Family Planning Policy in China: Measurement and Impact on Fertility*

Fei Wang

University of Southern California

First Draft: June 2011

This Draft: December 2011

Preliminary and Incomplete. Not for Citation or Circulation

Abstract

The family planning policy in China has been widely studied, but its impact on fertility is still an issue with controversial conclusions. One source of the disagreement is the difference in measuring the family planning policy. Most previous measures generally did not completely capture the secular and cross-sectional policy variations, failed to heterogeneously express people's exposures to the policy, and sometimes suffered from the problem of endogeneity. This paper reviews the history of family planning policy in China, and based on that constructs a new measure which can embody the policy variations more completely, heterogeneously, and exogenously. Using a cross-sectional data of the China Health and Nutrition Survey, this paper estimates the impact of family planning policy on the number of children ever born to a woman based on the new measure, and finds that the measure can well fit the complete history of family planning policy, can explain a sizable fertility change in the data, and can serve as a qualified instrumental variable for fertility. Moreover, women with more years of education, living in richer provinces, *or* whose first child is son, tend to actively control their births and face fewer restrictions from the policy. The paper also shows that, women tend to delay the first marriage and childbearing in response to the family planning policy.

Key Words: Family Planning Policy, Fertility, China

JEL Codes: J13, J18, O22

* I am grateful to the advice from John Strauss, and to the helpful comments from Jeffery Nugent, the participants of the 86th Annual Conference of the Western Economic Association International, and my friends at the University of Southern California. All errors are my own. Comments are welcome to wangfei@usc.edu

I. Introduction

For decades, the family planning policy (FPP) in China has been drawing wide attention and discussions in demographic and economic studies. The very first issue, the impact of FPP on fertility, however, remains ambiguous. A large number of papers concluded that FPP explains a sizable proportion of the fertility decline in China, like Lavely and Freedman (1990), Yang and Chen (2004), and Li et al (2005). While many other studies argued that, the role that China's FPP has played in population control has been overstated, like Shultz and Zeng (1995), McElroy and Yang (2000), Narayan and Peng (2006), and Cai (2010).

The different conclusions may originate from the use of different data sets, different empirical strategies, and more essentially, different measures of FPP. Moreover, different measures of FPP may also lead to various second-stage results in the studies where FPP is treated as an instrumental variable for fertility, sex ratio, and etc., such as Li and Zhang (2007, 2008), Edlund et al (2008), Qian (2009), Banerjee et al (2010), Islam and Smyth (2010), and Wu and Li (2011). Therefore, it's crucial to appropriately measure FPP, which is what the paper tries to achieve.

Roughly, FPP can be expressed by specific measures or constructed measures. A few data sets offer specific information on FPP, such as the availability of family planning services (Shultz and Zeng, 1995), monetary penalties on "over-quota" births (McElroy and Yang, 2000; Li and Zhang, 2008), and subsidies for one-child families (Li and Zhang, 2008). Because such specific data are not widely available, more papers utilize basic demographic information, such as birth year, living area, and ethnicity which are easy to obtain in normal survey data, to construct measures of FPP. Such measures need to be based on the history of FPP in China, particularly the historic policy variations over periods and across different groups of people. Most previous constructed measures, however, tended to ignore part of the history of FPP, and failed to completely capture the policy variations. In addition, most previous measures were created too roughly to express heterogeneous exposures to FPP. For instant, two women at different ages should be differently exposed to the FPP of a particular period, due to their different biological capability of childbearing, while such heterogeneity was not sufficiently displayed. Moreover, some papers measure the exposure to FPP with a dummy variable indicating if a woman bears a child in a period of FPP, which may suffer from the problem of endogeneity.

Aware of the problems of previous measures, this paper tries to propose a new constructed measure, and make it more complete, heterogenous, and exogenous. Completeness must build on a complete history of China's FPP. Over time, the FPP in China experiences four periods: the period without FPP (1949-1962), the period with mild and narrowly implemented FPP (1964-1970), the

period with strong and widely implemented FPP (1971-1979), and the period of the strictest one-child policy (1980-now). Cross-sectionally, the enforcement of FPP differs between urban and rural areas, and varies from ethnic majority, known as Han, to minority, denoted by non-Han. Based on the history, the new measure will try to embody the secular and cross-sectional policy variations as completely as possible.

The basic idea of constructing the new measure comes from the intuition that a woman is more likely to be affected by the FPP of some period if she is more likely to biologically bear a child at her age during the period. Following the idea, the measure successfully maps different ages under an FPP to heterogeneous exposures to the FPP. Furthermore, as the measure is created only based on relatively exogenous variables, such as birth year, the year of interview, urban dummy, and Han dummy, it greatly lowers the risk of endogeneity.

This paper uses a cross-sectional data set of the China Health and Nutrition Survey (CHNS) to construct a new measure of FPP, and based on the measure estimate the impact of FPP on the number of children ever born to an ever-married woman, and on the probability of getting married early for both ever-married and never-married women, and on the probability of bearing the first child early for ever-married women. First, the empirical results imply that the new measure well fits the history of FPP. Second, the new measure of FPP can explain a sizable part of the fertility change in the sample. The first marriage and childbearing tend to delay in response to FPP. The measure has statistically strong effect on fertility, and may serve as a qualified instrumental variable for the level of fertility. Third, women with different characteristics may respond to FPP differently. The women who are more educated, living in richer provinces, or whose first child is son, tend to actively control their births and are less likely to be pushed or bound by FPP.

The paper is arranged as follows. Section II briefly introduces the history of FPP in China since 1949. Section III reviews the previous measures of FPP, and lights the direction of progress. Section IV introduces the data set, the new method of measuring FPP, and the empirical specifications for estimating the effect of FPP on fertility. Section V shows and remarks on the empirical results. Section VI concludes.

II. A Short History of Family Planning Policy in China[1]

The family planning policy (FPP) in China experiences four periods: the period without FPP (1949-1962), the period with mild and narrowly implemented FPP (1963-1970), the period with strong and widely implemented FPP (1971-1979), and the period with the strictest one-child policy

[1] The paper studies the family planning policy of the People's Republic of China, founded in 1949. Section II, therefore, only introduces the history after 1949.

(1980-now). Over periods, the general policy gets stronger and stronger. Within each period, the policy differs between urban and rural areas, and varies from ethnic majority to minority. For convenience, denote the four periods by period 0 to 3. Most history presented below is cited or summarized from Yang (2004, 2010).

(I) Period 0: Without Family Planning Policy (1949-1962)

On the eve of the foundation of the People's Republic of China, Mao Zedong, the supreme leader, publicly argued that China preferred a large population,[2] which also fit China's traditional concept of fertility, *Duo Zi Duo Fu* (more children, more happiness). Moreover, China was deeply influenced by a birth-encouraging policy of the Soviet Union.[3] Consequently, from 1949 to 1953, China strictly limited birth control, and financially subsidized families with a large number of children.[4]

With a rapid growth of population, China began to abolish or relax certain restrictions on birth control in 1954.[5] Meanwhile, some influential scholars, such as Shao Lizi and Ma Yinchu, publicly promoted FPP.[6] Thereafter, the knowledge of birth control to some extent spreaded through public media.[7] No official FPP, however, was conceived.

In 1958, with the onset of the Great Leap Forward campaign which aimed to use China's *vast population* to rapidly transform the country from an agrarian economy into a modern communist society through the process of agriculturalization, industrialization, and collectivization,[8] the way to an FPP was blocked.[9] The campaign was followed by a great famine (1959-1961) which led to a dramatic decline in total fertility rate, from 6.679 in 1958 to 3.287 in 1961.[10] Under such circumstances, FPP was rarely discussed. After the famine, the total fertility rate compensatorily rose from 3.287 in 1961 to 6.023 in 1962.[11]

(II) Period 1: Mild and Narrowly Implemented Family Planning Policy (1963-1970)

Pressured by the birth boom in 1962, the Chinese government issued an instruction on the implementation of family planning in December 18th, 1962, known as the No. [62]698 document. It marks the start of China's FPP.[12]

---

[2] Mao said: "A large population is preferred in China. No matter how large it is, we can always handle it with production...Human being is the most valuable resource of the world...Human can create any miracle." (Yang, 2004, page 43)
[3] All mothers bearing enough many children in the Soviet Union were awarded an honorary title *Mother Heroine*. (Yang, 2004, page 44)
[4] China strongly restricted sterilization and abortion, and strictly controlled the production and sale of contraceptives whose import was banned. (Yang, 2004, page 44-45)
[5] In 1954, China canceled the restriction on contraception and the sale of contraceptives, and relaxed the restriction on abortion. Sterilization was, however, still under strong control. (Yang , 2004, page 47-48)
[6] Shao and Ma both supported contraception and late marriage, but differed on the issue of abortion. (Yang , 2004, page 48-50, 52)
[7] The knowlege of birth control only spreaded to some cities of some provinces. (Yang, 2004, page 50, 53, 54, 58)
[8] "Great Leap Forward", *Wikipedia*, http://en.wikipedia.org/wiki/Great_Leap_Forward
[9] Yang (2004), page 59.
[10] Yang (2004), page 61.
[11] Yang (2004), page 61-62.
[12] Yang (2010), page 27. Because the document was released in the late 1962, I assume it came into effect from 1963.

The period-1 FPP, in general, featured the setting of a population growth target,[13] late marriage,[14] the establishment of family planning institutions,[15] and the dissemination of family planning knowledge and technology.[16] Specific FPP varied by province. For instance, Shandong's FPP could be informally stated as "one (child) is not few, two are just right, three are too many". Shanghai's policy suggested that, a couple should not bear more than 3 children, the birth interval should be at least 4 years, and a woman's age of bearing the first child should exceed 26.[17]

The period-1 FPP is milder than the period-2 policy on at least two aspects, the restriction on the number of births per couple, and the ways of enforcing FPP. Details will be discussed below in Part (III).

The period-1 FPP was designed to be narrowly implemented, only in urban areas.[18] Moreover, it only constrained ethnic majority people, known as Han people.[19] Ethnic minority people, denoted by non-Han people, were generally not impacted by FPP in this period.[20] Simply put, the period-1 FPP mainly affected urban Han people. Because rural population took a large proportion, over 80%, the *overall* total fertility rate stayed at high levels in period 1.[21]

From 1966, the function of family planning institutions and the implementation of FPP were negatively shocked by the Cultural Revolution. The FPP was, however, not abolished, and fertility rate in urban areas remained at relatively low levels.[22] Meanwhile, contraceptive pills were actively researched and developed.[23]

(III) Period 2: Strong and Widely Implemented Family Planning Policy (1971-1979)

Concerned for the negative impact of the Cultural Revolution, the Chinese government issued a report on family planning in 1971, known as the No. [71]51 document, to re-emphasized the importance of FPP. The report signified that FPP recovered from the Cultural Revolution, and stepped

[13] The annual population growth rate targets are 2%, 1.5%, and 1% for the Recovery Period (1963-1965), the 3rd Five-year Plan (1966-1970), and the 4th Five-year Plan (1971-1975), respectively. (Yang, 2004, page 62)
[14] The Ministry of Health proposed late marriage in a national conference in 1963, and later received the central government's approval. (Yang, 2004, 62)
[15] The national family planning institution was established in 1964, and local agencies were founded from 1963. (Yang, 2004, 65)
[16] Contraceptive knowledge and devices were available in local hospitals. Restrictions on abortion and sterilization were basically removed. (Yang, 2004, page 65-67)
[17] Yang (2004), page 68.
[18] The policy was initially designed to work in populous rural areas as well, but most of those areas were *actually* not covered by the policy (Yang, 2004, page 69). Moreover, no FPP had ever covered the urban areas of the five ethnic minority regions, Guangxi, Inner Mongolia, Ningxia, Tibet, and Xinjiang, by 1970s. (Yang, 2004, page 144-145)
[19] Yang (2004), page 62. The 2010 Census of China indicates that 91.51% of Chinese are Han people.
[20] Part of the urban non-Han people who were *not* living in the five ethnic minority regions might be affected by FPP, particularly when they identified the ethnicity of their children as Han. (Yang, 2004, page 144-145)
[21] Yang (2004), page 69.
[22] Yang (2004, page 71) claims that the implementation of FPP was actually ceased, but no evidence well supports this argument. Urban total fertility rate kept falling in this period only expect in 1968. Moreover, family planning institutions started to recover in 1969 (Yang, 2004, page 75).
[23] Yang (2004), page 70.

into a new stage.[24]

Similar to period 1, the period-2 policy also involved a population growth target, and technological supports. Moreover, the FPP became more nationally uniform, known as "late, long, few". "Late" means late marriage and childbearing. The recommended age of marriage was 25 or above for men, and was 23 or above for women; women were suggested having children after 24. "Long" means the birth interval should be at least 3 years. "Few" means a couple can at most bear two children.[25]

The period-2 policy was stronger than that in period 1. First, a couple could at most have 2 children in period 2, while 3 children were allowed, though not encouraged, to bear in period 1. Second, the enforcement of period-2 policy was stronger. Before the FPP appeared, for the first time, in the 1978 Constitution of China, the enforcement was generally in a form of political or social movements which were expected to be highly effective in those periods of China driven by political powers. Mao Zedong, the supreme leader, promoted FPP harder in period 2,[26] and thus greatly strengthened the enforcement. In addition, family planning institutions took administrative measures to enforce the FPP in both period 1 and period 2. For instance, families with fewer children would be subsidized in some way.[27] Penalties were applied, particularly in period 2, against too many births. For example, rural people who did not comply with FPP would face a loss in food distribution in 1970s.[28]

In 1971, the FPP spreaded to rural areas, which was consistent with a dramatic decline, starting from 1971, in the *overall* total fertility rate (see Figure 1).[29] The population growth target, however, differed between urban and rural areas. By 1975, the urban population annual growth rate should fall to 1%, and the rural growth rate should drop to 1.5%.[30] In this period, non-Han people were generally not constrained by FPP,[31] though they could get access to the knowledge and technological supports of family planning.

