THE UNIVERSITY OF CHICAGO

AN ESSAY ON THE HOUSEHOLD- AND AGGREGATE-LEVEL FERTILITY CONSEQUENCES OF CHINA'S ONE-CHILD POLICY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARMENT OF ECONOMICS

BY

JORGE LUIS GARCIA GARCIA MENENDEZ

CHICAGO, ILLINOIS
JUNE 2018

Dedication

To my parents, to my sister and Hugo, and to Anna.

Contents

List of Tables

List of Figures

Acknowledgements

Special thanks to Stéphane Bonhomme, James J. Heckman, and Alessandra Voena for patient guidance as part of my dissertation committee. I also thank funding from a Conacyt fellowship granted by the Mexican government; from Sherwin Rosen, Esther and T.W. Schultz, Immasche, and social sciences fellowships granted by University of Chicago; and from a scholarship for young researchers granted by the Dan David Foundation.

Abstract

The One-Child Policy is often perceived as a government-mandated quota system. Recent research acknowledges regional incentives allowing women to have more than one child. The policy was actually an individually tailored pricing system varying within each woman's life cycle. I document and exploit this variation to find that the policy: only affected women whose first child was a girl; did (not) reduce the number of girls (boys) born per woman; and caused a minor decrease in aggregate fertility. Data on ultrasound availability suggest that the second finding, but not the first, results at least partly from prenatal sex selection.

1 Introduction

The Chinese Economic Reform introduced market principles in 1979. Agriculture transitioned from collective- to household-based, the economy opened to foreign investment, and permits for opening private businesses were granted. In 1979, the Chinese government also launched the One-Child Policy. A common perception is that the policy was a government-mandated quota system. Some assert that the policy enforced uniform one-child families (Fong, 2004). Others add that ethnic minorities were exempt from the policy and that rural families with a female first child were allowed to have a second child (Hesketh, Lu, and Xing, 2005; Gracie, 2015). Recent research acknowledges regional incentives allowing women to have more than one child (McElroy and Yang, 2000; Ebenstein, 2010b).[1] None of these characterizations is accurate. Aligned with the other reforms implemented in 1979, the One-Child Policy was based on market principles and individual-level incentives. It was a pricing system allowing every woman and her partner to have more than one child—if they paid a price.

I construct a novel dataset that allows me to exploit the pricing system to study the effect of the policy on household- and aggregate-level fertility. For the first time in the Chinese setting, I construct a sample tracking the fertile segments of the life cycles of a nationally representative set of women during the years in which the policy was active. I integrate into this sample the historical bureaucratic documentation on the rules associated with the pricing system; community- and province-level information from statistical yearbooks and from surveys with community leaders describing the economy; and data on ultrasound technology from local gazettes.

[1] The characterization of provincial incentives in Ebenstein (2010b) is taken as given in subsequent studies in economics. See Zhang (2017) for a review.

The challenge in analyzing the impact of the One-Child Policy on fertility is isolating the effect of the policy from a decreasing trend in fertility. Sen (2015) and Whyte, Feng, and Cai (2015) argue that the decrease in fertility is a result of the demographic transition, rather than an outcome of the One-Child Policy. Based on evidence like that in Figure 1, they observe that after 1979 realized fertility continued to decrease smoothly in China.[2] Others contend that the draconian nature of the policy in fact drove most of the decrease in fertility, and claim that the policy prevented up to 400 million births.[3]

The analysis so far is inconclusive because it either relies on the inspection of aggregate trends in fertility or on incomplete characterizations of regional incentives associated with the policy. An indisputable fact is: There was no imposition of uniform one-child families. Had that imposition been implemented, the realized average number of children for women ages 15 to 45—denoted by the circle pattern in Figure 1—would have converged to 1. This and other measures of fertility, as the total fertility rate, have not converged to 1 even as recently as 2017 (The World Bank, 2017).

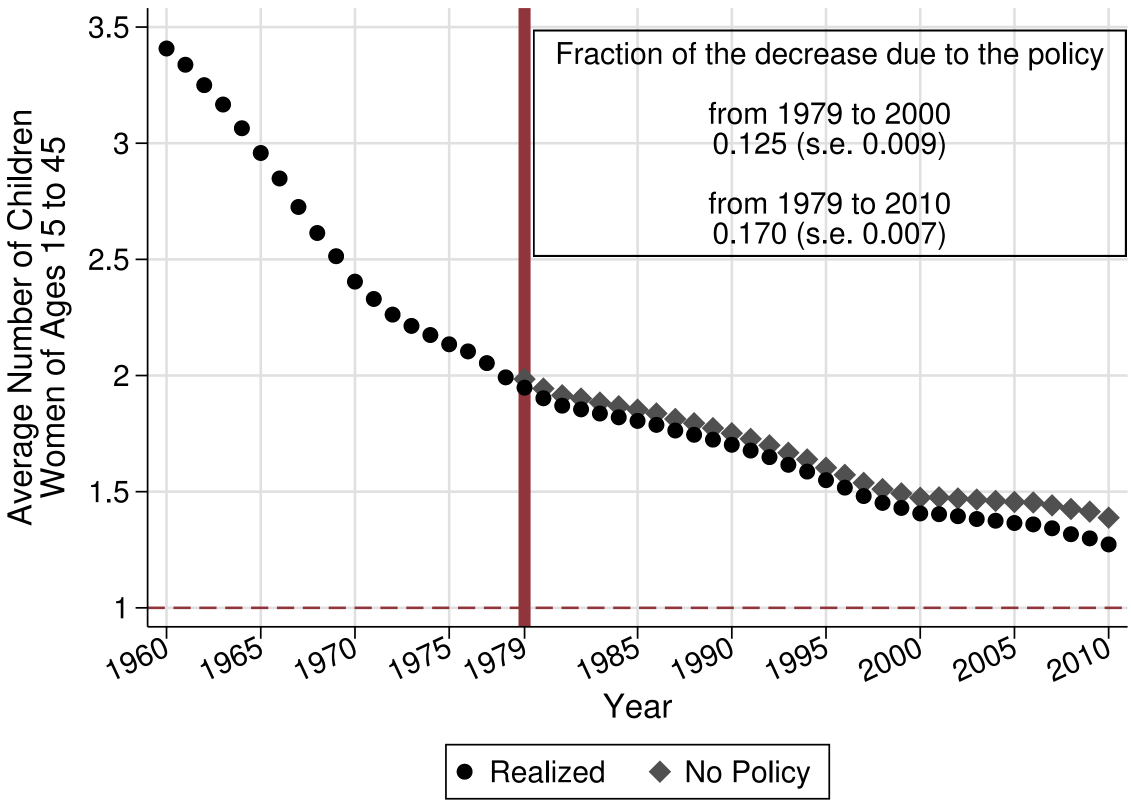
The One-Child Policy set rules that dictated the individual price associated with having more than one child. Women faced a two-part tariff. Every woman was allowed to have one child without a penalty. Having additional children resulted in a tax proportional to household labor income. The rules of the policy generated *de facto* variation in the tax across each woman's life cycle and across demographic groups and provinces.

Local officials announced the tax every year, generating province × year variation. Exemptions setting the tax to 0 were granted according to individual characteristics related to ethnicity, sex of the first child, age or health characteristics of the first child, scarcity of

[2]They also note that, during the years when the One-Child Policy was active, a decreasing trend was common across east-Asian countries.

[3]See Feng and Cai (2010) for a discussion of this and related estimates.

Figure 1: Aggregate Fertility in China



Note: This figure displays the average number of children for women with birth years between 1926 and 1972 ages 15 to 45 during the calendar year indicated in the abscissa (Realized); and the analogous number in a counterfactual scenario with no policy (No Policy), for which this paper explains the identification and estimation procedure. **Sources:** Author's calculations using the data described in Section 2.

males in the extended family, and job difficulties (e.g., risk associated with the jobs in which women and their spouses worked).[4] In total, there were 17 exemptions. Across provinces, not all 17 exemptions applied. Within provinces, different exemptions applied in different years while the policy was active. This generated within-woman tax variation far beyond the province $\times$ year variation. For instance, as a result of the policy, women (and their partners) had to pay, on average, 1 unit of average household labor income for each child that they had in addition to their first in 1980, independent of the sex of their first child. By 2000, this had increased to more than 2 units for women whose first child was a boy, while it had increased much less, to 1.5 units, for women whose first child was a girl.

I exploit this variation in a difference-in-differences framework to estimate the policy effect on the number of children at the household level. My strategy condenses individual-level event-studies. Past realizations of the policy have a negative and robust effect on current fertility, while future realizations have no effect. Women did not anticipate the policy when deciding current fertility levels, which provides a test of the "parallel trends" assumption that identifies policy effects in difference-in-differences frameworks.[5] Neither past nor future policy realizations have an effect on individual-level employment or on household-level labor income. This justifies discarding a spurious effect on fertility unrelated to the policy but related to fluctuations in labor income. Unique data on abortions and miscarriages allow me to further assess the validity of my empirical design.

I estimate that an increase of one standard deviation in my preferred measure of policy

[4]The taxes and the exemptions were determined by local family-planning officials in each province. The official's performance was evaluated by the central government. The evaluation was based on the evolution of concrete aggregate fertility statistics. Scharping (2003) documents the bureaucratic implementation of the policy in detail.

[5]See Suárez Serrato and Zidar (2016) and Fuest, Peichl, and Siegloch (2018) for recent studies using future policy realizations—as opposed to past—to verify the parallel trends assumption.

intensity—an increase which is equivalent to one half of annual average household labor income—caused a reduction of 0.101 (s.e. 0.014) children. I argue that this response is inelastic. While the pricing system was relatively ineffective at curbing fertility, a quota system would have been very welfare decreasing. The policy effect was solely driven by the subsample of women whose first child was a girl. This finding is not driven by sex selection at birth. Sex balance at the first parity has been documented in China (Ebenstein, 2010b; Chen, Li, and Meng, 2013), and is verified in my sample. Based on this, I conclude that the policy was not binding for women whose first child was a boy.

Although not relevant for the first parity, sex selection at birth does interact with the policy. I find that the effect of the policy was actually driven by a reduction in the average number of girls born per woman, while it had no effect on the average number of boys. Documentation of community-level availability of ultrasound technology suggests that the mechanism driving this effect is prenatal sex selection. I qualify this finding below because there could be non-random allocation of ultrasound technology across communities. Regardless, family-planning policies interact with parental preferences for the sex combination of children and could generate sex imbalance in the population. This finding provides a lesson for countries like India, with unfettered population growth and with documented preferences for having sons over daughters (e.g., Clark, 2000; Gaudin, 2011).

Using the household-level policy effects, I construct a counterfactual aggregate trend in fertility in the absence of the policy. In addition to requiring identification of the marginal effect of the policy on fertility, this construction also requires identification of the levels—i.e., the number of children that women did not have due to the policy. Identification in levels relies on strong assumptions that I explain and thoroughly test below. The diamond pattern in Figure 1 displays the result of this exercise. I estimate that 17% (s.e. 1%) of the decrease in the average number of children born to women ages 15 to 45 between 1979 and 2010 was caused by the policy.[6]

I contribute to the literature by providing a precise, individual-level documentation of the One-Child Policy in a longitudinal context. Exploiting this documentation, I provide novel estimates of the household- and aggregate-level effects of the policy on fertility. The data construction presented in this paper can be used to study other consequences of a policy that operated on an unprecedented scale, affecting approximately 500 million women between 1979 and 2010.[7]

My results are relevant for the following reasons. (i) Numerous studies in economics take as given the policy's effect on family size to study phenomena such as the quantity-quality trade-off or the effect of expected fertility on education (for a review, see Zhang, 2017). These studies are based on incomplete characterizations of the policy. (ii) The belief that the policy had a sizable effect on aggregate fertility is the basis for drawing strong conclusions, both in academic and popular circles. For example, The Economist claims that the One-Child Policy was the fourth main cause of the accumulated reduction in carbon emissions up to 2013.[8] (iii) Recent arguments assert that population control policies were a fundamental component in the global decline in fertility rates during the last 50 years (De Silva and Tenreyro, 2017), while other arguments suggest that the opportunity cost of raising children and the increasing demand for human capital are the main drivers of demographic transitions (Galor, 2012). It is important to clarify whether an iconic family-planning policy, implemented in the most populous country in the world, had an effect on aggregate fertility.

The rest of the paper unfolds as follows. Section 2 describes the institutional setting and the data construction. Section 3 explains the empirical strategy that I employ to identify

[6] I also display the fraction of the decrease between 1979 and 2000 because, as I explain below, my measures of policy intensity are most reliable between 1979 and 2000.
[7] See Section 2 for details on the calculation of the number of affected women.
[8] In "Curbing Climate Change: The Deepest Cuts" (The Economist, 2014).

household-level policy effects. Section 4 discusses these effects, justifies the identification assumptions, and explores robustness and heterogeneity. Section 5 explains how I use the household-level policy effects to estimate an aggregate policy effect. Section 6 explores the interaction of the policy with community-level ultrasound technology availability. Section 7 provides a discussion that contrasts my results with those obtained from using other policy characterizations. It also discusses other household-level outcomes that could have been affected by the policy. Section 8 concludes.

2 The One-Child Policy: Context and Data

Economic and Institutional Setting. The economic and institutional setting was inflexible during the era when the One-Child Policy was introduced. Although the government implemented market principles through a series of reforms, it still exerted great control over households' economic activities.

During the period of my analysis, the *hukou* (residency) dictated the location in which a person could live and assigned her either an agricultural or a non-agricultural status. Governmental rules, taxes, and subsidies were tied to each individual's *hukou*. The *hukou* system is so fundamental to the organization of the Chinese economy that the book detailing the rules associated with it is called *zhongguo diyi zhengjian* or "China's Number 1 Document" (Tian, 2003).

Until 2000, individuals with non-agricultural *hukou* belonged to a *danwei* (working unit). Membership to a *danwei* granted permanent employment and dictated total labor income—including access to food, health care, pensions and benefits, children's education, and housing according to statutes of the central government (Whyte and Parish, 1985; Tang and Parish, 2000). Most individuals remained in their *danwei* for life (Lu and Perry, 1997). After the liberalization of the economy in 1979, the *danwei* system continued to regulate individuals' economic lives (Naughton, 2007).

Individuals with an agricultural *hukou* mostly worked as farmers in collective systems until 1984. The collective to which they belonged dictated labor income through a point system. Individuals earned points per day worked and received basic in-kind payments. At the end of the harvest year, the government procured the grain, the surplus was sold at fixed prices, and the money was divided according to days worked (net of household-level fines and taxes). Between 1979 and 1984, the agricultural system was reformed. The land was divided as part of a contract between the collective and households. The contract also dictated grain prices. Because this system kept centralized control of the grain prices, it controlled household labor income (Cai, 2003). Agricultural activities gradually lost importance and individuals became part of town and village enterprises or other centralized activities, which regulated the economic life of individuals in a similar fashion as the *danwei* (Naughton, 2007).

Figure 2 illustrates how the inflexible setting affected individual economic decisions, using the data described in Section 2. In Panel 2a, I display the empirical probability of supplying full-time labor for females of childbearing ages in any given year, the year in which they gave birth, and one and two years before and after giving birth. There is virtually no variation in this probability, suggesting that the inflexible setting did not allow women to opt in or out of the labor force when considering their fertility decisions. This institutional feature is practical. Simultaneity of fertility and labor supply decisions is a standard concern when analyzing fertility empirically. In this setting, government mandates fixed female labor supply.[9]

As Panel 2b illustrates, in 80% of the communities across China, 10% or fewer household

[9]Similarly, the government dictated and provided established education levels (Fan, Heberer, and Taubmann, 2015; Postiglione, 2015).

heads ever migrated. This migration was mostly within communities, temporary, and regulated by government permits (Chan, 2001).[10] In my empirical strategy, evidence such as that in Figure 2 justifies the use of variables describing the economic environment as plausibly exogenous from women's perspectives.

