2$) distributed frailty converges, search for the location of the $(M+1)$'th point from a starting location to an ending location step by step. The range of searching region should be large enough to include possible locations of the new point. In this paper, the region is set to be from -100 to 100, with step length 0.01. Then, place a small mass (for example, 0.05) onto each point within the searching region. If the maximum increase in likelihood is greater than 0, the location corresponding to this maximum is used as the initial value of the new location, otherwise the procedure stops as adding a new point does not improve likelihood.

In this paper, supporting points of the two-mass-point frailty are estimated to be -2.4258 and 0.12376, with probabilities 0.0485 and 0.9515, respectively. Location of the third point is determined to be -1 with initial probability 0.05. Adding this new point increases log likelihood by 0.0799, the maximum increase among all points in the searching region. Then, fix the location and probability of the first point (-2.4258 with probability 0.0485), and correspondingly adjust the location and probability of the second point such that probabilities sum up to 1 and expectation of frailty is 0. Thus, initial values for estimating three-mass-point frailty are determined.

Supporting points of the three-mass-point frailty are finally estimated to be -5.6879, -1.6396, and 0.21167, with probabilities 0.0175, 0.0585, and 0.924, respectively. The maximum increase in log likelihood is negative (-0.0118) when searching for the fourth point. Therefore, the frailty eventually has three supporting points.

B Probability of having a certain number of births

Assume woman i appears in the sample from age 15 to A_i (A_i is the minimum between 49 and her age at the last interview). According to the model, woman i's probability of having a birth in the k^*th year of spell j, conditioning on that she did not have births before that year in that spell, is $1 - exp(-\theta_{ij}(k^*))$, where $\theta_{ij}(.)$ is the hazard function for spell j conditioning on $\mathbf{x}_i$ and v_i.[28] Then the probability of not having a birth that year, conditioning on not having births before that year, is $exp(-\theta_{ij}(k^*))$.

Given that woman i has had $(j-1)$ births, her probability of having the jth birth in the k^*th year after the $(j-1)$th birth is

$$\left(\prod_{k=1}^{k^*-1} exp(-\theta_{ij}(k))\right)(1 - exp(-\theta_{ij}(k^*))) = exp\left(-\sum_{k=1}^{k^*-1}\theta_{ij}(k)\right)(1 - exp(-\theta_{ij}(k^*))),$$

which measures the probability of not having any birth in the first $(k^* - 1)$ years of spell j, but having a birth in the k^*th year.

[28]Through this section, I omit $\mathbf{x}_i$ and v_i from hazard functions for simplicity.

Similarly, given that woman i has had $(j-1)$ births, her probability of not having more births in her rest life is

$$\prod_{k=1}^{K_{ij}} exp(-\theta_{ij}(k)) = exp\left(-\sum_{k=1}^{K_{ij}} \theta_{ij}(k)\right),$$

where K_{ij} corresponds to her last year in the sample.

Based on the two probabilities above, probabilities of having a certain number of births can be calculated as follows.

- **Probability of childlessness**:

$$\prod_{k=1}^{A_i-15+1} exp(-\theta_{i1}(k)) = exp\left(-\sum_{k=1}^{A_i-15+1} \theta_{i1}(k)\right)$$

- **Probability of having exactly 1 birth**. Assume the birth occurs at age a_{i1}, then the probability of not having the birth before a_{i1} but having it at a_{i1} is

$$exp\left(-\sum_{k=1}^{a_{i1}-15} \theta_{i1}(k)\right)(1 - exp(-\theta_{i1}(a_{i1} - 15 + 1))).$$

Having *exactly* 1 birth further requires having no more births after a_{i1}, whose probability is

$$\prod_{k=1}^{A_i-a_{i1}} exp(-\theta_{i2}(k)) = exp\left(-\sum_{k=1}^{A_i-a_{i1}} \theta_{i2}(k)\right).$$

Therefore, the probability of having exactly 1 birth at age a_{i1} is the multiplication of the two probabilities above:

$$exp\left(-\sum_{k=1}^{a_{i1}-15} \theta_{i1}(k)\right)(1 - exp(-\theta_{i1}(a_{i1} - 15 + 1)))exp\left(-\sum_{k=1}^{A_i-a_{i1}} \theta_{i2}(k)\right).$$

As a_{i1} may range from 15 to A_i, the probability of having exactly 1 birth is

$$\sum_{a_{i1}=15}^{A_i}\left(exp\left(-\sum_{k=1}^{a_{i1}-15} \theta_{i1}(k)\right)(1 - exp(-\theta_{i1}(a_{i1} - 15 + 1)))exp\left(-\sum_{k=1}^{A_i-a_{i1}} \theta_{i2}(k)\right)\right).$$