(IV) Period 3: One-child Policy (1980-now)

As a natural evolution of the period-2 FPP, the one-child policy was conceived in 1979, and was intensively propagandized in 1980.[32] The one-child policy, as the name suggests, restricted a couple to

[24] Yang (2004), page 73.
[25] The policy was first implemented in some parts of China, and then was extended to the whole nation in 1973. (Yang, 2004, 73)
[26] For example, in 1974, Mao said: "We must control the population." (Yang, 2004, 73)
[27] Yang (2004), page 64, 74.
[28] Yang (2004), page 80, 135.
[29] Yang (2004), page 77-79.
[30] Yang (2004), page 72.
[31] Yang (2004), page 143-145. The urban or rural non-Han people who were *not* living in the five ethnic minority regions might be affected by the period-2 FPP, particularly when they identified the ethnicity of their children as Han.
[32] In September 1980, the CPC central committee wrote an open letter, to expound the necessity of the one-child policy. This event was usually considered as the starting point of the one-child policy. (Yang , 2004, page 86)

having only one child. This policy was apparently stricter than any previous FPP.

The strictness of the one-child policy was also reflected by the enforcement. As mentioned in Part (III), before 1980s, the enforcement of FPP was mainly driven by political and administrative forces, not laws. In 1978, FPP appeared in the Constitution for the first time, and came up with more details in the 1982 amended Constitution. From the late 1980s, central and local governments successively legislated on FPP.[33] Legal measures, such as monetary penalties and subsidies,[34] ensured the effective enforcement of the one-child policy.

In the early 1980s, the one-child policy was successfully implemented in urban areas, but received large resistance from rural people.[35] Subsequently, in the mid-1980s, the one-child policy was relaxed for rural couples, and allowed them to have a second child in certain cases, for example, when the first child was daughter.[36]

In 1982, the FPP started to cover most non-Han people, but in more relaxed forms. In general, an urban non-Han couple could conditionally have two children, and a rural non-Han couple was conditionally allowed to have three or even more children. For an ethnic group with a small population size, the policy was even further relaxed.[37]

(V) History of Total Fertility Rate

Part (I)-(IV) briefly introduce the history of FPP, putting emphasis on the secular and cross-sectional policy variations. Over time, the general FPP gets stronger and stronger, in terms of the number of births allowed and the strength of enforcement. Cross-sectionally, the FPP spreads from urban to rural areas, and from Han to non-Han people, with weaker policies for rural or non-Han people in any period. Since the paper will focus on the impact of FPP on fertility, Table 1 summarizes the number of children allowed to bear over periods and across different groups of people,[38] and will be used to test if the newly constructed measure in the paper well fits the history of FPP.

[Table 1 is inserted to here.]

Figure 1 and 2 present the historic trend in total fertility rate (TFR) to provide more evidence for the history summarized above.

[Figure 1 is inserted to here.]

Figure 1 shows the overall, urban, and rural TFRs over time. The left, middle and right dashed

[33] Yang (2004), page 161.
[34] McElroy and Yang (2000), and Li and Zhang (2008).
[35] Yang (2004), page 86.
[36] Yang (2004), page 87. This case reflects a strong son preference in rural areas.
[37] Yang (2004), page 146-148.
[38] Through this paper, I will mainly use urban/rural and Han/non-Han to capture the cross-sectional variations of FPP. FPP also varied across provinces, but information is insufficient to specify the provincial differences. Instead, the following econometric analysis will simply control for province dummies to capture the provincial policy variations. Some recent policy change is not considered in the paper. For example, a couple who are both the only child to *their* families can have a second child, which was approved by Henan Province, the last province adopting this policy change, in November 2011.

lines mark the years of 1963, 1971 and 1980, the starting points of period 1 to 3 FPP.

Figure 1 tells that, before 1963, both urban and rural TFRs are high, except during the great famine (1959-1961). From 1963, urban TFR fell while rural TFR still remained high, consistent with the history that FPP was only implemented in urban areas in the period. After 1971, both urban and rural TFRs dropped, supporting the history that FPP covered both urban and rural areas after 1971, though to different degrees. From 1980, both urban and rural TFRs became stable and remained low under the one-child policy, with a clear urban-rural fertility gap. The overall TFR comoves with rural TFR due to the large proportion of rural population.

[Figure 2 is inserted to here.]

Figure 2 shows the overall, Han, and non-Han TFRs over time. The left and right dashed lines mark the years of 1971 and 1980, the starting points of period 2 and 3 FPP.

Before 1971, FPP mainly affected urban Han people, and non-Han TFR was supposed to be high in that period. Furthermore, due to the large proportion of rural Han population in the entire Han population, the Han TFR before 1971 would not significantly reflects the impact of period-1 FPP. Despite the unavailability of data, we can infer from the TFRs in 1971 that both Han and non-Han TFRs were indeed high before 1971.

From 1971 to 1979, Han TFR fell, consistent with the history that FPP spreaded from urban Han people to rural Han people in that period. Three possibilities could explain the drop of non-Han TFR in the period. First, according to the history, non-Han people could get access to family planning knowledge and technological supports that time, and part of them might *actively* control births. Second, part of the non-Han people who were *not* living in the five ethnic minority regions could be affected by the FPP of the period, particularly when they identified the ethnicity of their children as Han. Third, the change in other non-FPP factors, such as education, income, etc., might lead to an active control of births as well. Despite the decline in non-Han TFR, the Han-non-Han fertility gap became larger in the period.

Only a few years of TFRs are available after 1980. Around 1980, the Han-non-Han TFR gap was as high as 2. While in 1989, the gap narrowed to about 0.6, consistent with the history that FPP started to widely impact non-Han people after 1980. Due to the large proportion of Han population, the overall TFR comoves with Han TFR over time.

To summarize, Figure 1 and 2 well support the history of FPP. Only based on a complete history can we appropriately measure FPP.

III. A Short Review of Previous Measures of Family Planning Policy

The first step of exploring FPP related topics is to appropriately measure FPP. Before proposing a new measure, it is necessary to review previous measures to see why they are improvable. Roughly, there are specific measures and constructed measures.

Specific measures directly come from a data set which contains specific information on FPP. Schultz and Zeng (1995) use a specific individual level cross-sectional data for some rural areas of three provinces in China, collected in the 1985 In-Depth Fertility Survey (IDFS), to assess the effect of local family planning and health programs on fertility. FPP is measured by the availability of a family planning service station, a family planning outreach worker, a doctor or nurse, and a local clinic in a rural village.[39] McElroy and Yang (2000) use a specific household level cross-sectional data for some rural areas across ten provinces, collected in the 1992 Household and Economy Fertility Survey (HESF), to estimate the intensity of county-level FPP on the number of children per family. The HESF sample provides county-level monetary penalties imposed on "over-quota" births, and they are used to measure the county-level intensity of FPP. Li and Zhang (2008) use an individual level cross-sectional data, collected in the 1989 China Health and Nutrition Survey (CHNS), to study how birth behaviours of a woman are affected by the birth behaviours of her neighbours. In the first stage regression, simialr to McElroy and Yang (2000), they use community-level monetary penalties on "over-quota" births, and subsidies for one-child families, as a measure of the one-child policy, and use it to instrument the fertility of neighbours.

Specificity is one of the most notable advantages of such measures. They are so detailed that few irrelevant factors can be involved in. Such data sets are, however, relatively exclusive and difficult to acquire.[40] In addition, those measures may capture only part of the policy variations.[41] For example, according to the three papers, the enforcement of FPP at least relies on family planning services, penalties, and subsidies. Each paper tells part of the story.

Because a specific data set is not generally available, more papers utilize basic demographic variables, such as birth year, living area, ethnicity, etc., which are easy to obtain in normal survey data, to construct measures of FPP. Such measures need to be created based on the history of FPP, and to

[39] The availability is measured by dummy variables. The interaction of the dummy variables are also controlled in regressions.

[40] This paper will use a data set from the China Health and Nutrition Survey (CHNS) for empirical analysis, like Li and Zhang (2008). But the data used here doesn't include the community level sample from which Li and Zhang obtain the information on monetary penalties and subsidies. Different from other parts of the CHNS data, the community level sample can only be obtained after a formal application is approved.

[41] It could to some extent explain why the impact of FPP is relatively small in such papers. Schultz and Zeng (1995) conclude that, a representative woman in the younger sample residing in a village with a family planing service station, a family planning outreach worker, and a doctor or nurse, is expected to have a 0.25 lower fertility (drop by 21%) than a woman residing in a village with none of the three services. The effect is a little bit smaller in the older sample of women. McElroy and Yang (2000) conclude that, a removal of the existing penalties would increase fertility by 0.33 births per woman which appears to be small in view of the government's presumptions. Since Li and Zhang (2008) mainly focus on the external effect of birth behaviours, they do not specifically analyze the magnitude of the effect of FPP on fertility in the first stage regressions.

capture the secular and cross-sectional policy variations, as stated in Section II, as completely, heterogeneously and exogenously as possible. An incomplete measure ignores part of the history, and therefore captures only part of the policy variations and may under-estimate the impact of FPP. A homogeneous measure could assume equal exposures to FPP for heterogeneous women, which may be too rough to be reliable. An endogenous measure would lead to biased estimations in regressions.

The previous constructed measures could in general be improved in the completeness of capturing the history of FPP.

Some constructed measures reflect part of the secular policy variation, but fail to take cross-sectional variations into account. Yang and Chen (2004) use the 1992 HESF sample to assess the effect of FPP on fertility. They apply the year dummies of being married, from 1970 to 1989, to capture the various impacts of FPP for different marriage cohorts. Narayan and Peng (2006) use time series data and models to estimate the effect of FPP on fertility. They measure FPP with time dummies for two periods, 1970-1979 and 1980-2000. Similarly, Edlund et al (2008) measure FPP with a dummy variable of being exposed to the one-child policy one year prior to a mother's childbearing.

Some studies utilize cross-sectional policy variations, but fail to capture secular variations. Cai (2010) uses a county level cross-sectional data of Jiangsu and Zhejiang province, collected from the 2001 statistical year books of the two provinces, and the 2000 census compilations, to estimate the effect of FPP on fertility, measuring FPP with the percentage of population with agricultural *hukou*[42] and the percentage of Han population in each county.

More studies take both secular and cross-sectional policy variations into account, but either miss part of the urban-rural or Han-non-Han variations, or part of the policy change over time. Li et al (2005) apply a difference-in-difference approach to assess the impact of the one-child policy on fertility. The treatment and control groups are Han and non-Han people. The pre-treatment and post-treatment samples come from the 1982 and 1990 census, respectively. Li and Zhang (2007) use a provincial panel data involving 28 provinces over 20 years (1978-1998) to estimate the effect of birth rate on economic growth. In the first stage regressions, they use the percentage of non-Han people in each province/year to instrument the birth rate. Other than monetary penalties and subsidies, Li and Zhang (2008) also adopt a difference-in-difference approach in the first stage regressions, like Li et al (2005). Qian (2009) uses an individual level cross-sectional sample from the 1990 census and the 1989 CHNS, to test the quantity-quality trade-off hypothesis. In the first stage regressions, she makes use of cross-region policy variations and the policy evolution after 1970 to instrument family size. Islam and Smyth (2010) use the 2008 China Health and Retirement Longitudinal Survey (CHARLS) data to

[42] *Hukou* is a household registration system in China. In general, urban and rural people have non-agricultural and agricultural *hukou*, respectively. Many urban people who temporarily migrated from rural areas may still have an agricultural *hukou* as before.

estimate the effect of number of children on parental health. In the first stage regressions, they take advantage of urban-rural policy variations and the policy change after 1970 to instrument the number of children. Banerjee, Meng and Qian (2010) use an individual level cross-sectional data, collected in the 2008 Urban-Rural Migration in China and Indonesia Survey (RUMiCI), to study the impact of number of children on parental saving behaviours. In the first stage regressions, they use a dummy variable to capture the policy change in the early 1970s, and include the interaction between the dummy and the gender of the first child to capture the effect of son preference. Wu and Li (2011) use an individual level panel sample with 5 waves from the CHNS data to assess the effect of family size on maternal health. In the first stage regressions, they construct a time variable about the one-child policy and interact it with urban dummy and Han dummy, and use them to instrument family size.