Policy Evolution and Bureaucratic Implementation. After establishing the People's Republic of China in 1949, Mao Zedong encouraged population growth. Birth control, which would reduce the size of the workforce, was condemned and imports of contraceptives were banned. Abortions were illegal. After unfettered population growth, the government launched family-planning campaigns in the 1970s, encouraging the use of contraception, the delay of marriage, and the formation of smaller families (Powell, 2012). Towards the end of this decade, references to abortion were cleared from the constitution (Scharping, 2003). Before 1979, specific fertility limits were suggested but not enforced.[11]

In 1979, the government announced the One-Child Policy at the federal level. The mandate placed local authorities in charge of timing and implementation of the policy. Local authorities were incentivized to implement the policy because their effectiveness in curbing fertility factored into their job evaluations. Failing to effectively implement the policy could even have resulted in disaffiliation from the communist party (Scharping, 2003). An example evaluation form based on local aggregate fertility and citizens' knowledge of the policy is in Table A.8 in the Appendix.

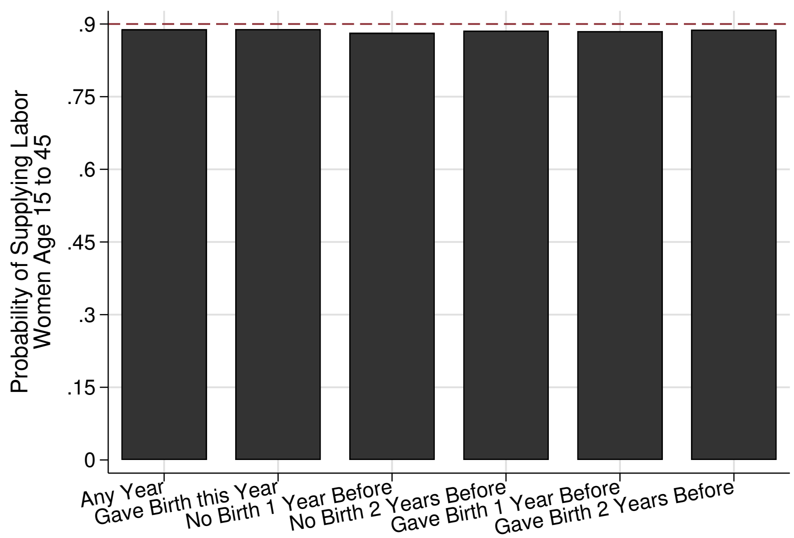
Local authorities also faced opposition to the policy from citizens (Hardee-Cleaveland and Banister, 1988). Arguably, the trade-off between conforming to national policy and mitigating social discontent at the local level, together with the contemporaneous economic reforms

[10]Massive waves of rural-to-urban migration began in the mid-1990s (Chan, 2001). These waves happened too late to affect the cohorts that I analyze.

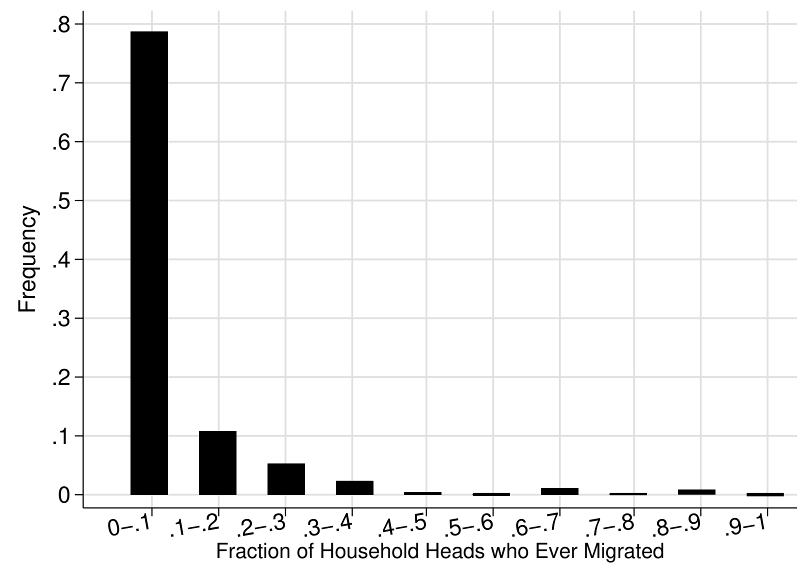
[11]In Section 4, I discuss the introduction of contraceptives and the phase-out of the illegality of abortions as potential confounders of the effect of the policy.

Figure 2: Economic Decisions in an Inflexible Setting

(a) Women's Labor Supply, Individual Level



(b) Household Heads' Migration, Community Level



Note: Panel (a) displays the empirical probability of supplying labor for women with birth years between 1926 and 1972 ages 15 to 45 in any given year, the year of giving birth, and one and two years before and after giving birth. Panel (b) displays the frequency of the fraction of household heads who ever migrated at the community level. **Source:** Author's calculations, created with information from the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

that introduced market principles, led officials to design the policy as a set of incentives dictating the price of having more than one child, instead of as an uncompromising imposition of one-child families.[12]

Quantifying Policy Intensity. The family-planning policies preceding the One-Child Policy established a bureaucracy for population control. Upon getting pregnant, each woman had to register her new child independent of her parity beginning in the early 1960s. After the enactment of the policy, permits for having a first child were generally granted—without a tax. Permits for having a second or a third child had a price. Permits for having a fourth child and beyond were generally denied. Permits for each child additional to the first were required (there were no discounts or joint contracts for second and third children).[13] The price per permit was a government-mandated proportional tax on household labor income that had to be paid for a number of years. Both the woman's and her partner's labor income were taxed.

In general, a woman i of age a had to register her pregnancy to the local birth-planning authority before her fourth month of pregnancy. If she was pregnant with her second or third child and wanted to keep the child, she needed to sign a contract with the government. The contract stated the fraction of household labor income to be taxed, κ_{ia}, and the number of years in which the tax would apply, L_{ia}. When entering the contract, households had

[12]Greenhalgh and Winckler (2005) provide examples that support this explanation. For instance, in 1989, Li Peng addressed provincial governors stating that: "To achieve substantial compliance, policy must be supplemented with more detailed management by objectives .... Targets should be evaluative." Another argument against a strict limit was that relatively small families would be economically damaged by the inability to have more than one child or that minorities would become underpopulated (Xu, 1984).

[13]Given the size of China's population, registering every single birth and collecting fines suggests the need for an immense bureaucracy. In fact, birth-planning expenditures grew from virtually 0 to 40 billion USD (2016) from 1970 to 1995. Employment in Birth-Planning Commissions grew from less than $5,000$ employees in 1970 to almost $400,000$ in 1994 (Scharping, 2003).

perfect certainty of both the fraction deducted from household labor income and the number of years in which the deduction would occur. If a woman decided to sign a contract with the government, the contract was not revised if the values κ_{ia} and L_{ia} evolved according to policy-rule changes. The averages of κ_{ia} and L_{ia} in my sample are 0.136 and 6.516.

I summarize the values of κ_{ia} and L_{ia} using the present value of the amount taxed that woman i would have been able to calculate at age a based on her current household labor income, y_{ia}:

$$\tau_{ia} := \sum_{\ell=1}^{L_{ia}} \beta^{\ell-1} \cdot \kappa_{ia} \cdot y_{ia}, \tag{1}$$

where β is the discount factor implied by a 2% discount rate. I refer to τ_{ia} as "the fine." τ_{ia} is my measure of age-a policy intensity. This measure has the virtue of condensing the information in κ_{ia} and L_{ia} into a scalar. However, it involves household labor income—potentially simultaneous to fertility decisions. The institutional setting that characterized the era of the One-Child Policy, mandating labor supply and household labor income, lessens concerns related to this construction. I provide tests for this in Section 4.

The calculation of τ_{ia} is based on y_{ia} and not on future values of household labor income. This is consistent with women having stationary expectations with respect to household labor income. If women had a different expectation process, τ_{ia} would be measured with error. I explore measurement error in Section 4.

Local officials announced the values of κ_{ia} and L_{ia} that applied during each year, generating province $\times$ year variation in τ_{ia}. Exemptions granted on an individual- and year-specific basis set κ_{ia} to 0 for the L_{ia} years.[14] An exemption implies $\tau_{ia} = 0$. The criteria qualifying women for exemptions varied across provinces and time. These criteria can be grouped into five categories: ethnicity, sex of the first child, age or health characteristics of the first child, scarcity of males in the extended family, and the risk associated with the parents' jobs. The exact details are in Tables A.1 to A.5 in the Appendix. The pricing system was tied to each woman's *hukou*. The *hukou* was (and remains) basically impossible to change.

As a starting point, I take the calculation of τ_{ia} from Ebenstein (2010b), which is based on province- and year-specific average household labor income, as opposed to household- and year-specific.[15] For example, Ebenstein (2010b) calculates that τ_{ia} was 1.21 years of province-specific average household labor income in 1980 in the province of Guangdong.[16]

The novelty in my quantification of policy intensity is that, as opposed to Ebenstein (2010b), I have the precise information to determine whether each woman i at age a qualifies for an exemption. Thus, I use the construction of τ_{ia} in Ebenstein (2010b) as a starting point and

[14]Scharping (2003) collates the historical documents that allow for a quantification of both the taxes and the exemptions. These come from a variety of sources made available by the federal government and province-level Birth Planning Commissions in charge of implementing the policy. The calculations in Ebenstein (2010a) rely on the historical documents in Scharping (2003).

[15]In Section 4, I document robustness to using household labor income. In the Appendix, I document my construction of household labor income. This involves challenges that are likely to introduce measurement error, which I tackle when using measures of τ_{ia} based on household labor income.

[16]The calculation uses province- and year-specific data on household labor income, demographic information from the China Health and Nutrition Study (CHNS) (Carolina Population Center and the National Institute for Nutrition and Health, 2009), and a 2% discount rate. Ebenstein (2010b) considers the relative exposure of the population to the policy. The calculation also accounts for urban-rural differences in the fines, and other differences within provinces. The final fine that he reports (1.21 years of household labor income in this example) is the weighted average of the fines reported in Scharping (2003) for different population groups, where the weights correspond to the relative density of the households exposed to the different fines. The calculations are available in Ebenstein (2010a). They produce a pattern of geographic variation similar to that of Baochang et al. (2007), who use prefecture-level, restricted access information on the enforcement of the policy.

set τ_{ia} to 0 when an exemption applied. After doing so, the average of τ_{ia} in my sample is 0.8 years of household labor income. It increased from 0 in 1979 to 2.3 in 2000. One year of household labor income is, on average, $1,300$ (2016 USD). It increased from 880 (2016 USD) in 1979 to $3,000$ (2016 USD) in 2000. Documentation on the policy is not available after 2000. This data restriction turns out to be relatively unimportant because the bulk of the decrease in aggregate fertility observed after 1979 happened before 2001.

Figure 3 illustrates the variation that I exploit in my empirical design. On average, τ_{ia} was around 1 unit of household labor income in 1980. It increased over time to around 2 units in 2000. This increase was much less for women whose first child was a girl; τ_{ia} decreased to almost 0 in 1990 for ethnic minorities and then increased again to 1 unit of household labor income in 2000. The variation that I document contrasts with simplified descriptions of the One-Child Policy as a mandate that uniformly enforced one-child families with an exception allowing rural families with a female first child to have two children (e.g., Gracie, 2015). It also contrasts with characterizations that use ethnic minorities as a "control group" that was uniformly exempted from the policy (e.g., Huang, Lei, and Zhao, 2013; Huang, 2016). Exemptions related to job difficulties and male scarcity also generated rich variation.[17]

The pricing system was the main nation-wide policy instrument (Scharping, 2003). In Section 4, I discuss other potential enforcement mechanisms. I estimate that 504.3 million women were potentially affected by the policy using the 1982, 1990, and 2000 census waves (Minnesota Population Center, 2017). This calculation is the sum of women who were between 15 and 45 years old for at least one year between 1979 and 2010.

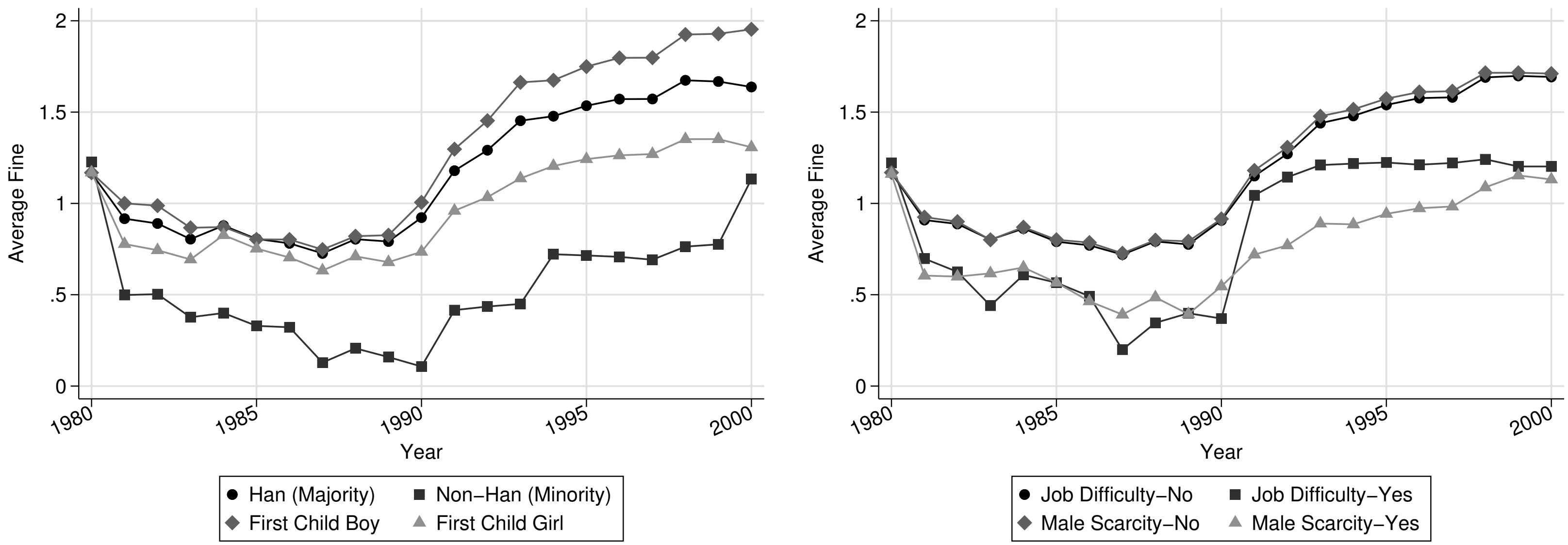
Constructing the Woman-Level Longitudinal Micro-Data. I link the fines with na-

[17]Figure A.2 in the Appendix further describes the identifying variation in the fine by displaying the distribution of τ_{ia} net of individual and age $\times$ year fixed effects. The standard deviation of this variable across women is 0.57. The average within-woman standard deviation is 0.48.

Figure 3: Fine to Have a Child Additional to the First in Units of Average Household Labor Income

(a) By Ethnicity and Sex of the First Child

(b) By Job Difficulty and Male Scarcity



Note: Panel (a) displays a time series of the average present value amount (fine) in exchange for a permit to have a child additional to the first in units of average household labor income by ethnicity and sex of the first child. Panel (b) is analogous to Panel (a) by job difficulty and male scarcity in the extended family. Job difficulty includes the following categories: one spouse is disabled and cannot work; a peasant couple lives in sparsely settled mountains, reclamation, or border areas; one spouse has been constantly working in underground mining for more than 5 years; and the couple has real economic difficulties or claims other (related) peculiar reasons. Male scarcity in the extended family includes the following categories: one spouse or both spouses are single children; only one child or one son has been born to a family for two generations; and the husband settles in the family of his wife which has daughters and no sons. Tables A.1 to A.6 of the Appendix provide more details on these categories. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

tionally representative data on women's fertility histories, which I construct from a retrospective survey that was part of the China Health and Retirement Study (CHARLS) (National School of Development, 2009-2017).[18] CHARLS is nationally representative and is the sister study of the US Health and Retirement Study (Zhao et al., 2009). The study's subjects lived in 443 communities within 27 provinces. The provinces excluded from the data collection were Chongqing, Hainan, Ningxia, and Tibet. This exclusion does not prevent the sample from being nationally representative.