- **Probability of having exactly 2 births**. Assume woman i has the first birth at age a_{i1} and the second birth at a_{i2} ($a_{i1} < a_{i2}$). Then, similarly, the probability of having the first birth at a_{i1} is

$$Pr(a_{i1}) = exp\left(-\sum_{k=1}^{a_{i1}-15} \theta_{i1}(k)\right)(1 - exp(-\theta_{i1}(a_{i1} - 15 + 1))).$$

Further, the probability of having the second birth at a_{i2} is

$$Pr(a_{i2}) = exp\left(-\sum_{k=1}^{a_{i2}-a_{i1}-1} \theta_{i2}(k)\right)(1 - exp(-\theta_{i2}(a_{i2} - a_{i1}))).$$

The probability of not having more births after age a_{i2} is

$$\prod_{k=1}^{A_i-a_{i2}} exp(-\theta_{i3}(k)) = exp\left(-\sum_{k=1}^{A_i-a_{i2}} \theta_{i3}(k)\right).$$

Therefore, the probability of having exactly 2 births at age a_{i1} and a_{i2}, respectively, is

$$Pr(a_{i1})Pr(a_{i2})exp\left(-\sum_{k=1}^{A_i-a_{i2}} \theta_{i3}(k)\right).$$

As a_{i1} ranges from 15 to $A_i - 1$, and a_{i2} is from $a_{i1} + 1$ to A_i, then the probability of having exactly 2 births is

$$\sum_{a_{i1}=15}^{A_i-1} \sum_{a_{i2}=a_{i1}+1}^{A_i} Pr(a_{i1})Pr(a_{i2})exp\left(-\sum_{k=1}^{A_i-a_{i2}} \theta_{i3}(k)\right).$$

- **Probability of having exactly 3 births**. Assume the three births occur at age a_{i1}, a_{i2} and a_{i3} ($a_{i1} < a_{i2} < a_{i3}$). Similarly, the probability of having exactly 3 births is

$$\sum_{a_{i1}=15}^{A_i-2} \sum_{a_{i2}=a_{i1}+1}^{A_i-1} \sum_{a_{i3}=a_{i2}+1}^{A_i} Pr(a_{i1})Pr(a_{i2})Pr(a_{i3})exp\left(-\sum_{k=1}^{A_i-a_{i3}} \theta_{i4}(k)\right),$$

where $Pr(a_{i1})$ and $Pr(a_{i2})$ are derived as before, and

$$Pr(a_{i3}) = exp\left(-\sum_{k=1}^{a_{i3}-a_{i2}-1} \theta_{i3}(k)\right)(1 - exp(-\theta_{i3}(a_{i3} - a_{i2}))).$$

- **Probability of having 4 or more births**. Similarly, the probability of having 4 or more births is

$$\sum_{a_{i1}=15}^{A_i-3} \sum_{a_{i2}=a_{i1}+1}^{A_i-2} \sum_{a_{i3}=a_{i2}+1}^{A_i-1} \sum_{a_{i4}=a_{i3}+1}^{A_i} Pr(a_{i1})Pr(a_{i2})Pr(a_{i3})Pr(a_{i4}),$$

where a_{i4} is the age for the fourth birth, and

$$Pr(a_{i4}) = exp\left(-\sum_{k=1}^{a_{i4}-a_{i3}-1} \theta_{i4}(k)\right)(1 - exp(-\theta_{i4}(a_{i4} - a_{i3}))).$$

In order to calculate the probability of having *exactly* 4 births, I need further to know

$$\prod_{k=1}^{A_i-a_{i4}} exp(-\theta_{i5}(k)) = exp\left(-\sum_{k=1}^{A_i-a_{i4}} \theta_{i5}(k)\right).$$

But birth spells beyond 4 are not included for estimation, so the probability of having exactly n ($n = 4, 5, ...$) births is not computable.

In order to reduce computation burden, the probability of having 4 or more births is simply calculated by $1 - P_0 - P_1 - P_2 - P_3$, where P_n is the probability of having exactly n births ($n = 0, 1, 2, 3$).

Each probability calculated above is conditional on one value of frailty. The final probability is the expectation of probabilities with different values of frailty, i.e,

$$P_n = \sum_{m=1}^{M} p_m P_n(q_m),$$

where the frailty follows a mass-point distribution with supporting points q_m and probabilities p_m $(m = 1, ..., M)$.

The probabilities are calculated for each woman. Average probabilities of a groupd women can be directly obtained by averaging probabilities for the women in that group.

Calculation of standard errors of such probabilities would be super tedious, and is not considered by the paper.