Most previous measures fail to capture women's heterogeneous exposures to FPP. For example, many measures define the exposure to the FPP of some period as dummy variables of whether bearing a child in the period. In other words, if two women both bear a child in some period, they would be treated as equally exposed to the FPP of the period. However, for the two women, the length of interval exposed to the FPP could differ, because one woman could enter the FPP at a young age while the other woman may almost reach her ending point of fertility. Wu and Li (2011) progress at this point by defining a measure based on the age at which a woman started to be affected by FPP. Because they assume the exposure as a function of the *length* of interval exposed to FPP *only*,[43] further improvement could be achieved.

Other than the issue of incompleteness and homogeneity, some measures are endogenously constructed, which may bias the effect of FPP and invalidate the potential role of FPP as an instrumental variable. As reviewed above, Yang and Chen (2004) use the year dummies of being married to capture secular policy variations, but the year of marriage is endogenous and might be correlated with unobserved factors related to fertility. Similarly, Edlund et al (2008), Qian (2009), Islam and Smyth (2010), and Banerjee, Meng and Qian (2010) all use whether some child was born in some period of FPP to measure the exposure to the policy, and the timing of childbearing may be endogenous.

To summarize, previous (constructed) measures are generally not complete, heterogeneous, and/or exogenous. This paper will concentrate on the improvement of constructed measures on the three aspects.

[43] For example, woman A was exposed to the policy from 20 to 30 years old, and woman B was exposed to the policy from 40 to 50 years old. The length of interval exposed to the policy was the same, while the actual exposures were reasonably different, because woman A was more fertile than B when they were affected by the policy.

IV. Data, New Measure of Family Planning Policy, and Empirical Specifications for the Impact of Family Planning Policy on Fertility

(I) Data: Introduction, Descriptive Statistics, and Representativeness

The paper mainly uses the birth history data from the China Health and Nutrition Survey (CHNS).[44] CHNS interviewed ever-married women who were below 52 and had ever born a child in 1991, 1993, 2000, 2004, 2006, and 2009, across 9 provinces,[45] and recorded their birth history, including the date of birth, gender, living arrangement, and date of death of *every* child that a woman has *ever* given birth to.

A woman may be tracked by the CHNS round by round. Over years, many faded out, and many faded in. The CHNS team combined the birth history data of all rounds, only kept the latest round of record for each woman who has ever been interviewed, and released the refined data online.[46] In other words, the data contains the birth history of a woman up to the latest round of interview for her. Furthermore, since the released birth history data only includes the women with experience of childbearing, I merged the data to another cross-sectional data for the ever-married women who had never born a child up to the latest round of interviews (see details in Appendix B2) to form a complete birth history data. Clearly, women in the data correspond to different rounds of interviews. In particular, among the 6745 women, 4.24%, 13.29%, 12.38%, 6.45%, 17.91% and 45.63% were interviewed in 1991, 1993, 2000, 2004, 2006, and 2009, respectively.

The data was restricted to women aged 15 or above during the interview. Although only the women under 52 should be interviewed, around 10% women in the sample were above 52 during the interview. I kept those observations to enlarge the sample, after checking the validity of these records. To rule out extreme cases, the women who ever bore a child when they were below 15 or above 49 were dropped from the data.[47]

To further study the impact of FPP on the age of first marriage, I merged the data to another data for never-married women up to the latest round of interviews. To summarize, the entire data used in the paper consists of ever-married women and never-married women up to their latest rounds of interviews; the part of ever-married women can be further divided into the women with and without experience of childbearing. For any empirical analysis about fertility, only the ever-married women

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

[45] The CHNS data were collected in 1989, 1991, 1993, 1997, 2000, 2004, 2006, and 2009, while no birth history was recorded in round 1989 and 1997. In round 1991, the interviewed women were under 50. Before round 2000, the survey covered 8 provinces, Guangxi, Guizhou, Henan, Hubei, Hunan, Jiangsu, Liaoning, and Shandong. Heilongjiang was included in round 2000 and thereafter.

[46] The data is named "m10birth", and was released in July 2011 on the official website of the CHNS. The data doesn't contain the information of round for any woman, therefore I merged the data to other ever-married women data (for example, the marriage history data for the same set of women) with the information of round, and mapped the latest round to each woman in the birth history data. Details will be clarified in Appendix B4.

[47] Only 20 observations were dropped.

subsample is used. While the analysis about the age of first marriage will be based on the entire data. The entire data was further merged with the individual characteristics of women, such as province, year of birth, ethnicity, living in urban or rural, education, etc. Details about the construction of data and variables are discussed in Appendix B.

Figure 3 displays the descriptive statistics of the data and key variables over cohorts.

[Figure 3 is inserted to here.]

The cohort ranges from 1925 to 1991. The large cohort range covers the entire history of FPP, and helps estimate the impact of FPP of different periods conveniently, which is why the paper does not use a cross-sectional data of a single year. Due to the small number of observations before cohort 1931 and after cohort 1988, I merged those observations to cohort 1931 and 1988, respectively, in Figure 3, to make lines smoother.[48]

Figure 3a plots the proportion of sample size to each cohort. It is roughly bell-shaped, and the women born in 1950-1970 take the largest proportion in the data.

Figure 3b shows the average number of children ever born to the women of each cohort. The number of children is small for the oldest cohorts, and then climbs up quickly, and finally gradually goes down. Three possibilities could explain the counterintuitive fertility levels of the oldest cohorts. First, the sample size of those cohorts is too small to generate a statistically reliable sample average. Second, the women bearing fewer children may be healthier, and be more likely to appear in the sample. Third, impacted by the great famine in 1959-1961, those cohorts may have fewer children[49]. If the fertility of the oldest cohorts, who were less likely to be exposed to FPP, is understated, the impact of FPP on fertility, for younger cohorts, tends to be under-estimated. This point should be kept in mind while explaining the regression results below.

Figure 3c presents the proportion of women living in urban areas to each cohort *during the interview*, and the proportion fluctuates between 20% to 40%. In order to capture urban-rural policy variations, it is better to know where a woman lived *before her childbearing*. In other words, the migration history of a woman matters. During the interview, a rural woman was less likely to originally come from urban areas,[50] while it was quite possible to encounter an urban woman who migrated from rural areas.[51] Therefore, identifying all women living in urban areas during the

[48] But I didn't do that in regression analysis.
[49] Cohort 1925 to 1940 were between 20 to 35 years old during the great famine, and therefore their fertility might be strongly impacted by the famine.
[50] A strict household registration system, known as the *hukou* system, was enforced after the foundation of the People's Republic of China. Under that system, the urban-rural migration was rare, until the Down-to-the-countryside Movement in the late 1960s. Many urban youth were sent down to rural areas to learn from farmers and workers during that period, and most of them returned to urban areas in the late 1970s. Since this movement was the only notable from-urban-to-rural migration in the history of the P.R.C., and only a small part of the urban youth finally stayed in rural areas, it is less likely to encounter a rural woman originally coming from urban areas during the CHNS interviews.
[51] As introduced in the footnote above, a woman interviewed by the CHNS might be a returned sent-down youth who could have been deeply influenced by the rural life. More notably, large waves of from-rural-to-urban migration occurred in 1980s,

interview as urban women would under-estimate the impact of FPP in urban areas and narrow the urban-rural policy gap. This point will be considered while explaining the regression results below.

Figure 3d displays the proportion of Han women to each cohort, and the proportion stays around 80%-90%. Similar to the argument above, it is possible to change the ethnicity so that the ethnicity during the interview might be different from that before childbearing. According to Wu and Li's (2011) argument, this possibility is fairly small. The lines in Figure 3c and 3d are less reliable for the oldest and the youngest cohorts due to small sample sizes.

Figure 3e plots the average age of women during the interview to each cohort. Since the women of the same cohort might be interviewed in different rounds, age is not always a decreasing function of cohort. Clearly, many women were still at fertile age when they were interviewed. Therefore, it's crucial to control women's age during the interview when we assess the effect of FPP on fertility. Figure 3f indicates that the completed years of education dramatically increases over cohorts. Again, the years of education measures the education level during the interview, not necessarily the level before childbearing. In the sample, only about 2.8% women completed education after the first birth. Therefore, the years of education during the interview can be reasonably assumed to be the education before fertility. From Figure 3b to 3f, dashed lines represent the upper and lower bounds of 95% confidence intervals.

Since most analyses are about fertility, Figure 3 only shows statistics based on the ever-married women data. After including the part of never-married women, Figure 3d and 3e are almost the same; The lines in Figure 3a , 3c and 3f are a bit higher for the youngest cohorts.[52]

Given the counterintuitive fertility levels of the oldest cohorts in Figure 3b, the representativeness of the data might be questioned. In order to further check for the representativeness, I calculated a sample total fertility rate, and compared it to the population total fertility rate, as shown in Figure 1 and 2.

The formula of the total fertility rate used here is

$$TFR_t = \sum_{a=15}^{49} B(a,t) / W(a,t) \qquad (1)$$

where TFR_t means the total fertility rate in year t; a means age, from 15 to 49; B(a,t) means the total number of children born in year t AND to the women aged a in *that* year; W(a,t) means the total number of women aged a in year t.

and are still ongoing today. Most of the migrants temporarily stayed in urban areas for higher-paid jobs.

[52] Given that most never-married women are the youngest cohorts, Figure 3a will assign more weight on those cohorts. Urban young cohorts are less likely to be married than rural young cohorts, so the proportion of urban young cohorts in Figure 3c would be larger after including never-married women. The never-married young cohorts may be pursuing higher education in school, therefore the level of education for those cohorts would be higher after including the never-married women in the data.

Since cohort ranges from 1925 to 1991 in the data, I can only calculate the total fertility rate for 1974-2006,[53] as shown in Figure 4.

[Figure 4 is inserted to here.]

The sample total fertility rate well fits the population counterpart, and the correlation coefficient is as high as 0.97. Therefore, the data proves to be representative, at least in terms of the total fertility rate.

As Olsen (1994) points out, the year-by-year total fertility rate may not be a good indicator when fertility pattern greatly changes over time, like in China. In this case, a fertility rate calculated for each cohort would be preferred, as shown in Figure 5. The formula of the cohort-adjusted total fertility rate is

$$TFR_c = \sum_{a=15}^{49} B(c,a) / W(c,a) \qquad (2)$$

where TFR_c means the total fertility rate of cohort c; a means age, from 15 to 49. B(c,a) means the total number of children born to cohort-c women when they were aged a; W(c,a) means the total number of cohort-c women when they were aged a. Because the women born after 1960 had not biologically completed fertility by the last round 2009, Figure 5 only ranges from cohort 1931 to 1960, with cohorts older than 1931 merged to the 1931 cohort as in Figure 3. Clearly, Figure 5 is similar to Figure 3b over the same range of cohorts, and also requires cautiousness while explaining the total fertility rate of the oldest cohorts.

[Figure 5 is inserted to here.]

(II) A New Measure of Family Planning Policy and Empirical Specifications for the Impact of Family Planning Policy on Fertility

An empirical model for estimating the impact of FPP on fertility can be written as

$$N_{ict} = \alpha + \sum_k \beta_k X_{kict} + \sum_{j=1}^{3} \gamma_j FPP_j(c,t,D_{urban},D_{Han}) + \mu_c + \delta_t + \eta_{ct} + \varepsilon_{ict} \qquad (3)$$

In equation (3), i, c, and t indicate individual i, cohort c, and the year of interview t. N_{ict} is the number of children ever born to a cohort-c woman i interviewed in round t. X_{kict} involve a set of individual level variables related to fertility, such as age, completed years of education, Han dummy, urban dummy, and province dummies.

Function FPP_j measures the period-j FPP, with period 0 as the base period. First, FPP_j is a function of cohort c and the year of interview t. Different cohorts were exposed to the period-j FPP differently; the year of interview t helps measure the length of interval exposed to the last period of FPP, the

[53] The older cohorts are absent for the calculation of the total fertility rate before 1974, while the younger cohorts are not complete for calculating the total fertility rate after 2006.

ongoing one-child policy. Second, FPP_j is also a function of urban dummy and Han dummy. As Table 1 shows, the enforcement of FPP differed between urban and rural areas, and between Han and non-Han women.

μ_c, δ_t, and η_{ct} are the effects of cohort, year of interview, and cohort/interview year interactions. These effects are used to capture part of the unobserved factors related to fertility, such as cohort and/or year specific socio-economic variables, health status, benefit and cost of bearing and raising a child, preference on fertility, round-specific sampling bias, etc. ε_{ict} is an idiosyncratic error term.

The core issue is to specify FPP_j. I assume a quasi-separable form for FPP_j as follows.

$$FPP_j(c,t,D_{urban},D_{Han}) = \theta_0 FPP_j(c,t) + \theta_1 D_{urban} FPP_j(c,t) + \theta_2 D_{Han} FPP_j(c,t) \qquad (4)$$

where D_{urban} and D_{Han} are urban and Han dummies. In other words, secular policy variations are captured by a cohort-year specific policy measure $FPP_j(c,t)$, and cross-sectional policy variations are reflected by the interactions between $FPP_j(c,t)$ and group dummies. $FPP_j(c,t)$ can be further specified as follows.