CHARLS began in 2011 with an initial sample of 8,919 female household heads born between 1925 and 1972. I analyze fertility between 1970 and 2000 using this sample. My main analysis stops at 2000 because precise details on policy implementation are only available from 1979 to 2000. In Section 5, I discuss the assumptions required to include the years 2001 to 2010 in the analysis. For each woman in the sample, I characterize each year of her fertility history (i.e., each year between age 15 and 45) with the total number of children (dead or alive), the total number of miscarriages and induced abortions, and the fines that applied in all of the relevant ages.[19]

To complement these data, I construct and match a dataset characterizing the economic environment in which women made fertility decisions. This construction combines two sources: province- and year-specific data from Chinese Statistical Yearbooks available in China Data

[18] I show evidence indicating that the retrospective number of children is accurately reported in Figure A.1 in the Appendix.

[19] Table A.9 in the Appendix provides descriptive statistics for the main sample of analysis. Statistics on fertility, policy-related variables, and employment are provided when relevant throughout the text. The average birth year in my sample is 1954. The sample becomes relatively old as my year of analysis progresses. This is not a concern in my empirical analysis for two reasons. First, I exploit within-woman variation in the policy in my empirical analysis, and all of my analysis is conditional age $\times$ year fixed effects. I document that my estimates of policy effects are robust to focusing on subsamples of various cohorts. Second, most of the aggregate decrease in fertility occurred before 2001. Before then, in 2000, there are 3,850 observations of women between ages 15 and 45 in my sample.

Center (2017) and community- and year-specific data based on retrospective surveys of community leaders available in CHARLS. Table A.10 in the Appendix provides a full list of the measures. In my empirical analysis, I aggregate each economic category into dedicated factors (e.g., employment, education, agriculture, trade).

I also construct and match community-level data on ultrasound technology availability. I obtained these data from two sources: the community-level "New China Local Gazetteers" and a report resulting from the conference "Reviews of Ultrasound Medicine in China." The gazetteers began to be published after 1949, and cover local developments in agriculture, public health, public security, etc. I use this source for most communities and complement it with information from local health-specific gazetteers. Information for four major cities (Beijing, Ningbo, Shenzhen, and Chongqing) is available in the conference reviews. The data collection consists of inspecting each gazetteer after the invention of ultrasound technology to determine when availability began. Data on 68 communities within 23 provinces are available. In Section 6, I explain my use of these data and explain the consequences of the limited information on ultrasound availability. More details on this part of the data construction are in the Appendix.

3 Empirical Strategy

Research Design and Identification. My outcome of interest is the number of children of woman i at age a in levels, n_{ia}. In this section and the next, I focus on the effect of the policy on this outcome at the household-level. This allows me to to construct a counterfactual aggregate fertility trend in the absence of the policy in Section 5. I start by analyzing the cohorts born in 1926 to 1972 between the years 1970 and 2000.

I consider the fertile years during the life cycle of each woman. I do not observe menarche or menopause so I include women in the age span 15 to 45. This generates an unbalanced panel of $8,023$ women.[20] My analysis is based on a woman-level event study designed as follows:

$$n_{ia} = \lambda_i + \phi(a,t) + \sum_{j=-J}^{J} \gamma_j \tau_{ia+j} + \varepsilon_{ia}, \tag{2}$$

where λ_i is an individual fixed effect, $\phi(a,t)$ is a full interaction of age and year indicators, and ε_{ia} is an error term. The independent variables of interest are τ_{ia+j}, in which j indicates the years of distance from the current age a and $j \in [-J, \ldots, J]$ for some $J \in \mathbb{R}^{++}$.

The events of interest are the realizations of τ_{ia+j}. The coefficients associated with the lags of the fine are $[\gamma_{-J}, \ldots, \gamma_0]$, while the coefficients associated with leads of the fine are $[\gamma_1, \ldots, \gamma_J]$. A sufficient condition for identifying $[\gamma_{-J}, \ldots, \gamma_J]$ is strict exogeneity.[21] The design of Equation (2) allows me to justify the parallel trends assumption implied by strict exogeneity.

The individual-level criteria qualifying women for policy exemptions could cast doubt on the validity of this assumption. In Section 4, I discuss how these criteria could induce violations to this assumption. Broadly, the justification of this assumption is the following. If future fines have no effect on current fertility, $[\gamma_1, \ldots, \gamma_J] = \mathbf{0}$, it is likely that women did not anticipate future realizations of the policy and, thus, for the assumption of parallel trends to hold. Trends would be parallel because policy information that is not public when women make fertility decisions would have no effect on the current number of children in the realized scenario or in a counterfactual scenario with no policy.

[20] This is the final sample for my analysis. In Table A.11 in the Appendix, I argue that the difference in observations between the initial sample reported in Section 2, $8,919$ women, and the final analysis sample does not bias my estimates of the policy effect.

[21] That is, $\mathbb{E}\left[\varepsilon_{ia} | \tau_{ia-J}, \ldots, \tau_{ia+J}, \lambda_i, \phi(a,t)\right] = 0$.

Even if $[\gamma_1, \ldots, \gamma_J] = \mathbf{0}$, a systematic relationship between the fines and household labor income would cast doubt on my empirical strategy. Especially because the fines, my measures of policy intensity, are a function of province-level labor income and because household labor income *per se* is a usual determinant of fertility decisions in economic models (for a review, see Hotz, Klerman, and Willis, 1997). The fact that household labor income is largely fixed by governmental policies, as I discuss in Section 2, alleviates this concern.

To further test this, I provide three categories of specification checks. (i) I estimate the coefficients characterizing event studies analogous to that in Equation (2) using individual-level female employment and household-level labor income as outcomes of interest. There is no relationship between past and future fines and these two variables.[22] (ii) I analyze the robustness of the functional form specification in Equation (2). For instance, I provide sensitivity analysis to the functional form assumptions to account for individual, age, and year effects. (iii) I control for a host of lagged, current, and lead values of factors describing the economic environment. In Section 4, I document that my estimates remain robust to these checks.

Unique data on abortions and miscarriages allow me to perform additional checks. Past fines have a positive effect on abortions. This finding is economically sound because abortions were one of the contraception methods available to women after the enactment of the policy. Future fines have no effect on abortions. It would be concerning if they did because it would reveal a systematic relationship between the fines and abortions. In that case, the fines could contain information related to the the phase-out of the illegality of abortions and, more generally, the availability of contraception methods that began in the late 1970s.

Observing miscarriages is a useful counterpart to observing abortions. Conditional on $\phi(a, t)$,

[22]The very low variation in individual-level male employment does not allow me to perform an analogous, credible test for male labor income.

the past and future fines should have no effect on miscarriages because miscarriages are not explicit economic decisions.[23] Fluctuations in miscarriages should be part of the age × time trend (e.g., due to improved public health services). I verify that this is the case in Section 4.

Inference. My empirical strategy relies on province × birth year variation. I use the corresponding two-way clustering when calculating standard errors as a baseline strategy. I document robustness to other clustering options: individual × age (homoscedastic), individual (block-homoscedastic), and province-level.

4 Main Results

Main Event Studies. I start by providing estimates of the coefficients characterizing Equation (2) binning the fines in two-year averages (Panel 4a). The coefficient at 1 is associated with the average fine of periods $a+1$ and $a+2$, the coefficient at 0 is associated with the average fine of periods a and $a-1$, the coefficient at -1 is associated with the average fine of periods $a-2$ and $a-3$, and so on. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. For instance, an increase of one standard deviation at 0 decreases the number of children at age a, on average, by 0.01 children. I present a thorough discussion on the magnitude of this and the effects in other time periods after justifying specification decisions.

There is an unlimited amount of lags of the fine that I could consider because the fines before 1979 were uniformly 0 for every women. The number of the leads of the fine is limited by the fact that information on the policy is only available up to 2000. I consider at most 5 two-year averages in Panel 4a.

[23] In fact, Table A.9 documents that the average likelihood of suffering a miscarriage remained low and stable across the years of my analysis.

Leads of the fine have no effect on current fertility. Period-specific and joint tests confirm this. In addition, a specification dropping leads of the fine yields very similar estimates for the coefficients associated with lags of the fine. At 0, there is a clear break. Fines that are public at age a (i.e., lags of the fine) have negative and precise effects on the number of children at age a. Fines that are not public (i.e., leads of the fine) have no effect. This supports the parallel trends assumption.

Panel 4b provides estimates of coefficients characterizing an equation analogous to Equation (2) with individual-level female employment and household-level labor income as dependent variables. There is no relationship between past and future fines and these two variables. This lessens concerns of the fines being based on province-level labor income, thus inducing a spurious effect on fertility related to fluctuations in labor income but unrelated to the policy.

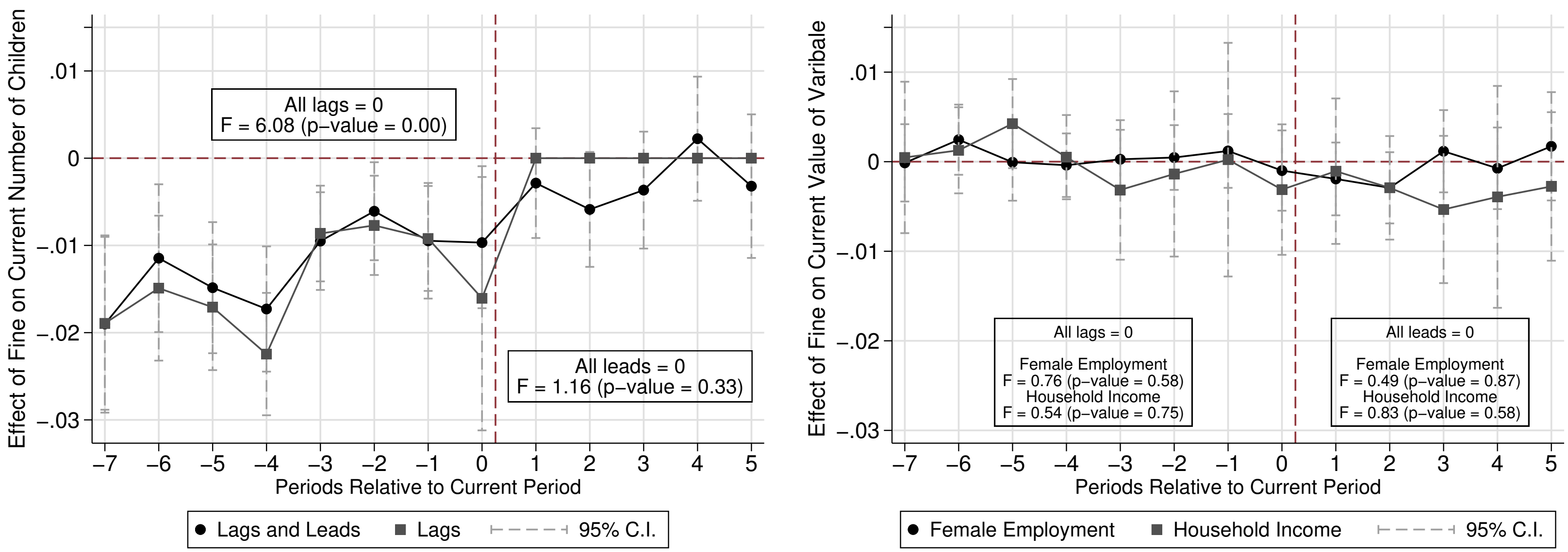
Event Studies with Refined Periods. Panel 5a refines the evidence in Panel 4a. It bins all leads of the fine into one average and all lags of the fine into three averages. This refinement makes the tests more precise. The break at 0 becomes clearer, the joint test on the lags of the fine indicates a stronger effect of the policy, and the specification dropping leads of the fine still yields very similar estimates for the coefficients associated with lags of the fine. The magnitude of the coefficient increases because I restandardize the fine measures. An increase of a standard deviation has a larger economic magnitude and causes a larger decrease in the number of children.

There could be measurement error in the fines and that is why averaging makes the estimates more precise, both when failing to reject the null effect of future fines and when rejecting the null effect of past fines. Below, I actually consider the effect of the average of all lags of the fine and the effect of the average of all leads of the fine. This allows me to condense

Figure 4: Effect of Past and Future Fines, Event Study

(a) Fertility

(b) Female Employment and Household Income



Note: Panel (a) displays estimates of the coefficients associated with lags of the fine, $[\gamma_{-J}, \ldots, \gamma_0]$, and with leads of the fine, $[\gamma_1, \ldots, \gamma_J]$, in Equation (2) binning 15 lags of the fine and 10 leads of the fine in two-year bins. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. The "lags" specification sets the coefficients on leads of the fine to 0. The dependent variable is the number of children at age a and the controls are individual and age $\times$ year fixed effects. Inference is province $\times$ birth year two-way clustered. Asymptotic confidence intervals are displayed. Panel (b) is analogous to Panel (a) with individual-level female employment and household-level labor income as dependent variables. The sample is an unbalanced panel of $8,023$ women from the cohorts born in 1926 to 1972 who were between ages 15 to 45 in the years 1970 to 20000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), China Data Center (2017), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

the information of my estimates into one statistic to test for parallel trends and one statistic to test the effect of the policy. Before discussing such estimates, I discuss one important specification issue.

It is economically sound that lags of the fine have an effect on current fertility. The current number of children is a stock accumulated over the years and fines in the past disincentive accumulation of children. A question is: How many lags of the fine should be considered? The sample contains women between ages 15 and 45. For relatively old women, several lags of the fine could be relevant. For women between ages 15 and 25, it would be concerning if fines very far in the past had an effect on the current number of children. This could mean that the fines contain information, for example, about regional fertility patterns, instead of information about policy intensity. To discard this potential issue, Panel 5b shows that fines in the far past, composed of the average fines 14 to 20 years before the current period, have no effect on the current number of children for women ages 15 to 25.

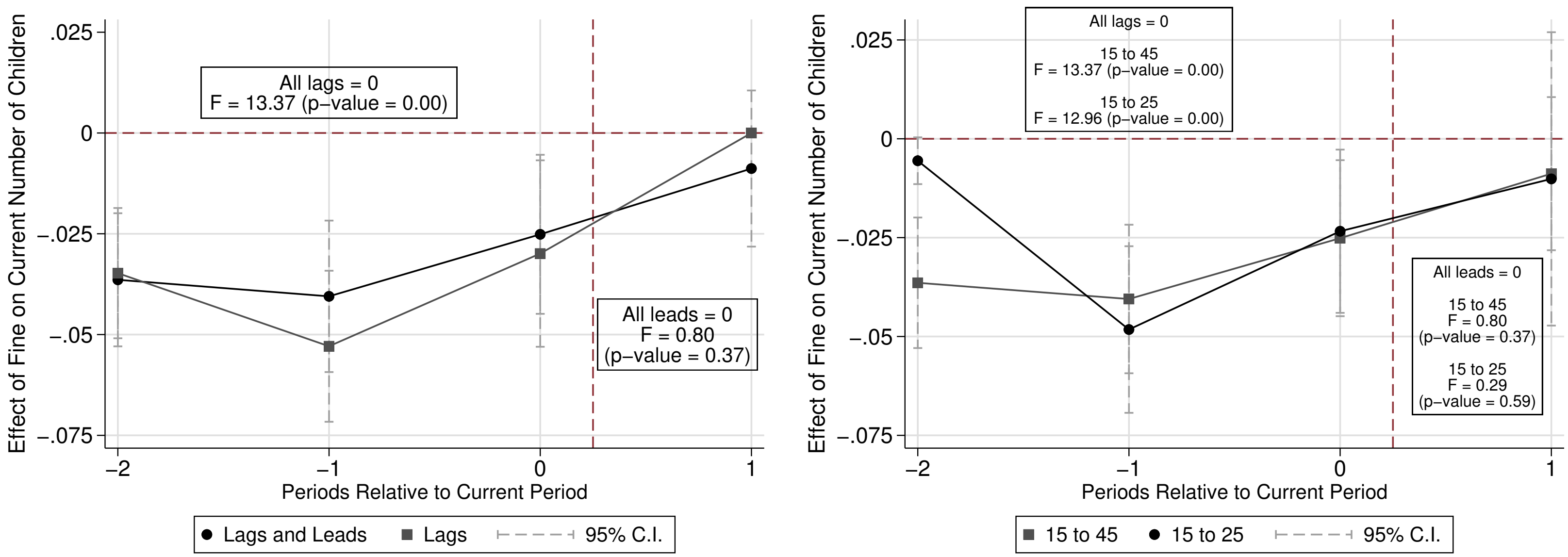
Condensed Event Studies on Children, Abortions, and Miscarriages. I consider the effect of the policy when condensing lags and leads of the fine. When doing so, I average 20 fines in the past and 10 fines in the future. The average of past fines is my baseline measure of policy intensity. The mean of this measure amounts to 0.412 (s.d. 0.476) units of a year of average household labor income. Recall that this is the average of the present value amount that a woman (and her spouse) would have had to pay in exchange for a permit to have a child additional to their first. When displaying results, I standardize the binned averages as detailed above.

Panel (a) of Table 1 displays the results with the number of children as the dependent variable. The first column considers a specification with both lags and leads of the fine. The second column drops the leads of the fine. The leads of the fine have no effect on fertility

Figure 5: Effect of Past and Future Fines, Refining Period Bins

(a) Fertility

(b) Fertility by Age Groups



Note: Panel (a) displays estimates of the coefficients associated with lags of the fine, $[\gamma_{-J}, \dots, \gamma_0]$, and with leads of the fine, $[\gamma_1, \dots, \gamma_J]$, in Equation (2) binning 20 lags of the fine in three bins and 10 leads of the fine in one bin. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. The "lags" specification sets the coefficients on leads of the fine to 0. The dependent variable is the number of children at age a and the controls are individual and age $\times$ year fixed effects. Inference is province $\times$ birth year two-way clustered. Asymptotic confidence intervals are displayed. Panel (b) is analogous to Panel (a) for two samples: women 15 to 45 years old (full sample) and women 15 to 25 years old. The full sample is an unbalanced panel of $8,023$ women from the cohorts born in 1926 to 1972 who were between ages 15 to 45 in the years 1970 to 20000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

and dropping them has little effect on the coefficient estimate associated with lags of the fine. An increase of one standard deviation in my baseline measure of policy intensity (lags) decreases the current number of children by 0.100 (s.e. 0.014) children. That is, increasing the average of the fine in the last 20 years by one half of a unit of average yearly household labor income decreases the number of children by one tenth of a unit. This response is inelastic showing that the pricing system was relatively ineffective at curbing fertility. In a log-log specification analogous to that in the first block (second column) of Panel (a) of Table 1, the policy elasticity is -0.080 (s.e. $= 0.011$, $N = 8,023$, $R^2 = 0.866$). The inelastic response indicates that a quota system would have been very welfare decreasing.

I consider these estimates by sex of the first child. Policy effects by sex are a starting point to explore the interaction of the policy with a preference for sons that characterizes Chinese families.[24] The effect of the policy is driven solely by the women whose first child was a girl. This is not driven by prenatal sex selection at first parity. In Figure A.4 in the Appendix, I show that the sex ratio at birth of the first child is balanced in my sample.[25]

The policy is not binding for women whose first child was a boy. The restriction was imposed on women whose first child was a girl. Women with a son preferred not to have another child. This is economically sound: If deciding to have another child, women whose first child was a boy faced the risk of having a second son. A second son would require the parents to provide him with a second household at the time of marriage, among other costs. Second sons would then be a *zhong fudan* or a "heavy burden" (Greenhalgh, Chuzhu, and Nan, 1994). Women whose first child was a girl faced a trade-off: Have a second child to try for a boy and satisfy social standards or avoid the fine. That is where the policy restriction kicked

[24]This preference has strong precedents in Chinese society and is part of Confucian tradition (Arnold and Zhaoxiang, 1986). Qian (2008) and Almond, Li, and Zhang (2013) study the interaction of land-reform and agricultural policies with this preference.

[25]Ebenstein (2010b) and Chen, Li, and Meng (2013) document a similar finding using various census waves and other sources.

Table 1: Baseline Effect on Number of Children, Abortions, and Miscarriages

	All		First Child Girl		First Child Boy	
Panel (a). Children						
Leads	-0.004		-0.011		0.022	
	(0.011)		(0.013)		(0.014)	
Lags	-0.086	-0.101	-0.084	-0.097	0.022	0.013
	(0.015)	(0.014)	(0.021)	(0.020)	(0.019)	(0.019)
N	8,023	8,023	3,746	3,746	4,037	4,037
R^2	0.879	0.879	0.879	0.878	0.887	0.887
Panel (b). Abortions						
Leads	-0.002		0.003		-0.007	
	(0.003)		(0.004)		(0.004)	
Lags	0.009	0.010	0.023	0.024	-0.003	-0.004
	(0.005)	(0.005)	(0.008)	(0.007)	(0.007)	(0.007)
N	6,892	6,892	3,267	3,267	3,531	3,531
R^2	0.687	0.687	0.669	0.669	0.701	0.701
Panel (b). Miscarriages						
Leads	0.002		0.003		0.003	
	(0.001)		(0.002)		(0.002)	
Lags	-0.002	-0.001	0.002	0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
N	6,896	6,896	3,267	3,267	3,532	3,532
R^2	0.802	0.802	0.799	0.799	0.804	0.804

Note: Panel (a) displays estimates of the coefficients associated with lags of the fine, $[\gamma_{-J}, \ldots, \gamma_0]$, and with leads of the fine, $[\gamma_1, \ldots, \gamma_J]$, in Equation (2) binning 20 lags of the fine in one bin and 10 leads of the fine in one bin. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. The columns labeled "All" show results from a specification considering both lags and leads of the fine and from a specification setting the coefficient associated with leads of the fine to 0. The dependent variable is the number of children at age a and the controls are individual and age $\times$ year fixed effects. Inference is province $\times$ birth year two-way clustered. Asymptotic standard errors are in parentheses. The subsequent columns present the analogous results when restricting the sample to women whose first was a girl or a boy. Panels (b) and (c) are analogous to Panel (a) with abortions and miscarriages as the dependent variables. The sample is an unbalanced panel of $8,023$ women from the cohorts born in 1926 to 1972 who were between ages 15 to 45 in the years 1970 to 20000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

in and generated an effect.

The effects by sex of the first child allow me to provide a further specification check using data on abortions and miscarriages. Abortions were one of the contraception methods available during the era of the One-Child Policy. Thus, the policy should have a positive effect on abortions, although smaller in magnitude compared to the effect on fertility because other contraception methods were available. To be consistent with the policy effects on the number of children, this should be true for women whose first child was a girl, while the effect on women whose first child was a boy should be 0. As I discuss in Section 3, parallel trends in abortions are important to test due to preceding policies phasing out the illegality of abortions. As discussed above, the policy should have no effect on miscarriages independent of the sex of the first child. Panels (b) and (c) of Table 1 verify these results.

Specification Checks. Table 2 displays a battery of sensitivity checks. I replicate several blocks analogous to the first block of estimates in Panel (a) of Table 1. The estimates remain robust to the checks in the following sense. (i) The estimates of the coefficient associated with the lags of the fine remain similar across specifications. (ii) This is true when including and when excluding leads of the fine, both within and across specifications. (iii) The estimates of the coefficients associated with leads of the fine, which allow me to test the parallel trends assumption, remain close to 0. When they are not close to 0, they are economically insignificant.

Panel (a) includes several specifications for capturing age and year effects. The most inclusive specification is that in my baseline strategy, where I account for age $\times$ year fixed effects.[26] The other specifications produce similar estimates.

[26]This full interaction is meant to capture all possible age $\times$ year information. It is not meant to identify age and year effects, which are not non-parametrically identified in this context (Heckman and Robb, 1985).

Panel (b) considers province and community-level characteristics. Recall that there are 443 communities within 27 provinces in my sample. The estimates remain robust to including age $\times$ year $\times$ province fixed effects. This allows me to discard the concern that any province-specific fertility patterns affecting women of specific ages during specific years drive fine variation and induce a spurious policy effect.[27] My estimates are also robust to controlling for the year-specific, community-level average measure of policy intensity (binned lags of the fine) and to limiting the sample to communities with preceding fertility policies.[28] This discards the concern that the policy effect is driven by peer effects or by women being located in a community with a tradition of fertility control.

Panel (c) considers estimates of the policy effects for different birth cohorts. There are two cases when $\tau_{ia} = 0$. First, $\tau_{ia} = 0$ before 1979. Second, beginning in 1979, τ_{ia} could have been 0 if women qualified for an exemption. I treat the two cases equivalently. This could be problematic. Old cohorts, facing $\tau_{ia} = 0$ between ages 15 and 45, are potentially selected because they survived historical events with large death tolls (e.g., Chinese Civil War, Great Famine). The estimates by cohorts show that, when dropping the older cohorts, the results remain virtually identical. This is sensible because identification relies on within-woman policy variation and not on relatively old "pure controls" never exposed to the policy.

Panel (d) shows that the estimates are also robust to including controls describing the economic environment in which decisions were made (i.e., describing, for instance, the province- and community-level state of the economy, education, trade, infrastructure) as detailed in

[27] The number of observations decreases in this case because, when accounting for age $\times$ year $\times$ province fixed effects, multiple observations without effective longitudinal variation appear in the sample and are dropped from the estimation.

[28] There were communities that implemented preceding fertility policies (e.g., suggested age of marriage, birth spacing, and number of children) and communities that did not. Information to construct the sample of communities that did comes from surveys with community-leaders and is reported in National School of Development (2009-2017).

Table A.10 in the Appendix. Robustness to these specifications is consistent with the economic environment being plausibly exogenous from each women's perspective.

Finally, Panel (e) shows that the estimates of the policy effect remain significant under various inference clustering procedures.

Endogeneity and Measurement Error. Some of the criteria for the exemptions to the policy, listed in Table A.1 and a main driver of the variation that I exploit, could appear to be endogenous (e.g., death of the first child, remarriage, or sex of the first child). It is important to note, however, that not all of the exemptions applied across provinces. Within each province, there was variation in when the exemptions applied. If households planned ahead in order to qualify for an exemption, they should have also been able to predict whether the exemptions would apply for the time when the female household head gave birth. Had they been able to predict future exemptions, then future fines would have an effect on their current number of children. The multiple tests of parallel trends displayed so far accumulate evidence against this kind of anticipatory behavior.

To further inquire on the endogeneity of the fines, I exploit unique data on household-level labor income which I describe in the Appendix. I measure the fine as in Section 2 but using observed household-level labor income. Household-level labor income is likely endogenous and measured with error, especially given the institutional framework in China where in-kind transfers were included in labor income. I provide estimates analogous to those in Panel (a) of Table 1 using the province-level fine, used so far as the independent variable of interest, as an instrument for the fine measures based on observed household-level labor income. The estimate analogous to that in the first block (second column) of Panel (a) of Table 1 is -0.124 (s.e. $= 0.023$, $N = 8,019$, $R^2 = 0.876$). This close alignment provides evidence against endogeneity of the fines as a specification concern.

Table 2: Baseline Effect, Specification Checks

Panel (a).	Age and Year FEs		Age FEs & Year Polynomial		Age Polynomial & Year FEs		Age & Year Splines	
Leads	-0.002		0.004		-0.005		0.006	
	(0.011)		(0.011)		(0.012)		(0.011)	
Lags	-0.086	-0.101	-0.088	-0.102	-0.085	-0.100	-0.096	-0.108
	(0.015)	(0.014)	(0.015)	(0.014)	(0.015)	(0.014)	(0.013)	(0.013)
N	8,023	8,023	8,023	8,023	8,023	8,023	8,023	8,023
R^2	0.877	0.877	0.877	0.877	0.868	0.868	0.876	0.875
Panel (b).	**Age × Year FEs**		**Age × Year × Province FEs**		**Control for Community Fine**		**Comms. w/ Preceding Policies**	
Leads	-0.004		-0.030		-0.022		-0.007	
	(0.011)		(0.024)		(0.015)		(0.014)	
Lags	-0.086	-0.101	-0.080	-0.077	-0.077	-0.094	-0.083	-0.111
	(0.015)	(0.014)	(0.028)	(0.023)	(0.019)	(0.015)	(0.021)	(0.021)
N	8,023	8,023	3,674	3,674	8,023	8,023	3,904	3,904
R^2	0.879	0.879	0.884	0.884	0.879	0.879	0.880	0.880
Panel (c).				**Cohorts**				
	1935-1972		1945-1972		1955-1972		1965-1972	
Leads	-0.004		-0.002		0.018		0.051	
	(0.011)		(0.011)		(0.015)		(0.031)	
Lags	-0.086	-0.101	-0.083	-0.099	-0.073	-0.097	-0.045	-0.073
	(0.015)	(0.014)	(0.015)	(0.014)	(0.019)	(0.017)	(0.034)	(0.029)
N	7,622	7,622	6,465	6,465	3,850	3,850	870	870
R^2	0.874	0.874	0.847	0.847	0.801	0.800	0.769	0.768

Note: Panel (a) displays estimates of the coefficients associated with lags of the fine, $[\gamma_{-J}, \ldots, \gamma_0]$, and with leads of the fine, $[\gamma_1, \ldots, \gamma_J]$, in Equation (2) binning 20 lags of the fine in one bin and 10 leads of the fine in one bin. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. The columns labeled "Age and Year FEs" show results from a specification considering both lags and leads and from a specification setting the coefficient associated with leads of the fine to 0. The dependent variable is the number of children at age a and the controls are individual, age, and year fixed effects (unless otherwise noted). Inference is province × birth year two-way clustered. Asymptotic standard errors are in parentheses. The next columns present the analogous results with alternative specifications to account for age and year fixed effects. The rest of the panels are analogous with the specification differences as labeled. The sample is an unbalanced panel of $8,023$ women from the cohorts born in 1926 to 1972 who were between ages 15 to 45 in the years 1970 to 20000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), China Data Center (2017), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

Table 2: Baseline Effect, Specification Checks, Cont.

Panel (d).	Economy Controls (e.g., economy, agriculture, trade)							
	Current		+ 2 Yr Lags		+1 Yr Lag/Lead		+ 2 Yr Lag/Lead	
Leads	-0.023		-0.016		-0.023		-0.023	
	(0.009)		(0.009)		(0.009)		(0.010)	
Lags	-0.102	-0.117	-0.095	-0.112	-0.108	-0.125	-0.111	-0.129
	(0.013)	(0.013)	(0.013)	(0.012)	(0.014)	(0.014)	(0.016)	(0.015)
N	7,977	7,977	7,942	7,942	7,942	7,942	7,860	7,860
R^2	0.882	0.882	0.888	0.888	0.884	0.884	0.885	0.885
Panel (e).	Clustering							
	Individual × Age		Individual		Province × Cohort		Province	
Leads	-0.004		-0.004		-0.004		-0.004	
	(0.003)		(0.008)		(0.011)		(0.029)	
Lags	-0.086	-0.101	-0.086	-0.101	-0.086	-0.101	-0.086	-0.101
	(0.004)	(0.003)	(0.011)	(0.011)	(0.015)	(0.014)	(0.041)	(0.037)
Clusters	185,011	185,011	8,023	8,023	1,067	1,067	27	27

Note: Panel (a) displays estimates of the coefficients associated with lags of the fine, $[\gamma_{-J}, \dots, \gamma_0]$, and with leads of the fine, $[\gamma_1, \dots, \gamma_J]$, in Equation (2) binning 20 lags of the fine in one bin and 10 leads of the fine in one bin. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. The columns labeled "Age and Year FEs" show results from a specification considering both lags and leads and from a specification setting the coefficient associated with leads of the fine to 0. The dependent variable is the number of children at age a and the controls are individual, age, and year fixed effects (unless otherwise noted). Inference is province × birth year two-way clustered. Asymptotic standard errors are in parentheses. The next columns present the analogous results with alternative specifications to account for age and year fixed effects. The rest of the panels are analogous with the specification differences as labeled. The sample is an unbalanced panel of $8,023$ women from the cohorts born in 1926 to 1972 who were between ages 15 to 45 in the years 1970 to 20000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), China Data Center (2017), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

Other Endogeneity Concerns. Recent studies show that corruption contributed to the outcomes of important historical events in China such as the Great Famine (Sun et al., 2013; Meng, Qian, and Yared, 2015). If women with close ties to the Communist Party received special treatment, the exogeneity of the fines could be threatened. Figure A.3 in the Appendix displays fertility by father's political affiliation using an auxiliary, nationally representative sample (Institute of Social Science Survey, 2010). Once multiple cohorts were exposed to the policy, it is actually the case that women with close ties to the government had lower fertility. This suggests that, if anything, women with close affiliations to the government more strictly followed the policy.