For an individual from a given cohort, the degree of exposure to period-j FPP depends on how biologically possible the individual bears a child in period j. For example, if an individual was at her peak age of fertility in period j, she was more likely to be affected by FPP than individuals at other ages, and thus should have a larger value for $FPP_j(c,t)$. Therefore, I need to first figure out the probability of childbearing at every age.

Figure 6 plots the probability distribution of childbearing over age, denoted by f(age). At each age, f(age) is the total number of children born to the women at that age divided by the total number of children ever born in the entire data. I assume the probability of childbearing is 0 for the women aged below 15 or above 49. Clearly, probabilities sum up to 1. Figure 6 shows that, the probability rapidly climbs up after age 15, and reaches the peak around age 23, and then gradually falls.

[Figure 6 is inserted to here.]

Define age_{jc}^b and age_{jc}^e as the age of a cohort-c woman at the beginning and ending of the period-j FPP, respectively. Then, $FPP_j(c,t)$ is defined as

$$FPP_j(c,t) = \sum_{age=age_{jc}^b}^{age_{jc}^e} f(age) \qquad (5)$$

with

age_{1c}^b = 1963 - year of birth

age_{2c}^b = 1971 - year of birth

age_{3c}^b = 1980 - year of birth

age_{1c}^{e} = 1970 - year of birth

age_{2c}^{e} = 1979 - year of birth

age_{3c}^{e} = t - year of birth

Figure 7 presents $FPP_1(c,t)$, $FPP_2(c,t)$, and $FPP_3(c,t)$, defined above, over cohorts.

[Figure 7 is inserted to here.]

Clearly, the women born in 1940s, 1950s, and 1960s were mostly affected by the period-1, period-2, and period-3 FPP, respectively. Since the period-3 FPP, the one-child policy, is still on the stage, the interval exposed to the policy for a woman born after 1960 is the interval between age 15 (or the age at the beginning of the one-child policy) and the age during the latest interview.

Finally, the empirical specification becomes

$$N_{ict} = \alpha + \sum_{k} \beta_k X_{kict} + \sum_{j=1}^{3} [\gamma_{j0} FPP_j(c,t) + \gamma_{j1} D_{urban} FPP_j(c,t) + \gamma_{j2} D_{Han} FPP_j(c,t)] + \mu_c + \delta_t + \eta_{ct} + \varepsilon_{ict} \quad (6)$$

where $\gamma_{j0} = \gamma_j * \theta_0$, $\gamma_{j1} = \gamma_j * \theta_1$, and $\gamma_{j2} = \gamma_j * \theta_2$.

The newly constructed measure improves in completeness, heterogeneity, and exogeneity, compared to previous measures. First, the new measure takes the secular and cross-sectional policy variations of all periods into account, capturing the history more completely. Second, the new measure largely allows heterogeneous exposures to FPP, for women from different cohorts. Third, the new measure is constructed only based on birth year, the year of interview, urban dummy and Han dummy, and is sufficiently exogenous.[54]

Inspired by the spirits of defining $FPP_j(c,t)$, I construct two more control variables, the exposure to the great famine in 1959-1961 and the exposure to the pre-PRC period, and include them in X_{kict}.

In order to capture the fertility drop caused by the great famine in 1959-1961, I define an exposure to the famine by using equation (5), with age_c^b the age in 1959 and age_c^e the age in 1961.

I only discuss the family planning policy in the People's Republic of China (PRC, since 1949), and little information on population related policies is available for the pre-PRC period. Women born after 1933 became fertile (aged 15) in the PRC period, and therefore are assigned value 0 on the exposure. For women born before 1934, I define the exposure by using equation (5), with age_c^b 15 and age_c^e the age in 1948.

In equation (6), δ_t will be round dummies. μ_c can be defined in two ways, a high order polynomial of birth year trend, and birth year dummies. By definition, $FPP_1(c,t)$ and $FPP_2(c,t)$ are constant within

[54] FPP_j is also constructed based on the function f(.) which is endogenously derived from the sample. However, the function is equally imposed on all the women in the sample, and thus will not lead to a problem of endogeneity.

a cohort. Therefore, if I want to include these two variables, the high order polynomial of birth year trend should be used. The birth year trend is defined as the difference between birth year and 1925, the oldest cohort in the sample. Empirically, I'll include the birth year trend and its second, third, and fourth order terms. Birth year dummies will also be considered in regressions. Correspondingly, η_{ct} will be the interactions between birth year trend and round dummies when μ_c is the polynomial, and will otherwise be the interactions between birth year dummies and round dummies.

V. Empirical Results

(I) Impact of Family Planning Policy on Fertility

Table 2 presents regression results for the impact of FPP on the number of children ever born, based on the data of ever-married women. Column-[1] regression corresponds to equation (5). Other than FPP related variables, it controls urban dummy, Han dummy, completed years of education and the square of demeaned education (divided by 2), age during the interview and the square of demeaned age (divided by 2), exposure to the great famine, exposure to the pre-PRC period, the polynomial of birth year trend up to the fourth order, round dummies and their interactions with the birth year trend, and province dummies. Column [2] further controls $D_{urban}*D_{Han}*FPP_j$ to capture more cross-sectional policy variations. Based on Column [1], Column [3] replaces the polynomial of birth year trend with birth year dummies, and replaces the interactions between round dummies and birth year trend with the interactions between round dummies and birth year dummies. FPP_1, FPP_2, FPP_3, exposure to the great famine and exposure to the pre-PRC period are dropped due to perfect collinearity.[55] Column [4] further controls $D_{urban}*D_{Han}*FPP_j$ on the basis of Column [3]. Column [5] and [6] are logit regressions using the same specification with Column [1]. The dependent variable of Column-[5] regression is a dummy variable equal to 1 if the number of children ever born to a woman is greater or equal to 2. The dependent variable of Column-[6] regression is a dummy variable taking 1 if the number of children is greater or equal to 3. All the coefficients in Column [5] and [6] are marginal effects. Standard errors are reported in squared brackets of all columns.

[Table 2 is inserted to here.]

Table 2 shows that, the coefficients of FPP related variables are in general negative through all columns, implying that FPP has negative impact on fertility, and the effect of FPP is stronger in urban areas and for Han women. Detailed effects over time and across groups will be discussed in Table 3 and 4.

The effects of other variables can be easily read from Table 2. First, the number of children or the probability of having more children is generally a decreasing function of mother's education. Column

[55] They are perfectly collinear with birth year dummies, round dummies and their interactions.

[3] and [4] imply that fertility is an increasing and concave function of age. Second, more exposure to the great famine leads to fewer children ever born, and more exposure to the pre-PRC period results in more children ever born. The inclusion of the three-term interactions in Column [2] and [4] do not change much the effects of other FPP variables. Using birth year dummies instead of birth year trend in Column [3] and [4] also do not change much the effects of variables involving FPP.

The "F statistics for Joint Significance of FPP" stands for the F statistics for the joint significance test for *all* the variables involving FPP. Through all columns, F statistics are fairly large with p values 0.0000, even when FPP_1 to FPP_3 are dropped in Column [3] and [4]. It implies that the new measure of FPP can serve as a qualified instrumental variable for fertility. The Chi-squared statistics for the joint significance test for all the variables involving FPP are also sufficiently large with p values 0.0000. The R-squared and pseudo R-squared signify nice goodness-of-fit of all specifications.

Table 3 shows *marginal* effects of the FPP of all three periods on the number of children ever born to each group of women, with period 0 as benchmark in each group. For example, a rural Han woman would have had 0.355 fewer children if she was, to an *average* degree, exposed to the period-2 policy instead of not being affected by any FPP as in period 0, given all the other factors constant. The average degree of exposure to period-j policy is expressed by assigning FPP_j its own average value based on the women who had *positive* exposure to the period-j policy. The average values of FPP_1 to FPP_3 are 0.2241, 0.3101 and 0.6547, respectively. All marginal effects are derived from the coefficients of FPP related variables in Column [1] of Table 2. Taking the 0.355 mentioned above as an example,

-0.355 = (the coefficient of FPP_2 + the coefficient of $D_{Han}*FPP_2$) * the sample average of FPP_2

= (-0.457 + (- 0.686)) * 0.3101.

Stars represent the significance level of marginal effects. Three, two, and one star(s) indicate being significant at the level of 1%, 5%, and 10%, respectively

[Table 3 is inserted to here.]

If only 1% significant (3 stars) effects are treated to be non-zero, an urban Han woman started to be significantly affected by FPP in period 1, a rural Han woman started in period 2, and a rural non-Han woman started in period 3. The starting points of the three groups of women perfectly match the history summarized in Table 1. Only the urban non-Han woman is inconsistent, possibly because they were imposed on a similar FPP when urban Han women living in the same region were covered by an FPP, as noted in Section II and Table 1. Moreover, for each group of women, the marginal effect becomes stronger and stronger over periods, also matching Table 1. For each period, the marginal effect is smaller in magnitude for rural or non-Han women.

Table 4 displays the cross-sectional policy variations *within* each period more clearly, using the

same calculation approach as in Table 3. In Table 4, within each period, the FPP for urban Han women is generally the strongest, and that for rural non-Han women is the weakest. Rural Han and urban non-Han policies are in between, and the latter is empirically stronger than the former. The cross-sectional comparisons are consistent with the history as well.

[Table 4 is inserted to here.]

To summarize, the results of Table 3 and 4 well capture the characteristics of FPP introduced in Section II, implying that the new measure fits the history more completely. Moreover, taking the marginal effects of the FPP of all periods into account, the impact of FPP on fertility is fairly sizable. For instance, Figure 5 implies that the total decline in fertility is around 2 in the data, and FPP can in general explain over half of the decline according to Table 3, not to mention the possibility that the effect of FPP is under-estimated as discussed before. Despite the controlling of many variables, an omitted variable bias could still exist. More model specifications could be tried for further studies.

According to the history, FPP itself varies from urban to rural areas, and from Han to non-Han people. Other than living area and ethnicity, women with different characteristics may face different degrees of restrictions from FPP. Table 5 considers three such characteristics, completed years of education, gender of the first child, and whether living in eastern coastal provinces.

[Table 5 is inserted to here.]

Column [1] and [2] are for less and more educated women, respectively. Nine years of education indicates graduation from junior high school, and is also the current compulsory level of education in China. Column [3] and [4] distinguish women by the gender of the first child. Given the strong son preference in China, a woman whose first child is son might behave differently from a woman whose first child is daughter. Column [5] and [6] are for the women living in eastern coastal and non-eastern coastal provinces, respectively. In the data, Jiangsu, Liaoning and Shandong are identified as eastern coastal provinces. Eastern coastal province are generally richer provinces in China. Due to endogeneity and measurement problem, all regressions do not directly control income. Therefore, Column [5] and [6] are useful in estimating the impact of FPP on women with different income levels. Model specifications in Table 5 are the same with Column [1] of Table 2. Some variables are dropped because of perfect collinearity. Before running the regressions in Table 5, I test for equal coefficients for each pair of subsamples, and report the F statistics at the bottom of Table 5. All test results strongly reject equal coefficients, which further proves the necessity of Table 5 and related analysis.

Table 6 to 8 are derived from Table 5, using the same approach as for deriving Table 3. Table 6 shows similar patterns to Table 3. No matter for which subgroup of women, the impact of FPP becomes stronger and stronger over periods; no matter for which level of education, the impact of FPP is generally larger for urban or Han women. More importantly, less-educated women face stronger

restrictions from FPP, the story behind which could be that more-educated women tend to actively control their births, and are less likely to be pushed by FPP. Similarly, Table 7 implies that, a woman whose first child is son tends to actively control births, possibly because the mother is less likely to need to bear more children for enough sons, given the strong son preference in China. Table 8 shows that, women in richer provinces are less likely to be bound by FPP.

[Table 6, 7, 8 are inserted to here.]

To summarize, women with characteristics related to lower fertility, such as high education and high income, tend to actively control their births, and thus would face weaker restrictions from FPP, reflected by the smaller marginal effects of FPP.

(II) Impact of Family Planning Policy on Age of the First Marriage and First Birth

By the history, FPP did not only impact the number of children ever born, but also affected the age of marriage and thus the age of childbearing. Table 9 assesses how the age of the first marriage and first birth respond to FPP.

[Table 9 is inserted to here.]

The dependent variables are dummies indicating whether a woman's age of the first marriage or having the first child was below some level. Column [1] to [3] are based on the entire data including never-married women, and [4] to [6] restrict the data to ever-married women. Logit models are used through all columns, with the same set of independent variables as in Column [1] of Table 2. Age is dropped from Column [1] due to perfect collinearity. Marginal effects and their standard errors are reported. The impact of FPP on the probability of early marriage or childbearing will be further analyzed in Table 10 and 11. Other than FPP related variables, Table 9 shows strong negative correlation between education and the probability of early marriage or childbearing.