Heterogeneity. I describe heterogeneity in Table 3. First, I consider policy effects by *hukou*. Although it remains negative and significant, the effect of the policy is lower for women of agricultural status. This is not driven by urban-rural location because it is common for women who reside in urban areas to have an agricultural *hukou*.[29]

The effect is strongest for women who reside in rural areas (52% of the sample). Although this could imply that other enforcement mechanisms are omitted from my specification (e.g., enforcement being stronger in urban areas due to individuals belonging to working units that could more closely monitor compliance), the next block of estimates suggests that this is not a concern.

Even women who worked their whole lives in state-owned enterprises, where enforcement would be stronger, were sensitive to the fine. In fact, enforcement to a one-child limit was not practiced through state-owned enterprises. The average number of children among women who worked in state-owned enterprises their whole lives is 2.256. This suggests that the

[29] In the sample that I analyze, 81% of women have an agricultural *hukou*. Of these women, 74% reside in urban areas and 26% in rural areas. Virtually all women with non-agricultural *hukou* reside in urban areas.

Table 3: Baseline Effect, Heterogeneity

Panel (a).	Agricultural		Non-Agricultural		Rural		Urban	
Leads	0.001		-0.002		-0.005		0.017	
	(0.012)		(0.015)		(0.012)		(0.015)	
Lags	-0.056	-0.077	-0.110	-0.106	-0.077	-0.101	0.014	0.018
	(0.016)	(0.015)	(0.023)	(0.023)	(0.017)	(0.016)	(0.024)	(0.025)
N	6,421	6,421	1,545	1,545	4,837	4,837	3,185	3,185
R^2	0.879	0.878	0.894	0.894	0.878	0.877	0.883	0.883
Panel (b).	Always Worked in SOE's		Not Always Worked in SOE's		Age Span 15-25		Age Span 26-45	
Leads	0.033		0.000		-0.038		-0.012	
	(0.026)		(0.011)		(0.021)		(0.010)	
Lags	-0.042	-0.048	-0.073	-0.092	-0.269	-0.286	-0.058	-0.071
	(0.045)	(0.045)	(0.015)	(0.015)	(0.058)	(0.057)	(0.013)	(0.012)
N	475	475	7,546	7,546	6,298	6,298	8,023	8,023
R^2	0.892	0.892	0.879	0.879	0.738	0.737	0.916	0.916
Panel (c).	Mother's Education and Household Socioeconomic Status: No School		Some School		Not Wealthy		Wealthy	
Leads	-0.004		-0.003		-0.002		0.022	
	(0.014)		(0.013)		(0.012)		(0.018)	
Lags	-0.057	-0.072	-0.101	-0.117	-0.063	-0.081	-0.078	-0.084
	(0.019)	(0.018)	(0.020)	(0.020)	(0.017)	(0.016)	(0.024)	(0.024)
N	4,578	4,578	3,427	3,427	6,600	6,600	1,347	1,347
R^2	0.882	0.882	0.861	0.860	0.881	0.880	0.847	0.847

Note: Panel (a) displays estimates of the coefficients associated with lags of the fine, $[\gamma_{-J}, \ldots, \gamma_0]$, and with leads of the fine, $[\gamma_1, \ldots, \gamma_J]$, in Equation (2) binning 20 lags of the fine in one bin and 10 leads of the fine in one bin. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. The columns labeled "Agricultural" show results from a specification considering both lags and leads of the fine and from a specification setting the coefficient associated with leads of the fine to 0 for agricultural-*hukou* women. The dependent variable is the number of children at age a and the controls are individual and age $\times$ year fixed effects. Inference is province $\times$ birth year two-way clustered. Asymptotic standard errors are in parentheses. The rest of the panels are analogous for demographic groups as labeled. A household is classified as wealthy if it had broadband internet connection at the time of the CHARLS interview. Broadband internet connection is provided for a narrow set of privileged households in China (Zhu and Wang, 2005). "SOE's" stands for state-owned enterprises. The sample is an unbalanced panel of $8,023$ women from the cohorts born in 1926 to 1972 who were between ages 15 to 45 in the years 1970 to 20000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

urban-rural difference could be driven by different extents to which the policy was binding rather than by enforcement heterogeneity.

Estimates by mother's education and household wealth indicate that women with lower socioeconomic status were less sensitive to the policy. This corresponds with the fact that, often, women facing economic difficulties were exempt from the policy as I document in Tables A.1 to A.5 in the Appendix. Finally, younger women, who were most fecund, were much more sensitive to the policy compared to older women. I consider this life-cycle profile and the sex of the first child as the main sources of heterogeneity when providing estimates of the aggregate effect of the policy in the next section.

5 Aggregate Policy Effect

With the household-level policy effects available, I can now provide conditions to construct counterfactual aggregate statistics in the absence of the policy. Specifically, I provide the conditions to construct the average number of children for women ages 15 to 45 between 1979 and 2010 in a no-policy scenario (Figure 1). This requires me to estimate the number of children that women did not have due to the fines, which is a level. Identifying such a level would then allow me to combine it with the number of children in the realized scenario and obtain estimates in a no-policy counterfactual.

The model in Equation (2) allows me to estimate the marginal effects of the fines associated with the policy. When identifying these marginal effects, the fact that the policy could have imposed a fixed cost is not a specification concern. The derivative of the fixed cost would be 0, and thus carefully specifying this cost is not necessary. So far I have subsumed this cost into $\phi(a, t)$. When identifying the effect of the policy in levels, it is necessary to assess the fixed cost. A fixed cost could be economically relevant in this context. For example, apart from the fine, women could have been subject to public shaming or could have been fired from their jobs if they had more than one child.

Let ζ_0 denote this fixed cost. To separate the fixed cost from the age and time effects, parametric assumptions are required. As discussed before, $\phi(a,t)$ is not non-parametrically identified. I impose the functional form assumption that $\phi(a,t) := \phi_A(a) + \phi_T(t)$, where $\phi_A(a)$ is a function without a constant—consistent with the fact that women at ages 15 and 16 have no children in my sample—and $\phi_T(t) := \varphi_0 + \varphi_1 t + \varphi_2 t^2$ is a second-order degree polynomial. Imposing these parameterizations on Equation (2) yields

$$n_{ia} = \lambda_i + \zeta_0 + \varphi_0 + \phi_A(a) + \varphi_1 t + \varphi_2 t^2 + \sum_{j=-J}^{J} \gamma_j \tau_{ia+j} + \varepsilon_{ia}. \tag{3}$$

The number of children that women i at age a did not have due to the policy is $\zeta_0 + \sum_{j=-J}^{J} \gamma_j \tau_{ia+j}$. The counterfactual number of children in a no policy scenario is n_{ia} less this quantity. If $|\zeta_0| > 0$, calculating the counterfactual based on $\sum_{j=-J}^{J} \gamma_j \tau_{ia+j}$ would generate a biased estimate—off by the amount ζ_0. The issue at stake is that ζ_0 and φ_0 are not separately identified.

Unfortunately, there is no credible identification strategy for obtaining ζ_0. Even though a very small subset of communities were uniformly exempt from the policy for a number of years and could be a control group, they still could have been exposed to the fixed cost. Thus, I estimate the counterfactual number of children of women i at age a by adding back $\sum_{j=-J}^{J} \gamma_j \tau_{ia+j}$ to n_{ia} and provide sensitivity analysis to multiple parameterizations of the age and year effects.

Sensitivity analysis is provided in Panel (a) of Table 2. The effect of the fines remains robust to multiple parameterizations. If the effects were not robust, then it would be more likely for the age and year effects to contain ζ_0 and for the counterfactual to be biased. Alternatively, the results in this section could be interpreted as a scenario with no fines—but potentially with other costs imposed by the policy. The fines, as I explain in Section 2, were the main device used to implement the policy.

Once I add $\sum_{j=-J}^{J} \gamma_j \tau_{ia+j}$ back to n_{ia}, forming the no-policy counterfactual in Figure 1 is straightforward; it simply requires forming yearly averages. When estimating $\sum_{j=-J}^{J} \gamma_j \tau_{ia+j}$, I allow heterogeneity in γ_j by age and sex of the first child. These are the main sources of heterogeneity documented in Section 4.

The comparison between the realized and no-policy scenarios indicates that 12.5% (s.e. 1%) of the decrease in the average number of children for women ages 15 to 45 between 1979 and 2000 was due to the policy. Calculating the decrease between 1979 and 2010 requires one further assumption because data on the fines are not available between 2001 and 2010. To calculate the percentage of the decrease between 1979 and 2010, I assume that the policy remained constant after 2000. This assumption is consistent with informal documentation (White, 2006). Based on this assumption, I obtain an estimate of 17% (s.e. 1%).

The aggregate effect of the One-Child Policy is, perhaps, seemingly minor, especially so for an iconic policy that has even been publicly labeled as a humanitarian crisis (Keyi, 2015). The fact that the response to the policy was relatively inelastic does not mean that there was no response. A simple back-of-the-envelope calculation based on the average effects that I report in Section 4 suggests that millions of children were not born as a consequence of the policy. For example, the average prediction of the number of children that women aged 15 to 30 did not have (women who were mainly affected by the policy according to Table 3) times the number of women who were of this age between 1979 and 2010 yields 43.8 million fewer children as a result of the policy.[30]

6 Effects by Sex and the Interaction with Ultrasound

If the policy prevented millions of births, it is of interest to investigate the sex composition of the effects, epecially after inspecting the gendered effects in Table 1.

Table 4 displays the estimates for all children in Panel (a) of Table 1. It then presents analogous results for cases in which the dependent variables are the number of girls or the number of boys. The results indicate that the effect of the policy is driven by a reduction in the average number of girls born per woman, not in the average number of boys. This necessarily has to do with some form of parental sex selection. While there is no sex imbalance at the first birth, there is imbalance at higher-order parities—as documented in Figure A.4—and previous studies have pointed to the One-Child Policy as the likely culprit (Ebenstein, 2010b).

If prenatal sex selection drives the results in Table 4, then it has to be the case that these effects interact with the availability of ultrasound technology. Panel 6a shows the estimated effects for specifications with the number of girls as the dependent variable, dropping all leads of the fine, by availability of ultrasound technology (left) and the analogous effects for specifications with the number of boys as the dependent variable (right).

If the effects in Table 4 were completely driven by ultrasound availability, the effect on boys and girls would be the same when the technology is unavailable. If this technology were randomly allocated in the population, the sum of the effects on girls and boys should add up

[30]These are the women who were age 15 to 30 between 1979 and 2010. That is, the women who were potentially restricted by the policy between ages 15 and 45. I estimate this number of women to be 435.2 million, using the same strategy as in Section 2.

Table 4: Baseline Effect on Girls and Boys

	Girls and Boys		Girls		Boys	
Leads	-0.004		-0.009		0.009	
	(0.011)		(0.007)		(0.007)	
Lags	-0.086	-0.101	-0.066	-0.074	0.011	0.002
	(0.015)	(0.014)	(0.011)	(0.011)	(0.009)	(0.009)
N	8,023	8,023	7,660	7,660	7,899	7,899
R^2	0.879	0.879	0.853	0.853	0.863	0.862

Note: This table displays estimates of the coefficients associated with lags of the fine, $[\gamma_{-J}, \ldots, \gamma_0]$, and with leads of the fine, $[\gamma_1, \ldots, \gamma_J]$, in Equation (2) binning 20 lags of the fine in one bin and 10 leads of the fine in one bin. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. The first two columns labeled show results from a specification considering both lags and leads and from a specification setting the coefficient associated with leads of the fine to 0. The dependent variable is the number of children at age a (girls and boys) and the controls are individual and age $\times$ year fixed effects. Inference is province $\times$ birth year two-way clustered. Asymptotic standard errors are in parentheses. The next columns present the analogous results when the dependent variable is the number of girls or the number of boys. The sample is an unbalanced panel of $8,023$ women from the cohorts born in 1926 to 1972 who were between ages 15 to 45 in the years 1970 to 20000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

to the same quantity regardless of availability of the technology. Neither is the case, making the evidence suggestive. The results, however, are intuitive and do suggest a prenatal sex selection pattern. The effect on the number of boys is precisely zero where the technology is available and negative, although noisy, where it is not. The fact that the effect on boys and girls is not the same where the technology is unavailable could hint at women making decisions other than prenatal sex selection to obtain their desired number of children (e.g., giving girls up for adoption and not reporting them as ever being born).[31] The fact that the effects on boys and girls do not add up to the same quantity across regions with different availability could hint at differences in preferences about family size across these regions.

Panel 6b presents an alternative to Panel 6a by displaying the analogous results by community-level availability of formal health care. The government provided ultrasound machines to formal health facilities, so the availability of these facilities might approximate ultrasound availability. This proves useful because I only observe data on ultrasound technology availability for 48% households in my sample.[32] The results align, but in this case, the effects on girls and boys are very similar in the absence of health care facilities, making a stronger case for the interaction of the policy with prenatal sex selection.

This highlights an important policy implication of my results. Family-planning policies interact with parental preferences for the number of girls and boys. My results indicate that this interaction caused the aggregate decrease in fertility due to the policy to be driven by a reduction in the number of girls born, as opposed to a reduction in the number of boys,

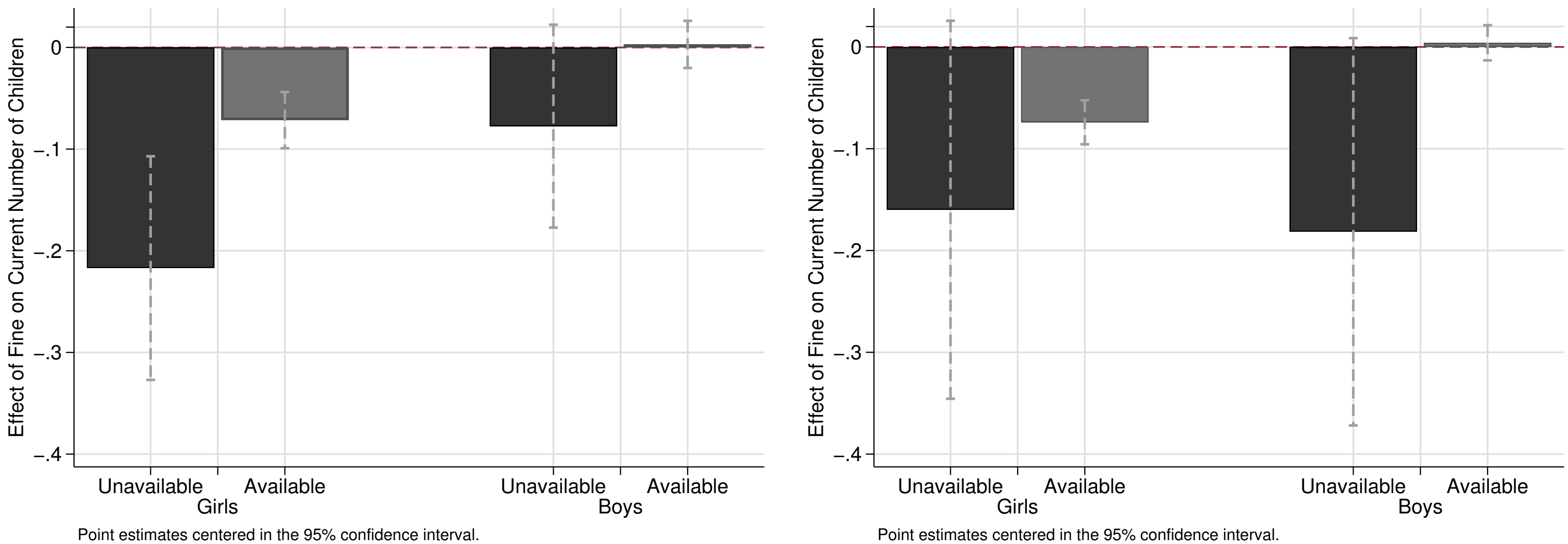
[31] Ebenstein (2010b) argues that formal and informal statistics on adoptions cannot reconcile sex imbalance at birth of children beyond the first parity. His arguments and results are consistent with Chen, Li, and Meng (2013).