Following a similar approach as for Table 3, Table 10 shows the marginal effect of FPP on the probability of being married early,[56] with three notable features. First, for any group of women and any of the three definitions of early marriage, the power of FPP against early marriage is stronger and stronger over periods. Second, within any particular group of women, those who initially tend to get married late respond to FPP earlier. For example, the period-1 FPP mainly changed the marriage behaviour of the women who initially got married around 23 years old, but had not impact on the women married around 19 years old. Only the strictest one-child policy could change the marriage behaviour of the women who initially got married early. Third, urban or Han women tend to actively delay marriage, and thus face fewer restrictions from FPP on the age of marriage, than rural or non-Han women. Table 3 and 10 together may imply different cross-sectional policy variations in

[56] Table 10 and 11 only show the magnitude of marginal effect, without computing the level of significance, because of the lack of the variance-covariance matrix for the marginal effects.

terms of the number of children allowed and the age of marriage. According to the history and laws of FPP, the number of children allowed explicitly differs between urban and rural areas, and varies from Han to non-Han people, with significantly stronger policies for urban or Han people which is why we could see stronger impact of FPP for urban or Han women in Table 3. No evidence, however, supports clear policy variations in terms of the age of marriage across urban and rural areas, and Han and non-Han people.[57] That's why, similar to Table 6 to 8, urban or Han women face fewer restrictions from FPP in Table 10. Table 11 clearly exhibits the first two features. To summarize, FPP indeed reduced the probability of early marriage or childbearing, though to different degrees for different groups of women.

[Table 10 and 11 are inserted to here.]

VI. Conclusion and Further Study

The family planning policy (FPP) in China has been widely studied, yet the impact of FPP on fertility remains ambiguous. A large number of papers concluded that FPP explains a sizable proportion of fertility decline in China, while many other studies argued that, the role that China's FPP has played in population control has been overstated.

One source of the controversial conclusions is different measures of FPP, which might not only lead to various effects on fertility, but could also cause indefinite second-stage results in the studies where FPP is treated as an instrumental variable. Roughly, FPP can be measured by specific variables or constructed variables. The former comes from specific information on FPP of some specific data sets, such as the availability of family planning services, monetary penalties on "over-quota" births, and subsidies for one-child families. The latter, more widely used, is relatively easier to construct and relies on the history of FPP. Most previous measures, however, failed to completely capture the policy variations, did not sufficiently express people's heterogeneous exposures to FPP, and sometimes suffered from endogeneity.

This paper tries to construct a new measure of FPP that can improve on the three aspects. Historically, FPP varies over time, and differs between urban and rural areas, and between Han and non-Han people. My measure is going to embody the secular and cross-sectional policy variations as completely as possible. The basic idea of constructing the new measure comes from the intuition that a woman is more likely to be affected by the FPP of some period if the woman is more biologically capable to bear a child at her age during the period. The method of constructing the new measure implies a progress in heterogeneity and exogeneity.

[57] Current laws set different "birth quota" for different groups of people, while the recommended age of marriage is uniform across the nation.

A cross-sectional data of the China Health and Nutrition Survey (CHNS) is used to construc the measure, and to estimate the impact of FPP on the number of children ever born to a ever-married woman, and on the probability of early marriage and early childbearing. First, the empirical analysis supports the appropriateness of the newly constructed measure, because its empirical features well fit the history. Second, based on the new measure, FPP can explain a sizable part of fertility change in the data, and the age of first marriage and childbearing tends to be largely delayed under FPP. FPP variables are significant enough to be qualifed as instrumental variables for fertility. Third, the women who are more educated, living in richer provinces, or whose first child is son, tend to actively control their births and face fewer restrictions from FPP.

Shortcomings must be mentioned. First, FPP differs cross provinces. Since little information on the provincial-level policy variations is available to me, only variations across urban/rural and Han/non-Han are considered. Second, the sample used in the paper might have too few observations for older generations. A data set with more observations for those cohorts could make results more reliable. Third, income is not controlled in regressions, due to its endogeneity and measurement problem. But it may at the same time lead to an omitted variable bias, even though many other variables have been included. Last but not least, a quasi-separable form is used to measure FPP, and a more general form could be desirable.

Several directions can be further explored. First, the paper analyzes the impact of FPP on the total number of children ever born up to the date of interview. It could be preciser to see the *dynamic* effect of the FPP of some period on the childbearing behaviour in that specific period. Second, a thorough review on how different measures of FPP could affect its impact on fertility and the second stage regression results could be done. Third, based on the empirical results, it would be interesting to have a fertility model with FPP to theoretically and generally explain people's fertility behaviours under FPP.

References

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

[2] Cai, Yong. 2010. China's Below-Replacement Fertility: Government Policy or Socioeconomic Development? *Population and Development Review*, 36(3): 419–440

[3] Edlund, Lena, Hongbin Li, Junjian Yi and Junsen Zhang. 2008. More Men, More Crime: Evidence from China's One-Child Policy. Working Paper

[4]Islam, Asadul, and Russell Smyth. 2010. Children and Parental Health: Evidence from China. Working Paper

[5] Lavely, William, and Ronald Freedman. 1990. The Origins of the Chinese Fertility Decline. *Demography*, Vol. 27, No. 3, pp. 357-367

[6] Li, Hongbin, and Junsen Zhang. 2007. Do High Birth Rates Hamper Economic Growth? *The Review of Economics and Statistics*, vol. 89, issue 1, pages 110-117

[7] Li, Hongbin, Junsen Zhang. 2008. Testing the External Effect of Household Behavior: The Case of the Demand for Children. *Journal of Human Resources*, 44(4)

[8] 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

[9] McElroy, Marjorie and Dennis Tao Yang. 2000. Carrots and Sticks: Fertility Effects of China's Population Policies. *American Economic Review*, 2000, vol. 90, issue 2, pages 389-392

[10] Narayan, P. Kumar, and Xiujian Peng. 2006. An Econometric Analysis of the Determinants of Fertility for China, 1952--2000. *Journal of Chinese Economic and Business Studies*, vol. 4, issue 2, pages 165-183

[11] Olsen, Randall. 1994. Fertility and the Size of the U. S. Labor Force. *Journal of Economic Literature*, Vol. 32, No. 1, pp. 60-100

[12] Qian, Nancy. 2009. Quantity-Quality and the One Child Policy: The Only-Child Disadvantage in School Enrollment in Rural China. Working Paper

[13] Schultz, T. Paul, and Yi Zeng. 1995. Fertility of Rural China: Effects of Local Family Planning and Health Programs. *Journal of Population Economics*, vol. 8, issue 4, pages 329-50

[14] Wu, Xiaoyu, and lixing Li. 2011. Family Size and Maternal Health: Evidence from the One-Child Policy in China. *Journal of Population Economics*

[15] Yang, Faxiang. 2004. Historical Research on Family Planning of Contemporary China. Ph.D. Dissertation in History. Zhejiang University

[16] Yang, Faxiang. 2010. China's Fertility Policy: 1949-2009. Working paper

[17] Yang, D. T., and D. Chen. 2004. Transformations in China's Population Policies. *Pacific Economic Review*, 9: 3 (2004) pp. 269–290

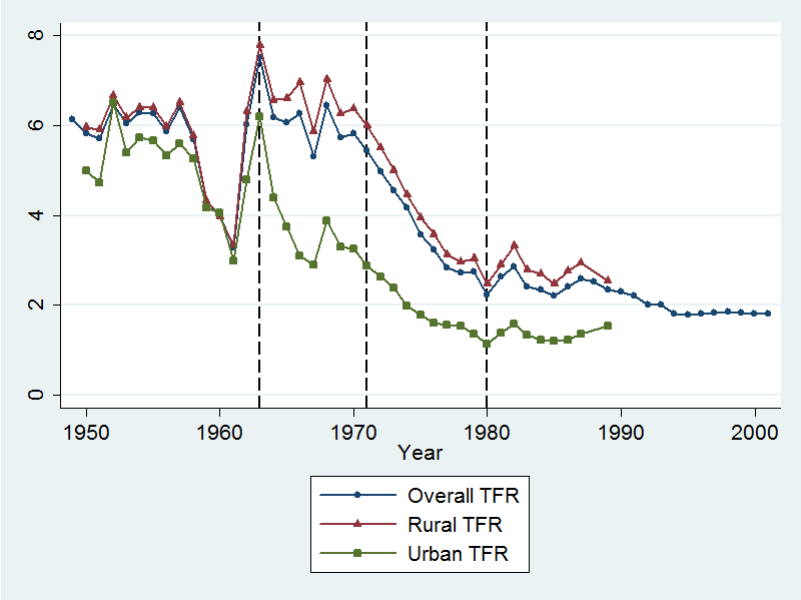
Appendix A: Figures and Tables

Table 1. Number of Children Allowed to Bear Over Periods and Across Groups of People[a]

	Period 0[b]	Period 1[b]	Period 2[b]	Period 3[b]
Urban Han	No restriction	Mild FPP allowing, though not encouraging, to have three children[c]	Strong FPP allowing, though not encouraging, to have two children	One-child policy allowing to have only one child
Rural Han	No restriction	No restriction	Milder policy than the upper cell	One-child policy conditionally allowing to have two children
Urban Non-Han	No restriction	No restriction[d]	No restriction[d]	Specific FPP conditionally allowing to have two children
Rural Non-Han	No restriction	No restriction	No restriction[d]	Specific FPP conditionally allowing to have three or even more children

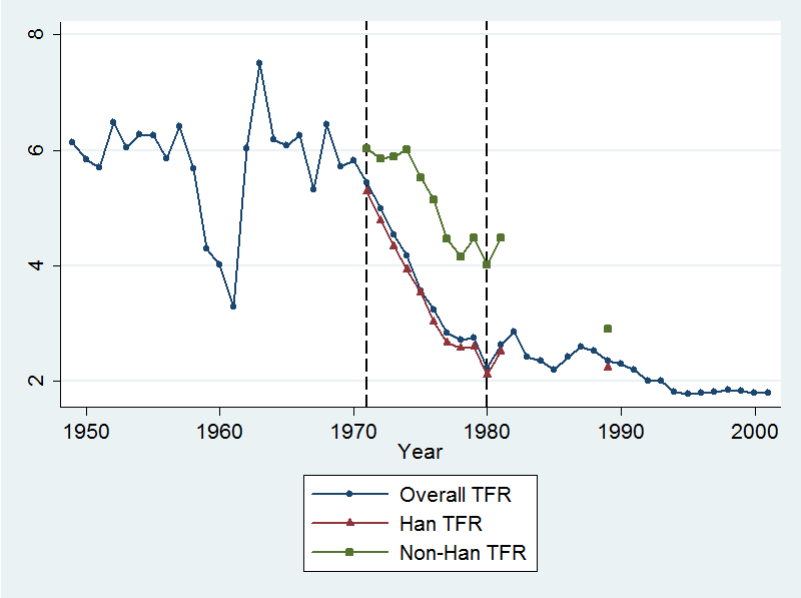
Notes: [a] The table is summarized based on the history introduced in Section II. Other content of FPP, such as the age of marriage, is not listed in the table.
[b] As stated at the beginning of Section II, period 0 to 3 indicate the period without FPP (1949-1962), the period with mild and narrowly implemented FPP (1963-1970), the period with strong and widely implemented FPP (1971-1979), and the period of the one-child policy (1980-now).
[c] The urban Han people living in the five ethnic minority regions were not covered by the FPP in period 1.
[d] Part of the non-Han people might be affected by FPP, particularly when they were not living in the five minority regions, and/or identified the ethnicity of their children as Han.

Figure 1. Urban and Rural Total Fertility Rates Over Time



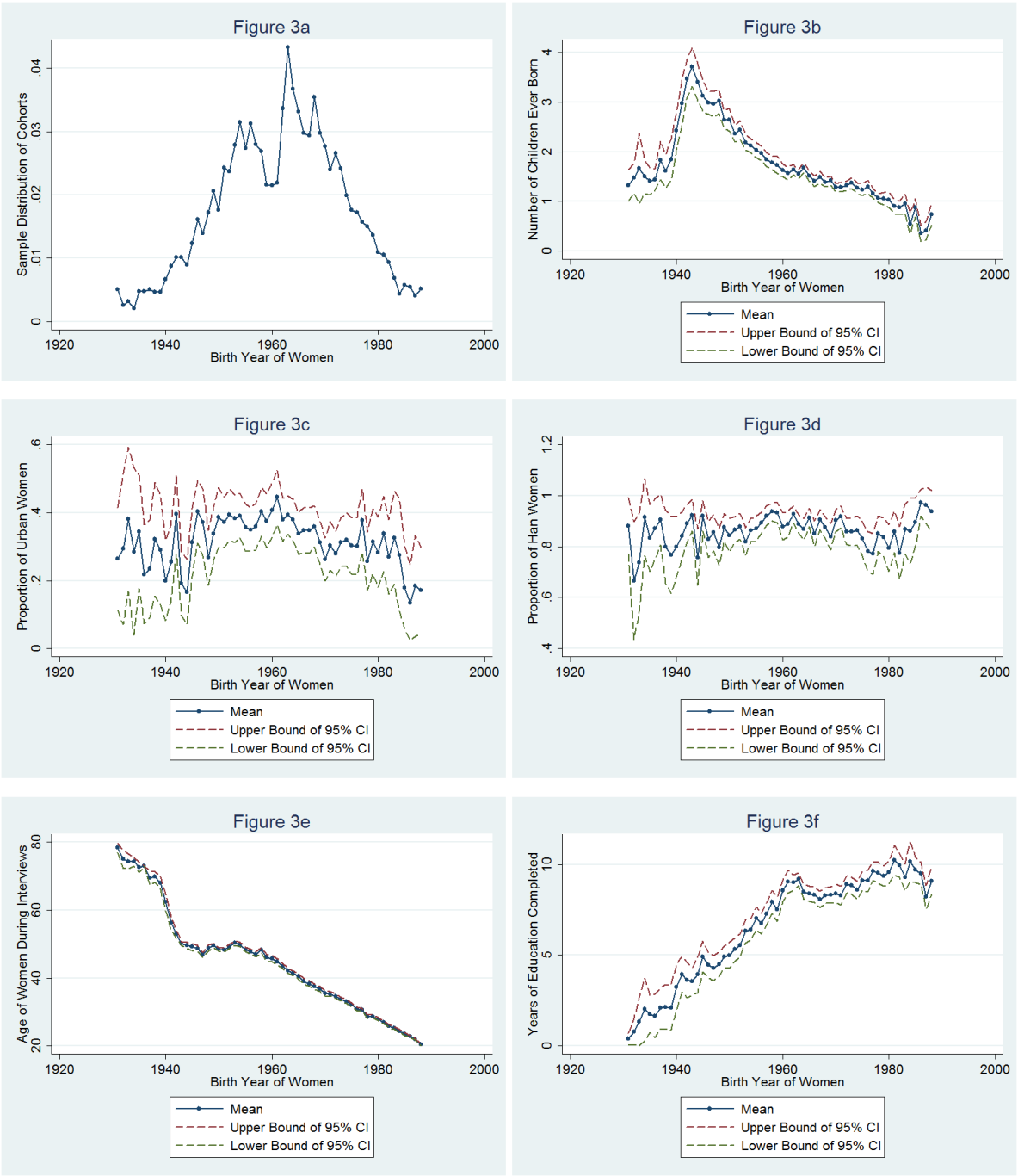
Notes. The overall total fertility rate is cited from Yang (2004, page 264-265). The urban and rural total fertility rates are cited from Yang (2004, page 134, 135, 139). The left, middle and right dashed lines mark the years of 1963, 1971 and 1980, the starting points of the period 1 to 3 family planning policies.

Figure 2. Han and Non-Han Total Fertility Rates Over Time



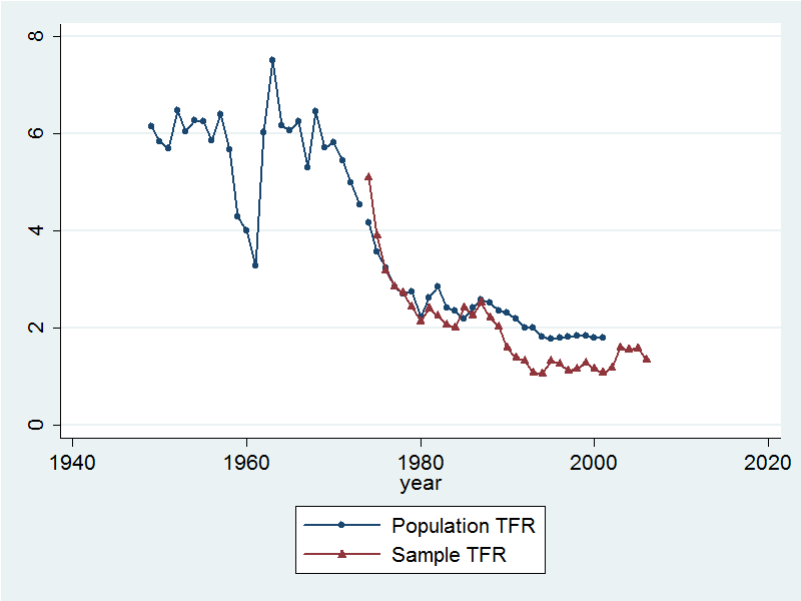
Notes. The overall total fertility rate is cited from Yang (2004, page 264-265). The Han (ethnic majority) and non-Han (ethnic minority) total fertility rates are cited from Yang (2004, page 145, 150). The left and right dashed lines mark the years of 1971 and 1980, the starting points of the period 2 and 3 family planning policies.

Figure 3. Descriptive Statistics of the Sample and Key Variable



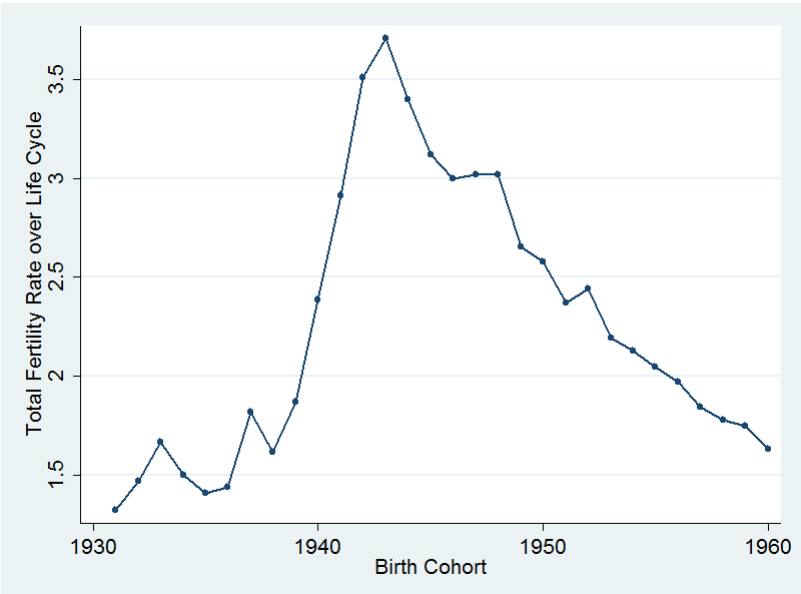
Notes. All the X-axes indicate the birth year of women. Due to the small sample size of the oldest and youngest cohorts, the cohorts before 1931 and after 1988 are merged to cohort 1931 and 1988, respectively, to make the lines smoother. Figure 3a is the proportion of sample size to each cohort. Figure 3b is the average number of children ever born to the women of each cohort. Figure 3c is the proportion of urban women of each cohort. Figure 3d is the proportion of Han women of each cohort. Figure 3e is the average age of a woman during her interview to each cohort. Figure 3f is the average completed years of education for women of each cohort. Dashed lines are the upper and lower bounds of the 95% confidence intervals. Figure 3 calculates statistics based on the ever-married women only. After including the never-married women, the figures presents similar features, except a little bit change for the youngest cohorts.

Figure 4. Population and Sample Total Fertility Rates Over Time



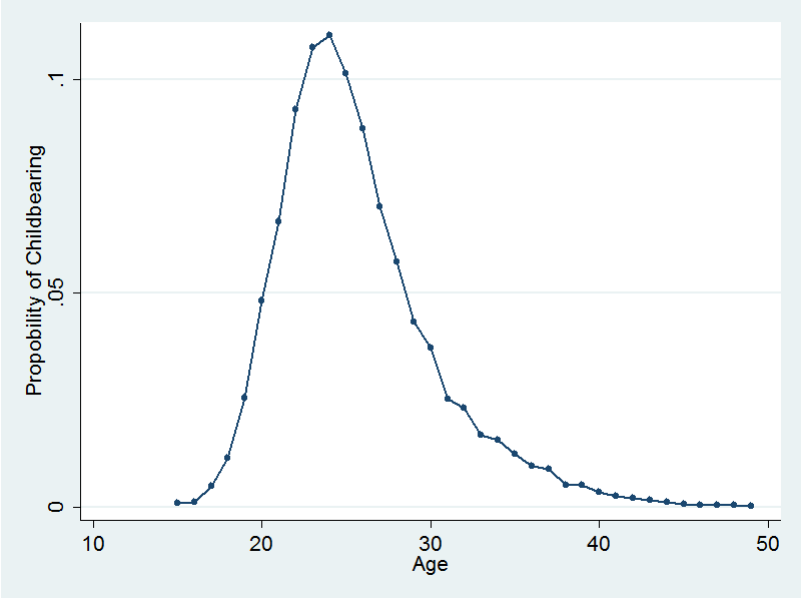
Notes. The population total fertility rate is cited from Yang (2004, page 264-265), and the sample total fertility rate is calculated by the author with the formula in equation (1). Given that the sample ranges from cohort 1925 to 1991 and the calculation of the total fertility rate requires the fertility information for the women aged from 15 to 49, the sample total fertility therefore can only cover 1974 to 2006.

Figure 5. Sample Total Fertility Rate Over Cohort



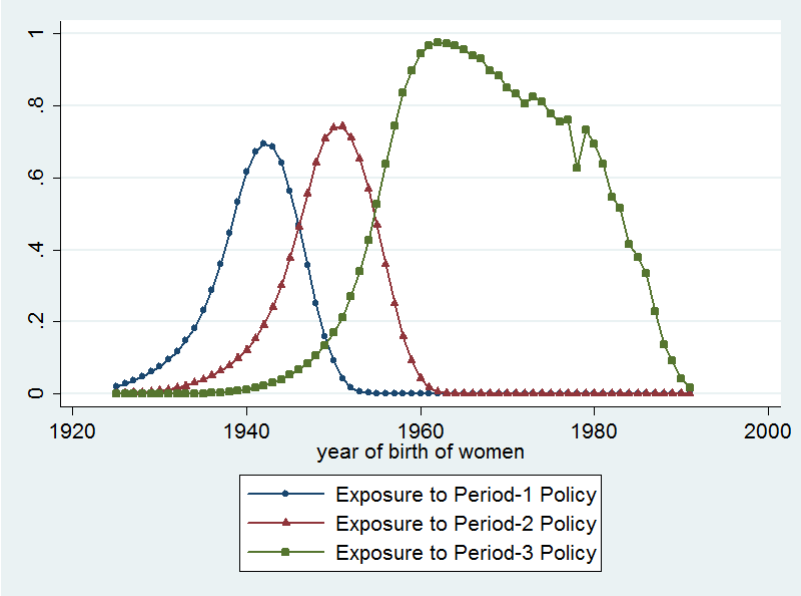
Notes. The cohort-adjusted sample total fertility rate is calculated by the author with the formula in equation (2). The cohorts older than the 1931 cohort were merged to the 1931 cohort. Since the last round of survey occurred in 2009 by which the women born after 1960 had not biologically completed fertility, the total fertility rate can only be calculated up to cohort 1960.

Figure 6. Probability Distribution of Childbearing Over Age



Notes. The probability of childbearing at an age is calculated as the ratio of the total number of children born to the women at that age and the total number of children ever born in the entire data. The distribution takes positive values from age 15 to 49, and 0 for other ages. Clearly, probabilities sum up to 1.

Figure 7. Exposure to Family Planning Policy Over Cohort



Notes. The exposure to the three periods of family planning policies are calculated based on equation (5). Clearly, the 1940s, 1950s, and 1960s cohorts are mostly exposed to the period-1, period-2 and period-3 policy, respectively. For the one-child policy is ongoing, the interval exposed to it for the younger cohorts is the interval between age 15 (or the age at which they started to be affected by the policy) and the age during the lastest interview.