[32] In Table A.12 in the Appendix, I verify that the results in Table 4 are very similar across subsamples with and without ultrasound availability data, reducing concerns that the sex differences in the policy effects are based on unobserved differences in the communities for which ultrasound data are available.

Figure 6: Investigating the Baseline Effect by Sex

(a) By Ultrasound Availability

(b) By Formal Health Care Availability



Note: Panel (a) displays estimates of the coefficients associated with lags of the fine, $[\gamma_{-J}, \ldots, \gamma_0]$, and setting the coefficients associated with leads of the fine, $[\gamma_1, \ldots, \gamma_J]$, to zero in Equation (2) binning 20 lags of the fine in one bin and 10 leads of the fine in one bin by availability of at least one ultrasound machine for sex detection at the community level. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. The dependent variable is indicated in the label (either number of girls or number of boys). Inference is province $\times$ birth year two-way clustered. Asymptotic confidence intervals are displayed. Panel (b) is analogous by availability of formal health care facilities at the community level. The sample is an unbalanced panel of $8,023$ women from the cohorts born in 1926 to 1972 who were between ages 15 to 45 in the years 1970 to 20000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), the sources explained in the Appendix for availability of ultrasound technology, and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

making the policy one of the causes for the "missing women" in China (for an introductory discussion, see Sen, 1991).

7 Discussion

Comparison to Readily Available Policy Quantifications. Part of the motivation and relevance of my study is that current estimates of the policy effect on fertility rely either on the inspection of aggregate trends or, when using microeconomic data, on incomplete characterizations of the policy. I compare my results with results obtained when using current, readily available characterizations of the policy. I do so by providing an estimate of the aggregate decrease in fertility that I would have obtained in Section 5 had I used these incomplete characterizations.

Table 5: Aggregate Decrease in Fertility, Comparing Strategies

Decrease Span	Estimate		
1979-2000	-0.016	0.043	0.125
	(0.022)	(0.008)	(0.009)
1979-2010	0.081	0.098	0.170
	(0.013)	(0.006)	(0.007)
Strategy for Calculating Household-Level Policy Effects			
Girl × Rural exemption	No	Yes	No
Documented exemptions	No	No	Yes
Individual FE	No	Yes	Yes
Age × Year FE	Yes	Yes	Yes

Note: This table shows the decrease in the average number of children for women with birth years between 1926 and 1972 ages 15 to 45 in the periods 1979-2000 and 1979-2010 due to the policy. The decrease is estimated as explained in Section 5 for the three strategies indicated in the table for calculating the household-level policy effects on fertility. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

First, I replicate the characterization in Ebenstein (2010b) when constructing τ_{ia}. That is,

(i) I do not use the exemptions to the policy documented in Section 2; and (ii) I rely on province × year cross-sectional variation in τ_{ia}, i.e., I do not exploit within-woman variation when estimating the household-level policy effects. These results are reported in the first column of Table 5, while the results in this paper are in the third column. This strategy actually predicts an increase in fertility for the period 1979 to 2000, which corresponds with τ_{ia} being endogenous to province × year fertility patterns. This bias decreases when widening the period of analysis to include the years 2001 to 2010. My strategy overcomes this bias by exploiting within-woman policy variation.

Second, I replicate the characterization in Qian (2009). That is, when constructing τ_{ia}, I only use the exemption documented in Section 2 related to the sex of the first child.[33] These results are reported in the second column of Table 5. If compared to my strategy, the strategy in Qian (2009) produces estimates that are substantially biased towards 0, which is sound because the only difference across strategies is the quality of measurement in policy intensity.[34]

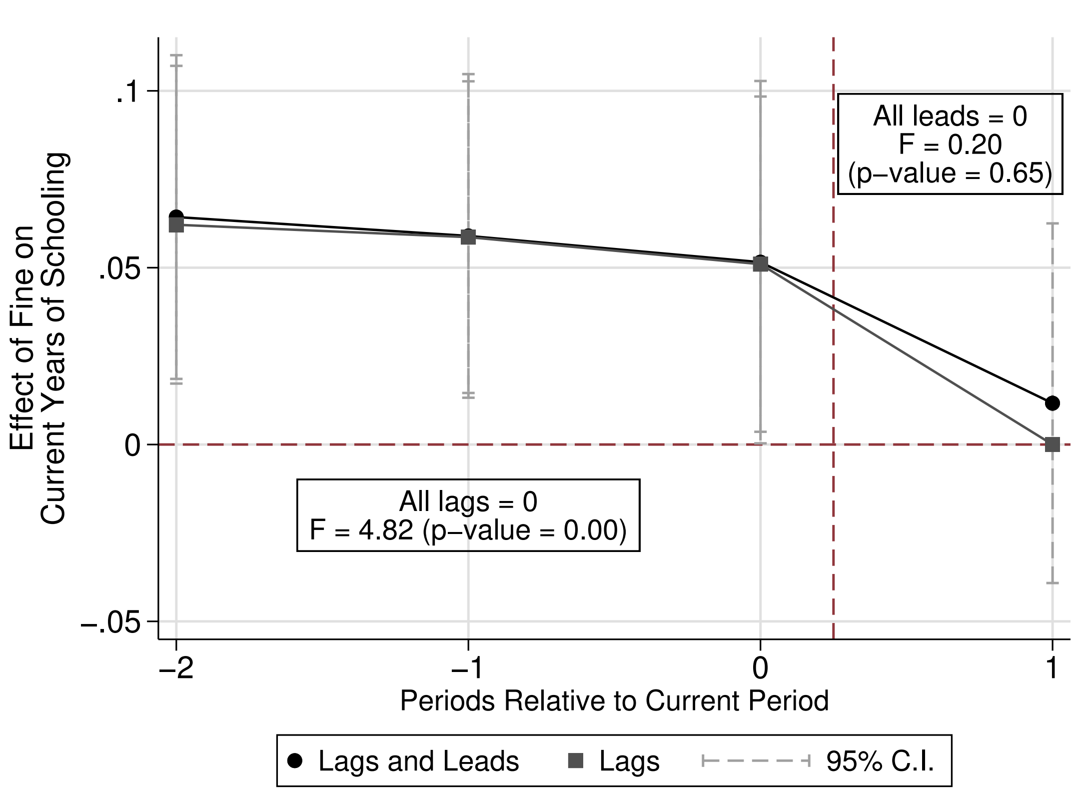
Other Outcomes Affected by the Policy. To finalize my empirical analysis I ask: What other outcomes could have been affected by the One-Child Policy? I focus on the intergenerational effect on education because it is an outcome considered to be related to fertility in economics since the seminal work of Becker and Lewis (1973).

I display results analogous to those in Figure 5 with accumulated years of child's schooling as an outcome. The unit of observation, in this case, is the child or the children of the mother

[33]Qian (2009) does not exploit within-woman variation. I interpret this empirical decision as induced by data restrictions and not by a characterization shortcoming. For a more informative comparison, I choose the replication exercise that I report.

[34]Note that neither Qian (2009) nor Ebenstein (2010b) focus on fertility as an outcome of analysis. They use the policy as a vehicle to study other, related economic phenomena—the quantity-quality trade-off and sex ratios at birth, respectively.

Figure 7: Effect of Past and Future Fines on Children's Education, Event Study



Note: This figure displays estimates of the coefficients associated with lags of the fine, $[\gamma_{-J}, \ldots, \gamma_0]$, and with leads of the fine, $[\gamma_1, \ldots, \gamma_J]$, in Equation (2) binning 20 lags of the fine in three bins and 10 leads of the fine in one bin. The binned averages of the fines are standardized to an in-sample mean and standard deviation of 0 and 1. The "lags" specification sets the coefficients on leads of the fine to 0. The dependent variable is the number of years of schooling for each child in a household when her or his mother is a years old and the controls are individual and child's age $\times$ mother's age $\times$ year fixed effects. Inference is province $\times$ birth year two-way clustered. Asymptotic confidence intervals are displayed. The sample is an unbalanced panel of $8,023$ women from the cohorts born in 1926 to 1972 who were between ages 15 to 45 in the years 1970 to 20000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

who was potentially restricted by the policy—multiple children per women could appear in the sample; I rely on within-child variation and account for child's age $\times$ mother's age $\times$ year fixed effects. The results indicate that household-level fertility restrictions caused a substantial and persistent increase in education. A one standard deviation increase in the fine at 0 increases accumulated years of child's schooling by 0.05 years—the analogous effect on the number of children is -0.01. Although not the focus of this paper, this indicates that the policy had effects on household-level outcomes beyond fertility. A more exhaustive investigation of these effects is a matter for future research and is possible to carry out given the data collection in this paper, which includes intergenerational outcomes such as education and labor income.

8 Conclusion

The conventional and practical characterizations of the One-Child Policy do not allow for a conclusive analysis of its effect on fertility. The One-Child Policy was neither a quota system nor a system of regional incentives. It was not a mandate of fertility limits exempting ethnic minorities or women whose first child was a girl. Rather, the policy was a pricing system allowing every woman and her partner to have more than one child—if they paid a price. In this paper, I provide a novel data construction that allows me to exploit the pricing system to study the effect of the policy on household-level fertility relying on within-woman variation in policy intensity. I then provide the conditions to use these estimates to quantify the fraction of the total decrease in fertility due to the policy.

I document three novel facts. The policy only affected women whose first child was a girl. That is, it was not binding for women whose first child was a boy. The policy reduced the average number of girls born per woman, but not the average number of boys. Suggestive evidence using data on ultrasound availability indicates that this was driven by prenatal sex selection. And, finally, of the observed decrease in fertility between 1979 and 2010, a relatively minor fraction, 17%, was due to the policy.

My results provide two policy implications. First, even in a context in which the government had thorough control over household economic decisions, the response to the policy was relatively inelastic. The pricing system known as the One-Child Policy was relatively ineffective at curbing fertility, while a quota system, as the policy is conventionally perceived, would have been very welfare decreasing. Second, any family-planning policy interacts with parental preferences for the number of daughters and sons, which could generate sex-ratio imbalances in the population.

Looking forward, what will be the effect on fertility after the national introduction of a uniform Two-Child Policy in 2015? After 2000, both average completed fertility and the total fertility rate have remained virtually constant and around 1.5 (The World Bank, 2017). Thus, the Two-Child Policy likely relaxes a non-binding constraint and will have little effect on fertility.

9 Appendix

Criteria for Additional-Child Permits. Table A.1 lists the different criteria to obtain free permits for children additional to the first. There was province and time variation in the criteria. Tables A.2, A.3, A.4, and A.5 document this variation. These criteria generate demographic, temporal, and spatial variation in the average fine for having more than one child. Table A.6 describes the empirical construction of the criteria in Table A.1.

Table A.1: Criteria for Acquiring Permits to Have More than One Child, 1979-2000

1. First child is disabled or dead	2. Pregnancy after long years of childless marriage and a subsequent adoption	3. In remarriage one spouse has been childless, the other spouse already had one or two children
4. One or both spouses returned to China from Hong Kong or Taiwan	5. One or both spouses belong to a national minority with less than 10 million members	6. **One spouse is disabled and cannot work**
7. **A peasant couple lives in sparsely settled mountain, reclamation, or border seas**	*8. One spouse is a deep-sea fisherman*	9. **One spouse has been constantly working in underground mining for more than 5 years**
10. *One spouse or both spouses are single children*	11. *Only one child or one son has been born to a family for two generations*	*12. Among brothers, only one is able to produce children*
13. *Husband settles in the family of his wife which has daughters and no sons*	14. One spouse is the (single) child of a revolutionary martyr	15. **Couple has real economic difficulties or claims other peculiar reasons**
16. First child is a girl (and couple has economic difficulties)	17. Three, four, or five years after birth of first child	

Note: This table lists the criteria that qualified women for additional-child permits. These criteria applied differently by province and time as indicated in Tables A.2, A.3, A.4, and A.5. When linking these criteria to household-level micro-data, I can determine whether a household qualified for an additional-child permit by satisfying any of the criteria except for the three cases in *italics* due to lack of data. For the statistics in the paper, I classify as "first child dead, disabled, or older" the categories in gray, as "job difficulty" the categories in **bold**, and as "male scarcity in the extended family" the starred categories (*). The categories in black (no italics) are classification themselves. **Source:** Adapted from Scharping (2003).

Table A.2: Criteria for Acquiring Permits to Have More than One Child 1979-2000, 1/4

Province	Years	Agricultural and Non-Agricultural Women	Additional Criteria for Agricultural Women
Beijing	1979–1981	none	none
	1982–1990	1,2,3,4,5	7,10,12,13
	1991–2000	1,2,3,4,5,15	6,(7),12,13,(16)
Tianjin	1979–1981	none	none
	1982–1987	none	1,2,3,6,10,12,15
	1988–1992	1,2,3,4,(5),10	6,12,13,(16)
	1993–1996	none	none
	1997–2000	1,2,3,4,(5),10	6,12,13,(16)
Hebei	1979–1981	none	none
	1982–1985	1,2,3	5,7,8,10,12,13
	1986–1988	1,2,3,9	5,7,8,10,12,13
	1989–2000	1,2,3,4,5,6,9,10,15	(7),8,13,(16)
Shanxi	1979–1981	none	none
	1982–1985	1,2,3,4,5	6,7,10,12,13,14
	1986–1988	1,2,3,15	10,12,13
	1989–2000	1,2,3,4,5,6	7,(12),13,(16)
Anhui	1979–1980	17	none
	1981–1983	1,2,3,5	none
	1984–1987	1,2,3,4,5,6,9,10	7,12,13
	1988–1991	1,2,3,4,5,6,9,10,12	13,16
	1992–2000	1,2,3,4,5,6,(9),10,12	13,16
Fujian	1979–1981	none	none
	1982–1987	1,2,3,5	13,15
	1988–1990	1,2,3,4,5,6,8,9,14	7,10,12,13
	1991–2000	(1),(2),(3),(4),(5),(6),(8),(9),(10),(14)	7,11,12,13
Jiangxi	1979–1982	none	none
	1983–1984	1,2,3	7,10,12,13
	1985–1989	1,2,3,4,5,9,10	7,10,12,13,16
	1990–1994	1,2,3,4,(5),(9),14,15	11,12,(13),16
	1995–1996	none	none
	1997–2000	1,2,3,4,(5),(9),14,15	11,12,(13),16
Shandong	1982–1987	1,2,3	10,12,13,14
	1988–1995	1,2,3,4,5,(6),8,(9),15	8,12,13,14,16
	1996–2000	1,2,3,4,5,(6),8,(9),10,15	8,12,13,14,16
Henan	1982–1984	1,2,3,7,12	none
	1985–1986	1,2,3,4,5,6,9,10,14,15	7,11,12,13
	1987–1989	none	(16)
	1990–2000	1,2,3,4,(9),14	5,(7),13,16

Note: This table lists the province and time variation in the criteria qualifying women for additional-child permits listed in Table A.1. When the criterion is in parentheses, the household had to be socio-economically disadvantaged for the permit to be granted. **Source:** Adapted from Scharping (2003).