Table 2. Impact of Family Planning Policy on Fertility

	Dependent Variable: Number of Children Ever Born				Y = 1 If Number of Children Ever Born >= 2	Y = 1 If Number of Children Ever Born >= 3
	[1]-OLS	[2]-OLS	[3]-OLS	[4]-OLS	[5]-LOGIT	[6]-LOGIT
FPP_1	-0.820**	-0.767*			-0.451	-0.283
	[0.408]	[0.412]			[0.432]	[0.188]
FPP_2	-0.457	-0.533*			0.17	-0.301**
	[0.299]	[0.303]			[0.317]	[0.141]
FPP_3	-0.802***	-0.825***			-0.0485	-0.292**
	[0.222]	[0.223]			[0.232]	[0.123]
$D_{urban}*FPP_1$	-0.628***	-0.867*	-0.574***	-0.877	0.368*	-0.0614
	[0.202]	[0.523]	[0.214]	[0.548]	[0.224]	[0.0912]
$D_{urban}*FPP_2$	-0.940***	-0.556**	-0.922***	-0.455*	-0.337**	-0.150**
	[0.130]	[0.260]	[0.137]	[0.265]	[0.133]	[0.0732]
$D_{urban}*FPP_3$	-0.384***	-0.293*	-0.376***	-0.299*	-0.0869	-0.108
	[0.128]	[0.157]	[0.136]	[0.164]	[0.131]	[0.0757]
$D_{Han}*FPP_1$	-0.132	-0.196	-0.0388	-0.116	-0.275	-0.0593
	[0.253]	[0.261]	[0.268]	[0.275]	[0.320]	[0.115]
$D_{Han}*FPP_2$	-0.686***	-0.592***	-0.531***	-0.418**	-0.569***	-0.0912
	[0.163]	[0.172]	[0.173]	[0.181]	[0.188]	[0.0835]
$D_{Han}*FPP_3$	-0.319**	-0.290*	-0.178	-0.154	-0.257	-0.0676
	[0.158]	[0.161]	[0.172]	[0.174]	[0.160]	[0.0870]
$D_{urban}*D_{Han}*FPP_1$		0.27		0.341		
		[0.516]		[0.539]		
$D_{urban}*D_{Han}*FPP_2$		-0.424*		-0.517**		
		[0.251]		[0.254]		
$D_{urban}*D_{Han}*FPP_3$		-0.0981		-0.082		

		[0.0980]		[0.0985]		
D_{urban}	0.119	0.115	0.104	0.101	-0.172	0.042
	[0.116]	[0.116]	[0.124]	[0.124]	[0.120]	[0.0851]
D_{Han}	0.383***	0.380***	0.264*	0.260*	0.232*	0.0528
	[0.142]	[0.142]	[0.155]	[0.155]	[0.132]	[0.0481]
Years of Education	-0.0675***	-0.0672***	-0.0661***	-0.0658***	-0.0602***	-0.0162***
	[0.00309]	[0.00309]	[0.00313]	[0.00313]	[0.00286]	[0.00135]
Demeaned Years of Education^2/2	0.00107	0.00106	0.000966	0.000958	-0.00589***	-0.00177***
	[0.00109]	[0.00110]	[0.00111]	[0.00111]	[0.00116]	[0.000438]
Age	-0.0249	-0.0147	0.0629**	0.0628**	0.106	0.0465
	[0.137]	[0.137]	[0.0245]	[0.0245]	[0.159]	[0.0522]
Demeaned Age^2/2	-0.00527	-0.00556	-0.00712***	-0.00709***	-0.00666	-0.00258*
	[0.00392]	[0.00392]	[0.00257]	[0.00257]	[0.00439]	[0.00148]
Exposure to 1959-1961 Famine	-2.863***	-2.861***			-0.960*	-0.36
	[0.636]	[0.636]			[0.563]	[0.316]
Exposure to Pre-PRC Period	7.430***	7.432***			1.553	2.89
	[2.319]	[2.319]			[3.117]	[2.065]
Other Control Variables	A	A	B	B	A	A
Constant	5.484	4.91	3.909***	3.921***		
	[7.920]	[7.924]	[1.155]	[1.155]		
Observations	6,353	6,353	6,353	6,353	6,353	6,353
R-squared	0.509	0.509	0.536	0.536		
F Statistics for Joint Significance of FPP	15.98	12.39	14.47	10.31		
Pseudo R-squared					0.369	0.404
Chi-squared Statistics for Joint Significance of FPP					36.33	31.87

Notes. Standard errors are in squared brackets. *** p<0.01, ** p<0.05, * p<0.1. "Other Control Variables A" includes birth year trend, (birth year trend)^2, (birth year trend)^3, (birth year trend)^4, round dummies, birth year trend*round dummies, and province dummies. "Other Control Variables B" includes birth year dummies, round dummies, birth year dummies*round dummies, and province dummies. FPP_1-FPP_3, Exposure to 1959-1961 Famine, and Exposure to Pre-PRC Period are dropped in Column [3] and [4], because they are perfectly collinear with birth year dummies, round dummies and birth year dummies*wave dummies. R-squared and Pseudo R-squared are reported for OLS and logit regressions, respectively. The F statistics for the joint significance test of all the FPP related variables are reported for the OLS regressions, and the Chi-squared statistics for the joint significance test of all the FPP related variables are reported for the logit regressions. All the p values related to the F and Chi-squared statistics are 0.0000.

Table 3. Marginal Effect of Three Periods of Family Planning Policy on Number of Children Ever Born to Four Groups of Women

	Period 0	Period 1	Period 2	Period 3
Urban Han	Benchmark of this Row	-0.354***	-0.646***	-0.985***
Rural Han	Benchmark of this Row	-0.214**	-0.355***	-0.734***
Urban Non-Han	Benchmark of this Row	-0.325***	-0.433***	-0.777***
Rural Non-Han	Benchmark of this Row	-0.184**	-0.142	-0.525***

Notes. The number for group i (row) and period j (column) women represents the number of *fewer* children a group i woman would have if she was to an average degree exposed to the period-j FPP, compared to not being affected by any FPP as in period 0. In other words, period 0 is the benchmark of each row. The average degree of exposure to period-j FPP is expressed as assigning FPP_j its average value based on the women having positive exposure to period-j FPP. Taking the rural Han woman in period 2 as an example: -0.355 = (the coefficient of FPP_2 + the coefficient of D_{Han}*FPP_2) * the sample mean of FPP_2 = (-0.457 + (- 0.686)) * 0.3101. All the coefficients are from Column [1] of Table 2. *** p<0.01, ** p<0.05, * p<0.1.

Table 4. Difference of Marginal Effect of Family Planning Policy between Groups of Women within Each Period

Rural Non-Han as Benchmark			
	Period 1	**Period 2**	**Period 3**
Urban Han	-0.058	-0.349***	-0.132***
Rural Han	0.056	-0.094***	0.042
Urban Non-Han	-0.114***	-0.255***	-0.174***
Urban Non-Han as Benchmark			
	Period 1	**Period 2**	**Period 3**
Urban Han	0.056	-0.094***	0.042
Rural Han	0.170 ***	0.161***	0.216***
Rural Han as Benchmark			
	Period 1	**Period 2**	**Period 3**
Urban Han	-0.114***	-0.255***	-0.174***

Notes. A number in a panel represents the difference of the marginal effect of FPP between the row group and the benchmark group in that panel within the column period. Calculation approach is similar to Table 3. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Impact of Family Planning Policy on Number of Children Ever Born for Subsamples

	Dependent Variable: Number of Children Ever Born					
	[1] Completed Years of Education<9	[2] Completed Years of Education>=9	[3] The First Child is Son	[4] The First Child is Daughter	[5] Eastern Coastal Provinces	[6] Non-Eastern Coastal Provinces
FPP_1	-1.629**	2.146***	-0.921*	-1.411**	1.176	-1.516***
	[0.665]	[0.781]	[0.532]	[0.646]	[0.746]	[0.512]
FPP_2	-1.200**	0.684	-0.527	-1.463***	0.531	-1.097***
	[0.475]	[0.425]	[0.388]	[0.478]	[0.497]	[0.382]
FPP_3	-0.872**	-0.253	-0.976***	-1.427***	-0.00192	-1.210***
	[0.343]	[0.326]	[0.283]	[0.369]	[0.410]	[0.275]
$D_{urban}*FPP_1$	-0.386	-1.201***	-0.669**	-1.008***	-0.693**	-0.594**
	[0.302]	[0.295]	[0.270]	[0.325]	[0.323]	[0.259]
$D_{urban}*FPP_2$	-0.595***	-0.977***	-0.859***	-1.256***	-1.016***	-0.974***
	[0.198]	[0.169]	[0.175]	[0.212]	[0.224]	[0.161]
$D_{urban}*FPP_3$	-0.379*	-0.443***	-0.353**	-0.795***	-0.465**	-0.424***
	[0.202]	[0.158]	[0.173]	[0.214]	[0.228]	[0.154]
$D_{Han}*FPP_1$	-0.239	-0.559	-0.0354	-0.302	-1.638***	0.36
	[0.336]	[0.553]	[0.332]	[0.395]	[0.570]	[0.296]
$D_{Han}*FPP_2$	-0.535**	-0.658**	-0.999***	-0.428	-0.986***	-0.389**
	[0.222]	[0.259]	[0.204]	[0.266]	[0.344]	[0.192]
$D_{Han}*FPP_3$	-0.412*	-0.0791	-0.512***	-0.156	-0.740**	-0.083
	[0.227]	[0.230]	[0.198]	[0.265]	[0.343]	[0.186]
D_{urban}	-0.0361	0.211	0.114	0.482**	0.198	0.153
	[0.177]	[0.146]	[0.159]	[0.197]	[0.212]	[0.139]
D_{Han}	0.581***	0.0415	0.604***	0.214	0.615*	0.204
	[0.197]	[0.209]	[0.179]	[0.240]	[0.319]	[0.164]
Years of Education	-0.111***	-0.105***	-0.0613***	-0.0717***	-0.0455***	-0.0812***
	[0.0216]	[0.0208]	[0.00402]	[0.00453]	[0.00439]	[0.00403]
Demeaned Years of Education^2/2	-0.0112**	0.0104**	0.00397***	-0.000283	0.00374**	-0.000849

	[0.00511]	[0.00494]	[0.00143]	[0.00164]	[0.00151]	[0.00145]
Age		-0.387	0.721***	-0.192		0.0267
		[0.243]	[0.178]	[0.218]		[0.181]
Demeaned Age^2/2	-0.00646***	0.00762	-0.0260***	-0.00106	-0.00456***	-0.00715
	[0.000451]	[0.00670]	[0.00510]	[0.00615]	[0.000577]	[0.00519]
Exposure to 1959-1961 Famine	-3.263***	-2.706	-2.430***	-3.750***	-1.803*	-3.901***
	[1.035]	[2.678]	[0.824]	[0.966]	[0.957]	[0.813]
Exposure to Pre-PRC Period	8.999***		9.941***	8.334**	11.12***	5.947*
	[3.326]		[2.788]	[3.850]	[2.845]	[3.457]
Constant	4.054***	26.38*	-38.07***	15.01	2.668***	3.289
	[0.893]	[14.88]	[10.31]	[12.68]	[0.825]	[10.44]
Observations	3,029	3,324	3,115	2,885	2,092	4,261
R-squared	0.483	0.357	0.523	0.492	0.477	0.519
F Statistics for Joint Significance of FPP	3.981	11.5	13.29	8.876	6.973	11.16
F Statistics for Equal Coefficients Test	6.32	6.32	4.67	4.67	12.24	12.24

Notes. Standard errors are in squared brackets. *** p<0.01, ** p<0.05, * p<0.1. Other control variables include birth year trend, (birth year trend)^2, (birth year trend)^3, (birth year trend)^4, round dummies, birth year trend*round dummies, and province dummies. Age is dropped in some columns, because it's perfectly collinear with birth year trend, round dummies and birth year trend*round dummies in those subsamples. Exposure to Pre-PRC Period is dropped in Column [2], because of little variation in that subsample. Eastern coastal provinces are Jiangsu, Liaoning and Shandong. All p values related to the F statistics for equal coefficients test and for joint significance test are 0.0000.

Table 6. Marginal Effect of Three Periods of Family Planning Policy on Number of Children Ever Born for Groups of Women (Subgrouped by Education)

	Years of Education	Period 0	Period 1	Period 2	Period 3
Urban Han	< 9	Benchmark of this Row	-0.572***	-0.847***	-0.857***
	>= 9	Benchmark of this Row	0.055	-0.228**	-0.601***
Rural Han	< 9	Benchmark of this Row	-0.474***	-0.631***	-0.662***
	>= 9	Benchmark of this Row	0.225**	0.006	-0.257
Urban Non-Han	< 9	Benchmark of this Row	-0.511***	-0.653***	-0.645***
	>= 9	Benchmark of this Row	0.134	-0.070	-0.539**
Rural Non-Han	< 9	Benchmark of this Row	-0.413**	-0.436**	-0.450**
	>= 9	Benchmark of this Row	0.304***	0.164	-0.196

Notes. Calculation approach is similar to that for Table 3. Just need to replace the sample average of FPP_j with subsample average of FPP_j.*** p<0.01, ** p<0.05, * p<0.1.

Table 7. Marginal Effect of Three Periods of Family Planning Policy on Number of Children Ever Born for Groups of Women (Subgrouped by the Gender of the First Child)

	Gender of the First Child	Period 0	Period 1	Period 2	Period 3
Urban Han	Son	Benchmark of this Row	-0.358***	-0.733***	-1.216***
	Daughter	Benchmark of this Row	-0.605***	-0.983***	-1.568***
Rural Han	Son	Benchmark of this Row	-0.211*	-0.469***	-0.983***
	Daughter	Benchmark of this Row	-0.381***	-0.591***	-1.044***
Urban Non-Han	Son	Benchmark of this Row	-0.350***	-0.426***	-0.878***
	Daughter	Benchmark of this Row	-0.538***	-0.850***	-1.466***
Rural Non-Han	Son	Benchmark of this Row	-0.203*	-0.162	-0.645***
	Daughter	Benchmark of this Row	-0.314**	-0.457***	-0.941***

Notes. Calculation approach is similar to that for Table 3. Just need to replace the sample average of FPP_j with subsample average of FPP_j.*** p<0.01, ** p<0.05, * p<0.1.