Table A.3: Criteria for Acquiring Permits to Have More than One Child 1979-2000, 2/4

Province	Years	Agricultural and Non-Agricultural Women	Additional Criteria for Agricultural Women
Hubei	1979–1986	none	none
	1987–1990	1,2,4	6,10,13,6
	1991–2000	1,2,3,4	6,11,13,16
Hunan	1979–1981	(17)	none
	1982–1988	1,2,3,4	(15)
	1989–2000	1,2,3,4,5,10	11,12,13,14,(16)
Guangdong	1979	none	none
	1980	(17)	none
	1981–1985	1,8,9	3,7,16
	1986–1991	1,2,3,4,5,9,10	(16),(17)
	1992–1996	1,2,3,(4),(5),9,10	(17)
	1997	none	none
	1998–2000	1,2,3,(4),(5),9,10	16
Guangxi	1979–1981	none	none
	1982–1984	1,2,3,6	5,15
	1985–1987	1,5,6,9,10	7,13,16
	1988–2000	1,2,35,6,19,14	7,12,13,16
Hainan	1979–1983	none	none
	1984	1,2,3,9,10	7,8,16
	1985–1988	1,2,3,9,19	16
	1989–1994	1,2,3,4,5,(6),(9),(14)	15
	1995–2000	1,2,3,4,5,15	none
Chongqing	1979–1996	none	none
	1997–2000	1,2,3,4,10	6,7,11,12,13,14,(16)
Sichuan	1979–1981	none	none
	1982–1983	1,2,3,4,15	15
	1984–1986	none	6,7,12,13,14,(16)
	1987–1992	1,2,3,4,10	6,7,12,13,14,(16)
	1993	none	none
	1997–2000	1,2,3,4	6,7,12,13,(16)
Guizhou	1979–1981	17	none
	1982–1983	1,2,3	5,15
	1984–1986	1,2,3,5,6,10	11,12,13,15
	1987–1997	1,2,3,4,5,10	13,15
	1998–2000	1,2,3,4,10	13
Yunnan	1979–1985	none	none
	1986–1989	1,2,3,4,5,(6),10,13	5,7,17
	1990–2000	1,2,3,4,(6),10	(5),(7),15

Note: This table lists the province and time variation in the criteria qualifying women for additional-child permits listed in Table A.1. When the criterion is in parentheses, the household had to be socio-economically disadvantaged for the permit to be granted. **Source:** Adapted from Scharping (2003).

Table A.4: Criteria for Acquiring Permits to Have More than One Child 1979-2000, 3/4

Province	Years	Agricultural and Non-Agricultural Women	Additional Criteria for Agricultural Women
Inner Mongolia	1979–1981	none	none
	1982–1984	1,2,3,5	15
	1985–1987	1,2,3,4,5,9,10,15	6,7,12,13,16
	1988–1989	1,2,3,4,5,9	6,7,12,13,16
	1990–2000	1,2,3,5,9,15	6,16
Liaoning	1979	17	none
	1980–1981	1,3	none
	1982–1983	1,3	5
	1984	1,2,3,10	5,8,12,13
	1985–1986	1,2,3,19	5,8,12,13,(16)
	1988–2000	1,2,3,(5),19,15	5,8,12,13,16
Jilin	1988–1992	1,2,3,4,5,10,15	6,11,12,16
	1993–1996	1,2,3,4,5,10	(6),12,16
	1997–2000	1,2,3,4,5,10,15	(6),12,16
Heilongjiang	1979	1,17	none
	1983–1988	1,2,3,5	10,12,13
	1989–1993	1,2,3,4,5,6,10	12,13,16
	1994–1998	1,2,3,4,5,15	7,10,12,16
	1999–2000	1,2,3,4,5,10,15	7,16
Shanghai	1979–1980	none	none
	1981	1,2,3	none
	1982	none	none
	1984	1,2	3,6,10,12,13
	1987	1,2,3,4,9	5,6,8,10,12,13,14
	1992	1,2,3,4,(5),10,15	6,(8),(13)
Jiangsu	1979–1981	1,17	none
	1982–1984	1,2,3	10,12
	1985–1989	1,2,3,4,8,10,14,15	7,12,13
	1990–1994	1,2,3,4,(8),(9),10,11,14,15	7,12,13,(16)
	1995–2000	1,2,3,4,(9),10,11,14,15	7,(8),12,13,(16)
Zhejiang	1979–1981	none	none
	1982–1984	1,2,3,5	(15)
	1985–1988	1,2,3,4,5,9,10,14,15	6,(7),(8),12,13,(15)
	1989–2000	1,2,3,4,5,(9),10,14,15	(7),(8),11,13,(16)

Note: This table lists the province and time variation in the criteria qualifying women for additional-child permits listed in Table A.1. When the criterion is in parentheses, the household had to be socio-economically disadvantaged for the permit to be granted. **Source:** Adapted from Scharping (2003).

Table A.5: Criteria for Acquiring Permits to Have More than One Child 1979-2000, 4/4

Province	Years	Agricultural and Non-Agricultural Women	Additional Criteria for Agricultural Women
Shaanxi	1979–1980	17	none
	1981	1,2,3,5,7	none
	1982–1984	1,2,3,4,5	7
	1985	1,2,3,4,5,6	7,11,12,13
	1986–1987	1,2,3,4,5,10,15	6,7,12,13
	1988–1990	15	(7),(16)
	1991–2000	1,2,3,4,5,10	6,7,13,16
Gansu	1979–1981	none	none
	1982–1984	1,2,3	15
	1985–1988	1,2,3	15,16
	1989–1996	1,2,3	5,13, (7 and 16)
	1997–2000	1,2,3,4	5,13,14,15,16
Qinghai	1979–1981	none	none
	1982–1984	3,15	none
	1985	3,4,6,10	none
	1986–2000	1,2,3,4,5,6,10	none
Ningxia	1979	none	none
	1980–1981	17	none
	1982	1,2,3	17
	1986	1,2,3,4,5,6,10,15	17
	1990	1,2,3,4,5,9,10	17
Xinjiang	1979–1980	none	none
	1981–1987	5	none
	1988–1990	5	none
	1991–2000	1,3,4,5,6,9,10	(17)

Note: This table lists the province and time variation in the criteria qualifying women for additional-child permits listed in Table A.1. When the criterion is in parentheses, the household had to be socio-economically disadvantaged for the permit to be granted. **Source:** Adapted from Scharping (2003).

Table A.6: Empirical Construction of Eligibility for the Criteria in Table A.1, 1/2

	Wave and Relevant Data File	Construction Summary
1	Wave 2013, Child	Year of birth of all children provided; possible to indicate whether or not first child died
2	Wave 2013, Child	Year of birth and biological origin of children provided
3	Wave 2013, Child	Number of non-biological children available
4	N/A	Previous residence in Hong Kong or Taiwan not available
5	Wave 2014, Demographic	Data on ethnicity available
6	Wave 2011, Work, Retirement, Pension	Data on reason for unemployment available (e.g., disability)
7	Wave 2011, Community	Type of landscape provided; classify as sparse if population density $<$ within-sample 5^{th} percentile
8	N/A	Unable to classify as deep-sea fisherman

Note: This table describes the empirical procedure for assigning women as complying to the criteria in Table A.1, based on the questions in the China Health and Retirement Longitudinal Survey (National School of Development, 2009-2017). When the conditions in Tables A.2 to A.5 are in parentheses, the authorities required, in addition, for households to be in socio-economic disadvantage. I classify a household as in socio-economic disadvantage if the community where it resides is below the within-sample 1^{st} percentile of average household income. The relevant data files (e.g., Child, Work, Retirement) are labeled as when provided by the China Health and Retirement Longitudinal Survey. **Source:** Author's creation.

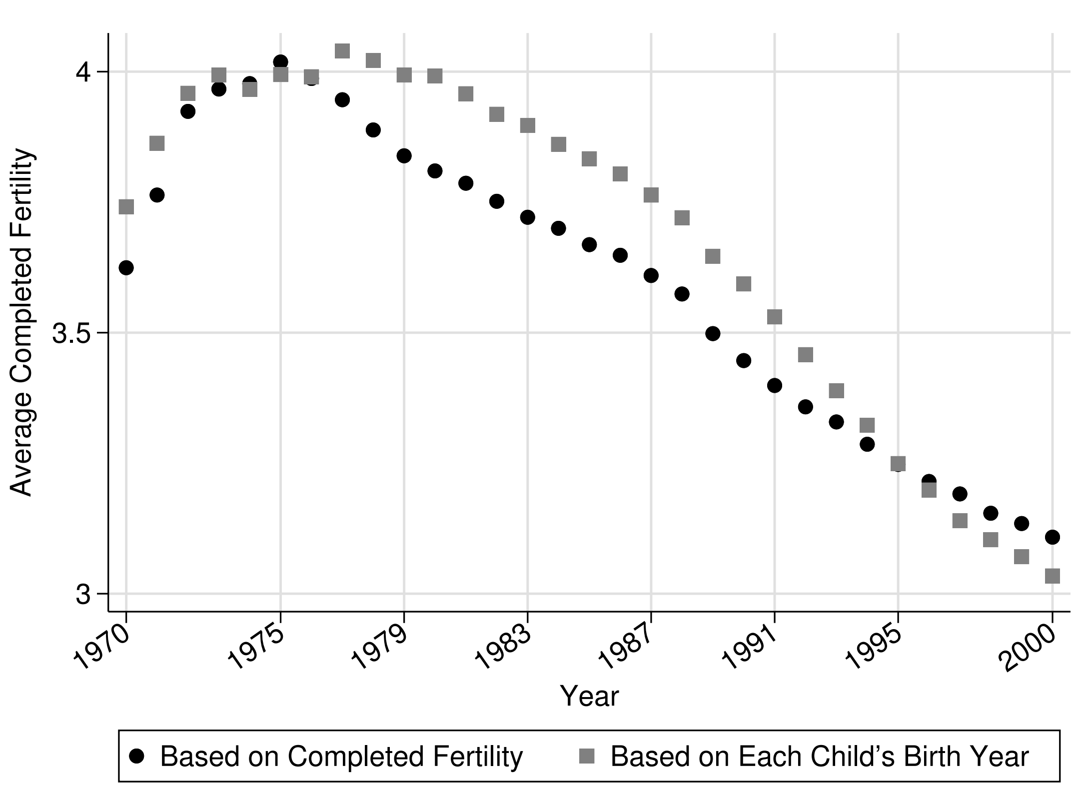
Table A.7: Empirical Construction of Eligibility for the Criteria in Table A.1, 2/2

	Wave and Relevant Data File	Construction Summary
9	Wave 2011, Community	Main job individuals seek available; mining available
10	Wave 2011, Family	Number of respondent's and spouse's siblings available
11	Wave 2011, Family	Number of respondent's and spouse's siblings available
12	N/A	Detailed information on brothers' fertility is not available
13	Wave 2011, Family	Number and sex of respondent's and spouse's siblings available
14	Wave 2014, Wealth History	Information of martyr status of parent available
15	Wave 2011, Community	Classify as being in economic disadvantage if community of residence < within-sample 1st percentile of average household income
16	Wave 2011, Child	Sex of all children provided; economic disadvantage as in 15
17	Wave 2011, Child	Year of birth of all children provided; use three year spacing

Note: This table describes the empirical procedure for assigning women as complying to the criteria in Table A.1, based on the questions in the China Health and Retirement Longitudinal Survey (National School of Development, 2009-2017). When the conditions in Tables A.2 to A.5 are in parentheses, the authorities required, in addition, for households to be in socio-economic disadvantage. I classify a household as in socio-economic disadvantage if the community where it resides is below the within-sample 1st percentile of average household income. The relevant data files (e.g., Child, Work, Retirement) are labeled as when provided by the China Health and Retirement Longitudinal Survey. **Source:** Author's creation.

Retrospective and Realized Number of Children. A concern when analyzing fertility longitudinally is that I use a single retrospective survey asking about each child's birth year to construct individual-level fertility histories. The survey took place a number of years after women completed their fertility. Figure A.1 compares yearly average completed or total fertility using two sources. The first source is based on a single question asking about completed or total fertility, which is unlikely to be measured with error. The second source is based on constructing completed fertility from the questions that I use to construct each woman's fertility history (a set of questions asking about each child's birth year), which could be measured with error if women do not recall the birth years of their children. The two sources are closely aligned lessening measurement error concerns.

Figure A.1: Retrospective and Realized Total Number of Children



Note: This figure displays the average completed or total number of children in the China Health and Retirement Longitudinal Survey (National School of Development, 2009-2017) using two sources: (i) A single question asking about completed fertility; and (ii) A set of questions asking about each child's birth year. **Source:** Author's calculations, created with information from the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

Ultrasound Technology Data Availability. I collected information on ultrasound technology availability for 68 communities in 23 provinces from two sources: city-level "New China Local Gazetteers" and the "Reviews of Ultrasound Medicine in China." The New Local Gazetteers began to be published after 1949, and cover all general facts related to specific geographical areas such as economy, public security, agriculture, public health, natural science, and industry.

The New Local Gazetteers are available on the national, provincial, county, and city levels. In addition to the general gazetteers, there are topic-specific gazetteers for larger cities such as the Local Health Gazetteer. For 64 communities, I use the city-level Local Health Gazetteer if available, or the Local Gazetteer if not, to find information on when the first ultrasound machine was available.

Four major cities/communities (Beijing, Ningbo, Shenzhen, Chongqing) were chosen to be a part of the Reviews of Ultrasound Medicine in China conference, and thus there are detailed essays of ultrasound usage in the respective cities. The China Society of Ultrasound Medical Engineering hosted the conference in 2008 to celebrate 50 years of ultrasound diagnosis in China.

For each community, I record the year that ultrasound machines were first introduced. The sources that I use and the list of corresponding citations are available upon request. Section 6 in the main paper explains how I use these data.

Evaluation Form for Local Birth-Planning Officials. Table A.8 presents an example of the forms used to evaluate birth-planning officials' performance by the community-level government. The performance depended on local aggregate fertility and citizens' knowledge of the policy.

Table A.8: Evaluation Form for Birth-Planning Officials, 1992

	Maximum credit points		
A. Birth-planning targets	35		
1. Target formulation		12	
according to upper-level mandate, by scientific reasoning			12
self-decreed, without sufficient basis			6
transmitted by upper level			4
no targets existing,			0
2. Target fulfillment		23	
fulfilled, birth rate < 15%			23
not fulfilled, birth rate 15-16%			16
not fulfilled, birth rate 16-18%			10
not fulfilled, birth rate > 18%			4
B. Birth-planning work	35		
1. Propaganda work		12	
> 80% of couples of reproductive age grasp present birth policy			12
70-80% of couples of reproductive age grasp present birth policy			9
60-70% of couples of reproductive age grasp present birth policy			6
< 60% of couples of reproductive age grasp present birth policy			3
2. Technical services		12	
contraceptive prevalence (married women of reproductive age) > 80%			12
contraceptive prevalence (married women of reproductive age) 70-80%			9
contraceptive prevalence (married women of reproductive age) 60-70%			6
contraceptive prevalence (married women of reproductive age) < 60%			3
3. Policy implementation		11	
penalty rate for above-quota births > 95%			11
penalty rate for above-quota births 80-95%			8
penalty rate for above-quota births 65-80%			5
penalty rate for above-quota births < 65%			2
C. Economic benefits of birth planning	15		
good economic benefits			15
satisfactory economic benefits			10
medium economic benefits			6
poor economic benefits			2
D. Social benefits of birth planning and image of birth-planning personnel	15		
> 90% of canvassed population have good opinion of birth-planning dept.			15
80-90% of canvassed population have good opinion of birth-planning dept.			10
70-80% of canvassed population have good opinion of birth-planning dept.			6
< 70% of canvassed population have good opinion of birth-planning dept.			2
Head of evaluation team (signature)	Total credit points:		

Note: This tables reproduces an evaluation form for birth-planning officials in towns and townships. **Source:** Adapted from Li Jinfeng (1992).