Table 8. Marginal Effect of Three Periods of Family Planning Policy on Number of Children Ever Born for Groups of Women (Subgrouped by Geographic Location)

	Geographic Location	Period 0	Period 1	Period 2	Period 3
Urban Han	Coastal Provinces	Benchmark of this Row	-0.096*	-0.295***	-0.774***
	Non-Coastal Provinces	Benchmark of this Row	-0.393***	-0.767***	-1.120***
Rural Han	Coastal Provinces	Benchmark of this Row	-0.038	-0.091	-0.476**
	Non-Coastal Provinces	Benchmark of this Row	-0.260**	-0.464***	-0.843***
Urban Non-Han	Coastal Provinces	Benchmark of this Row	0.040	-0.097	-0.299
	Non-Coastal Provinces	Benchmark of this Row	-0.474***	-0.646***	-1.066***
Rural Non-Han	Coastal Provinces	Benchmark of this Row	0.098	0.107	-0.001
	Non-Coastal Provinces	Benchmark of this Row	-0.340***	-0.342***	-0.789***

Notes. Calculation approach is similar to that for Table 3. Just need to replace the sample average of FPP_j with subsample average of FPP_j.*** p<0.01, ** p<0.05, * p<0.1.

Table 9. Impact of Family Planning Policy on the Age of First Marriage and First Birth

	[1]-LOGIT	[2]-LOGIT	[3]-LOGIT	[4]-LOGIT	[5]-LOGIT	[6]-LOGIT
	Y=1 If Age of First Marriage <= 19	Y=1 If Age of First Marriage <= 21	Y=1 If Age of First Marriage <= 23	Y=1 If Age of First Birth <= 20	Y=1 If Age of First Birth <= 22	Y=1 If Age of First Birth <= 24
FPP_1	0.264	-0.153	-1.155***	0.0684	-0.22	-1.658***
	[0.172]	[0.356]	[0.402]	[0.156]	[0.321]	[0.449]
FPP_2	-0.0189	-0.640**	-1.754***	-0.166	-0.965***	-2.199***
	[0.134]	[0.264]	[0.311]	[0.111]	[0.233]	[0.357]
FPP_3	-0.436***	-1.292***	-1.860***	-0.534***	-1.530***	-2.258***
	[0.0927]	[0.192]	[0.255]	[0.0738]	[0.175]	[0.301]
$D_{urban}*FPP_1$	0.0433	0.165	0.277*	-0.0285	-0.1	0.278
	[0.0618]	[0.130]	[0.147]	[0.0622]	[0.139]	[0.199]
$D_{urban}*FPP_2$	0.0588	0.0963	0.256**	-0.0345	-0.148	0.0208
	[0.0450]	[0.0884]	[0.100]	[0.0454]	[0.104]	[0.152]
$D_{urban}*FPP_3$	0.0535	0.159*	0.316***	-0.0281	-0.108	0.183
	[0.0429]	[0.0844]	[0.105]	[0.0451]	[0.106]	[0.161]
$D_{Han}*FPP_1$	0.141*	0.453***	0.549**	0.107	0.163	0.14
	[0.0732]	[0.172]	[0.219]	[0.0739]	[0.165]	[0.250]
$D_{Han}*FPP_2$	0.0944*	0.121	0.189	0.0586	0.0831	-0.173
	[0.0522]	[0.119]	[0.157]	[0.0511]	[0.121]	[0.190]
$D_{Han}*FPP_3$	0.0903*	0.238**	0.394**	0.0152	0.0812	-0.0713
	[0.0498]	[0.118]	[0.168]	[0.0487]	[0.123]	[0.203]
D_{urban}	-0.0742**	-0.176**	-0.354***	-0.00453	0.0512	-0.216
	[0.0322]	[0.0689]	[0.101]	[0.0401]	[0.101]	[0.150]
D_{Han}	-0.106	-0.201*	-0.237***	-0.00543	-0.0563	0.0909
	[0.0679]	[0.112]	[0.0770]	[0.0450]	[0.121]	[0.194]
Years of Education	-0.0219***	-0.0424***	-0.0321***	-0.0172***	-0.0361***	-0.0376***
	[0.00137]	[0.00222]	[0.00187]	[0.00138]	[0.00210]	[0.00218]
Demeaned Years of Education^2/100	-0.00339***	-0.00599***	-0.00434***	-0.00275***	-0.00524***	-0.00525***

	[0.000497]	[0.000802]	[0.000656]	[0.000461]	[0.000754]	[0.000779]
Age		-0.135	-0.105	0.0304	-0.003	-0.0807
		[0.0929]	[0.0929]	[0.0418]	[0.0832]	[0.105]
Demeaned Age^2/100	-0.000794***	0.00135	0.000744	-0.00198	-0.00282	-0.00139
	[0.000267]	[0.00281]	[0.00277]	[0.00123]	[0.00243]	[0.00303]
Exposure to 1959-1961 Famine	0.291	-0.556	-1.933**	0.198	-0.853	-3.503***
	[0.392]	[0.873]	[0.841]	[0.404]	[0.855]	[1.021]
Exposure to Pre-PRC Period	-1.085	-2.541	1.496	0.328	1.948	4.552
	[0.989]	[2.513]	[2.560]	[1.322]	[2.952]	[3.334]
Observations	6,283	6,175	6,071	6,311	6,276	6,230
Pseudo R-squared	0.128	0.121	0.13	0.125	0.133	0.168
Chi-square Statistics For Joint Significance of FPP	67.61	107	133.1	101.4	150.9	138

Notes. All coefficients are marginal effects. Standard errors are in squared brackets. *** p<0.01, ** p<0.05, * p<0.1. Other control variables include birth year trend, (birth year trend)^2, (birth year trend)^3, (birth year trend)^4, round dummies, birth year trend*round dummies, and province dummies. Age is dropped in Column [1] because of perfect collinearity. All p values related to the Chi-squared statistics are 0.0000.

Table 10. Marginal Effect of Family Planning Policy on the Probability of Early Marriage

	Age of First Marriage	Period 0	Period 1	Period 2	Period 3
Urban Han	<= 19	Benchmark of this Row	0.101	0.041	-0.182
	<= 21	Benchmark of this Row	0.104	-0.130	-0.558
	<= 23	Benchmark of this Row	-0.074	-0.404	-0.718
Rural Han	<= 19	Benchmark of this Row	0.091	0.023	-0.216
	<= 21	Benchmark of this Row	0.067	-0.160	-0.658
	<= 23	Benchmark of this Row	-0.136	-0.483	-0.915
Urban Non-Han	<= 19	Benchmark of this Row	0.069	0.012	-0.239
	<= 21	Benchmark of this Row	0.003	-0.168	-0.707
	<= 23	Benchmark of this Row	-0.197	-0.462	-0.964
Rural Non-Han	<= 19	Benchmark of this Row	0.059	-0.006	-0.272
	<= 21	Benchmark of this Row	-0.034	-0.197	-0.807
	<= 23	Benchmark of this Row	-0.259	-0.541	-1.161

Notes. Given one of the four groups of women and one of the three definitions of early marriage, the number in the column of period j is the marginal effect of period-j FPP on the probability of early marriage to that group of women and under that definition of early marriage, compared to a situation without FPP as in period 0. Levels of significance for the marginal effect are not reported, due to the lack of the variance-covariance matrix of the marginal effects.

Table 11. Marginal Effect of Family Planning Policy on the Probability of Early Childbearing

	Age of First Childbearing	Period 0	Period 1	Period 2	Period 3
Urban Han	<= 20	Benchmark of this Row	0.033	-0.044	-0.358
	<= 22	Benchmark of this Row	-0.035	-0.320	-1.019
	<= 24	Benchmark of this Row	-0.278	-0.729	-1.406
Rural Han	<= 20	Benchmark of this Row	0.039	-0.033	-0.340
	<= 22	Benchmark of this Row	-0.013	-0.274	-0.949
	<= 24	Benchmark of this Row	-0.340	-0.736	-1.525
Urban Non-Han	<= 20	Benchmark of this Row	0.009	-0.062	-0.368
	<= 22	Benchmark of this Row	-0.072	-0.345	-1.073
	<= 24	Benchmark of this Row	-0.309	-0.676	-1.359
Rural Non-Han	<= 20	Benchmark of this Row	0.015	-0.051	-0.350
	<= 22	Benchmark of this Row	-0.049	-0.299	-1.002
	<= 24	Benchmark of this Row	-0.372	-0.682	-1.479

Notes. Given one of the four groups of women and one of the three definitions of early childbearing, the number in the column of period j is the marginal effect of period-j FPP on the probability of early childbearing to that group of women and under that definition of early childbearing, compared to a situation without FPP as in period 0. Levels of significance for the marginal effect are not reported, due to the lack of the variance-covariance matrix of the marginal effects.

Appendix B. Notes on the Construction of Data and Variables

B1. The Data of Ever-married Women with Experience of Childbearing

The birth history of ever-married women with experience of childbearing is stored in the data named "m10birth", released on the official website of the CHNS in July, 2011. This data contains the birth history up to the latest round of interview of each such ever-married woman (the possible latest-rounds don't include 1989 and 1997). The original data is in a panel form with the women and their children as the two dimensions. I construct the variable of the number of children ever born to a woman simply by counting the number of records of children to that woman. I reshaped the data to a cross-sectional sample.

Two forms of the birth date of each child were recorded, the western date of birth and the Chinese lunar date of birth. I use the former one. For the observation with missing western date of birth, I filled in an inferred date by adding one month to the Chinese lunar date of birth. Since the date of birth of children is essential to the analysis in terms of ordering the births and calculating the age of childbearing for mothers, if *some* child of a woman had missing value on the date of birth, I dropped the dates of birth of *all the other* children of that woman (only 26 observations were dropped). The gender of the first child was kept for the analysis related to son preference.

B2. The Data of Ever-married Women without Experience of Childbearing

The information of ever-married women who have never born a child is stored in a panel data of ever-married women, named "m10emw". To match the data in B1, I dropped the observations in round 1989 and 1997, and only kept the latest round of observation for each woman. If the latest round of a woman is 1991, the women without experience of childbearing can be targeted by looking at question S39, "Have you ever given birth to a child (even a child who lived only a few seconds)?" For round 1993, 2000, 2004, and 2006, a short form of questions about birth history were asked, stored in "m10emw", before the detailed birth history were recorded as in B1. The questions are S122 ("In total, how many children have you given birth to in your life?") and S47A ("According to my record, you have altogether given birth to __ children. Is that figure true?"). Then the women without experience of childbearing can be easily targeted. In round 2009, the women with experience of childbearing were directly interviewed about their detailed birth history without being asked the short form of questions. Therefore, the women without experience of childbearing were identified as the women appearing in "m10emw" but not included in "m10birth".

I merged the data in B1 and B2 to form a complete ever-married women sample. Some women did not answer the detailed birth history even though they had experience of childbearing according to question S122 or S47A. I treated these women as the ever-married women with experience of childbearing, but any analysis based on the birth history details, such as the date of birth of children, would not include these women.

B3. The Age of First Marriage for Ever-married Women

In the data "m10wed", it is easy to find the year and month of first marriage for each ever-married woman, even though they were re-married or widowed during the interview. In round 1989, women were directly asked a question about the age of first marriage, question A9, stored in the data "m10rst". The answer to question A9 from some woman was used to replace the missing information about the first marriage in "m10wed" for that woman.

The never-married women were included for the analysis on the age of first marriage. For the data "m10rst", I first dropped round 1989 and 1997, then kept the latest round for each woman, and finally deleted the women aged below 15 during the interview, to match the data of ever-married women. I targeted the never-married women by looking at question A8. The never-married women were then appended to the ever-married women. All women aged below 15, or born before 1925 or after 1991, were dropped (about 30 observations).

B4. The Latest Round

The information about the latest round is used to construct the FPP measure, and serves as necessary control variables in regressions. The data "m10birth" kept the latest round of observation for each woman but didn't tell which round it is.

Other than the birth history data "m10birth", the entire ever-married women data set also includes other parts, "m10emw", "m10wed", "m10preg", and "m10media". I went through these parts, dropped round 1989 and 1997, and kept the latest round for each woman in each part of the data. Then, I combined these parts, selected the latest of the lastest rounds of different parts, and merged them to the data of ever-married women constructed in B1 and B2.

B5. The Construction of Key Variables

The date of birth of women were handled as in B1. The urban dummy was defined based on if the interviewee was living in urban areas during the interview, but not on the *hukou* status. The variables above come from the data "c10mast".

The variable of education was derived from the data "m10educ". The original data provides the level of education, variable A11, and I transformed it to the completed years of education in the following way.

Level of Education	Illiterate	Primary School Grade 1-6	Junior High School Grade 1-3	Senior High School Grade 1-3	Training High School Grade 1-3	College
Completed Years of Education	0	1-6	7-9	10-12	10-12	12+years in college

Some women didn't answer question A11, but answered A12, the highest level of education attained. I replace the completed years of education for these women with the average completed years of education of other women at *the same* highest level of education. If both variables are missing, I used the education information in previous rounds to proxy the years of education in the latest round.