Descriptive Statistics. Table A.9 provides summary statistics for the main variables in my empirical analysis.

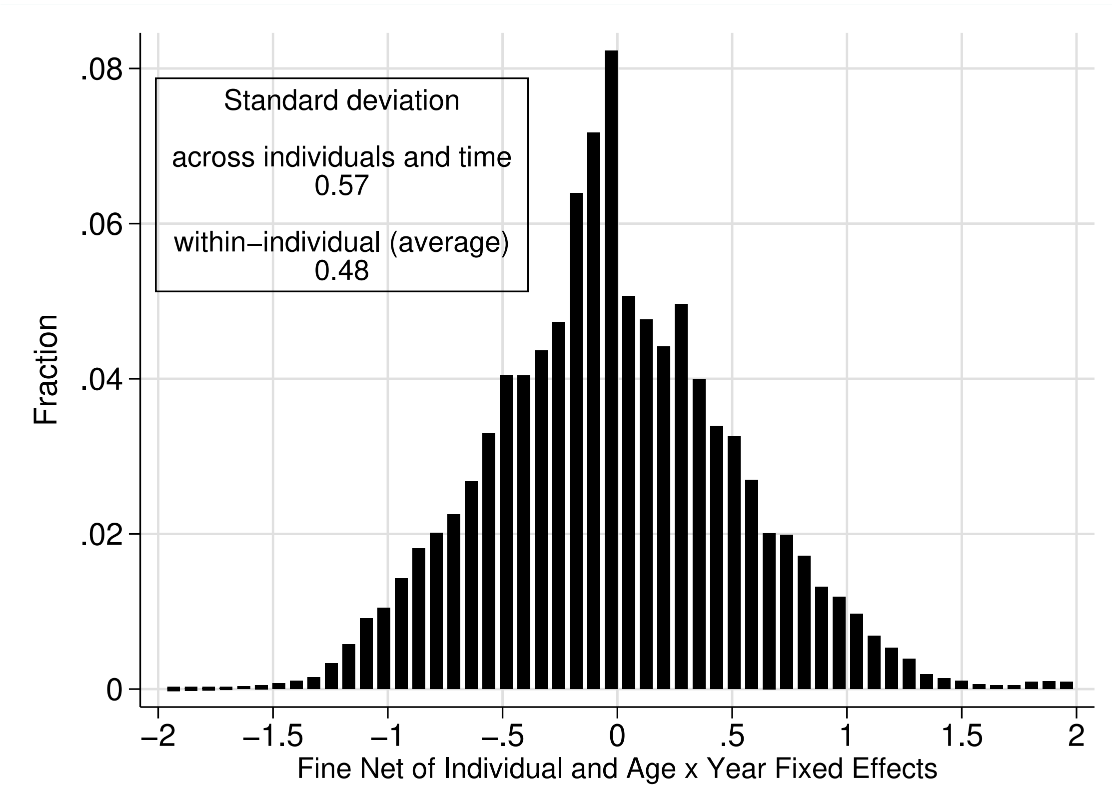
Table A.9: Descriptive Statistics, Female Household Heads Ages 15 to 45

	Year	1975	1985	1995	2005	1970-2000
Panel (a).	Birth year	1949.233	1955.216	1958.591	1963.737	1954.390
Basic		(7.477)	(7.464)	(5.368)	(2.443)	(7.983)
	Agricultural *hukou*	0.805	0.810	0.811	0.806	0.807
		(0.396)	(0.392)	(0.391)	(0.395)	(0.394)
	Urban	0.609	0.607	0.599	0.584	0.604
		(0.488)	0.488)	0.490)	0.493)	0.489)
	Education years	2.847	3.936	4.345	5.813	3.765
		(4.253)	(4.616)	(4.735)	(4.686)	(4.592)
Panel (b).	Children	1.603	1.890	2.202	1.996	1.870
Fertility		(1.799)	(1.583)	(1.166)	(1.052)	(1.562)
	Abortions	0.045	0.104	0.184	0.223	0.109
		(0.223)	(0.327)	(0.423)	(0.462)	(0.336)
	Miscarriages	0.040	0.052	0.062	0.065	0.051
		(0.220)	(0.241)	(0.250)	(0.250)	(0.235)
Panel (c).	Fine	0.000	0.760	1.564	2.266	0.835
Policy	(τ_{ia})	(0.000)	(0.462)	(1.290)	(1.059)	(0.963)
	Tax rate	0.000	0.096	0.342	0.651	0.136
	(κ_{ia})	(0.000)	(0.139)	(0.561)	(0.620)	(0.315)
	Tax tate span	0.000	10.893	5.760	4.009	6.516
	(L_{ia})	(0.000)	(3.782)	(4.532	(3.283)	(5.786)
	Exempted	1.000	0.184	0.266	0.261	0.424
	($\tau_{ia} = 0$)	(0.000)	(0.387)	(0.442	(0.439)	(0.494)
Panel (d).	Works	0.837	0.887	0.900	0.875	0.871
Employment		(0.369)	(0.316)	(0.300)	(0.331)	(0.335)
and	Income	55.937	114.305	136.687	144.896	102.002
Income		(53.354)	(118.358)	(133.458)	(126.233)	(110.978)
(2016 USD)	Husband's Income	46.853	139.128	152.335	183.085	117.087
		(42.707)	(199.405)	(128.010)	(134.143)	(166.579)
	Household Income	104.005	258.030	328.920	330.583	227.162
	(y_{ia})	(85.242)	(257.727)	(316.276)	(214.619)	(245.300)
Panel (e).	Formal Health Care	0.988	0.990	0.992	0.992	0.990
Community	Facility Available	(0.108)	(0.098)	(0.090)	(0.090)	(0.099)
Panel (f).	Data available	0.454	0.454	0.450	0.443	0.454
Ultrasound		(0.498)	(0.498)	(0.498)	(0.497)	(0.498)
	Technology available	0.419	0.761	1.000	1.000	0.713
		(0.494)	(0.427)	(0.000)	(0.000)	(0.453)
	Observations	5,672	7,117	5,411	2,456	185,118

Note: This table presents the mean and standard deviation (in parentheses) for the variables in my empirical analysis. The units of τ_{ia} are years of average household labor income. My main analysis is between the years 1970 and 2000. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), the sources cited in Section 2 for the data on ultrasound technology availability, and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

Variation in the Fine. Figure A.2 describes the identifying variation in the fine (τ_{ia}) by displaying its distribution net of individual and age $\times$ year fixed effects.

Figure A.2: Fine to Have a Child Additional to the First, Net of Individual and Age $\times$ Year Fixed Effects



Note: This figure displays the distribution of the fine, τ_{ia}, described in Section 2 net of individual and age $\times$ year fixed effects. **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

Measuring Actual Household Labor Income. To construct household labor income, I extract the individual-level work histories from the 2014 follow-up of the China Health and Retirement Longitudinal Study and follow these steps for each individual:

1. For each of the 13 possible listed jobs, I calculate an initial, middle, and final date for each job using the answers in the questionnaire.
2. I form a job-specific labor income profile by linearly interpolating the observed initial, middle, and final compensation for each job.
3. If any of the variables within the profile is missing after interpolation (e.g., when the initial compensation is missing), I impute the within-job mean of the available values.
4. If there is no within-job mean because all three compensations for a job are missing, I impute the individual-level mean across all jobs.
5. I sum the woman and male households heads' incomes and convert the values to 2016 USD, by using the 2000 Chinese price index and the (2016) exchange rate between the yuan and the dollar.
6. If the household labor income is still missing, I impute it using the province- and year-specific gross domestic product by agricultural status. I calculate this using statistical yearbooks available in China Data Center (2017).

When presenting results using this construction, the standard errors are bootstrapped and clustered as indicated when the result is presented. Each bootstrap procedure repeats the steps above to account for the sampling error that comes from constructing the measure.

Individuals who worked in collective agricultural systems had, roughly, the following payment system: (i) For each day of work, they received a number of points; (ii) The points accumulated during the year; (iii) At the end of the harvest year, the government procured the grains and compensated the collective; and (iv) Dividends were split according to the accumulated working points. Throughout the year, the workers were compensated in-kind with food, transportation, medical services, etc. For each job, I observe both the points and their realized monetary value. I also observe the monetary value individuals assigned to in-kind compensation. I sum both the monetary value of the points and the in-kind compensation. Recall that the estimation strategy that I employ when using this variable accounts for measurement error.

Measuring Fines Using Actual Household Labor Income. After I calculate the actual household labor income, I use this to calculate the fine. When the calculation in Ebenstein (2010b) indicates that the province- and year-specific fine is 1 unit, I assign the corresponding individual a unit of her household-specific labor income as a fine. I proceed similarly for other values of the fine.

Accounting for the Economic Environment. Table A.10 lists the full set of measures that I use to construct category-specific (dedicated) factors to account for the economic environment.

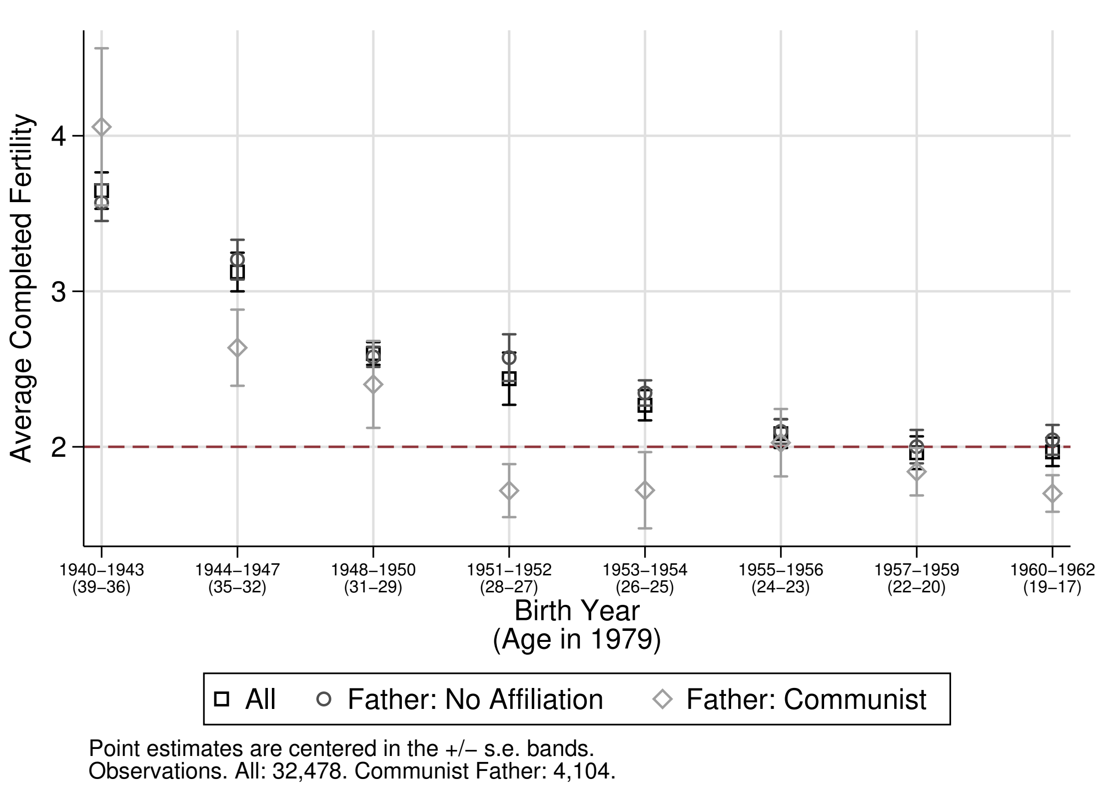
Table A.10: Controls Approximating the Economic Environment

Category (Factor)	**Measures**
Population	Total men, total women, total population, birth rate, death rate, first and second wave of migration permits
Agriculture	Grain crops (kg/hectare), cotton, oil-bearing, agricultural output, agricultural farming output, agricultural forestry output, agricultural husbandry output, agricultural fishery output, agricultural machinery, sown area, grain crops (value), grain tons, fruit tons, large animal heads, aquatic product tons
Infrastructure	Railways, highways, total passengers, railway passengers, highway passengers, freight traffic, freight traffic in railways, freight traffic in highways, civil motor vehicles, post and telecomm value, number of mailed letters
Industry	Industrial enterprises, industrial state enterprises, industrial collectives, industrial output value industrial collective output value, cloth industry, paper industry, cigarette industry, electricity industry, steel industry, steel product industry, cement industry, fertilizer industry
Employment	Employed, employed in urban areas, government staff, state enterprises workers, urban collective workers
Trade	Retail sales, exports net of imports, exports, investment in fixed assets
Government Banking	Local revenue, local revenue through taxes, local expenditure
Economic Organization	Agricultural organization (household or collective), households allowed to rent land for farming and non-farming activities, communal government
Subsidies	Unemployment subsidies, minimum allowance, farming subsidies, reforestation subsidies, pension if age 60 or older or age 80 or older, reformed or non-reformed rural pension scheme
Education	Higher education institutions, elementary schools, higher education institution teachers, middle school teachers, elementary school teachers
Economy	GDP, industry, primary industry, secondary industry, construction, tertiary industry, transportation, GDP per capita, yearly female and male labor income

Note: This table lists the categories of factors to account for the economic environment, using the measures listed for each category. I obtain these measures from China Data Center (2017). **Source:** Author's creation.

Fertility by Father's Affiliation to the Communist Party. Figure A.3 displays completed or total fertility by father's political affiliation using data from Institute of Social Science Survey (2010).

Figure A.3: Average Completed Fertility by Father's Affiliation to the Communist Party



Note: This figure displays the average completed or total fertility by calendar year and father's affiliation to the Chinese Communist Party. **Source:** Author's calculations, created with data from the China Family Panel Studies (Institute of Social Science Survey, 2010).

Balance in Sample Construction. The sample satisfying my age and cohort requirements in the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017) contains $8,919$ women. After restricting the sample to women for whom I observe all of the variables for my analysis, I obtain a final sample of $8,023$ women. I estimate an analogous block to the first block of Panel (a) in Table 1 using weights representing the inverse probability of belonging to the final sample based on urban-rural location and age—observed for the $8,919$ women in the initial sample. I do this by following Horvitz and Thompson (1952). Table A.11 displays the details of this estimation and shows that weighting has little impact on the estimate that I present in the main paper.

Table A.11: Baseline Effects Using Inverse Probability Weighting

	Baseline		IPW	
Leads	-0.004		-0.003	
	(0.011)		(0.011)	
Lags	-0.086	-0.101	-0.085	-0.101
	(0.015)	(0.014)	(0.015)	(0.014)
N	8,023	8,023	8,023	8,023
R^2	0.879	0.879	0.870	0.869

Note: This table presents the first block of Panel (a) in Table 1 (Baseline) and an analogous block using inverse probability weighting (IPW). For the latter case, weights represent the inverse probability of belonging to the final sample based on urban-rural location and age—observed for the $8,919$ women in the initial sample. I do this by following Horvitz and Thompson (1952). **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

Balance in Samples with and without Ultrasound Availability Data. Information on ultrasound availability at the community level is not available for all women in the sample. I estimate an analogous set of estimates as those in Table 4 using the samples with and without data on ultrasound availability. The estimates across the two samples are very similar, reducing concerns that the sex differences in the policy effects are based on unobserved differences in the communities for which ultrasound data are available.

Table A.12: Baseline Effect on Girls and Boys

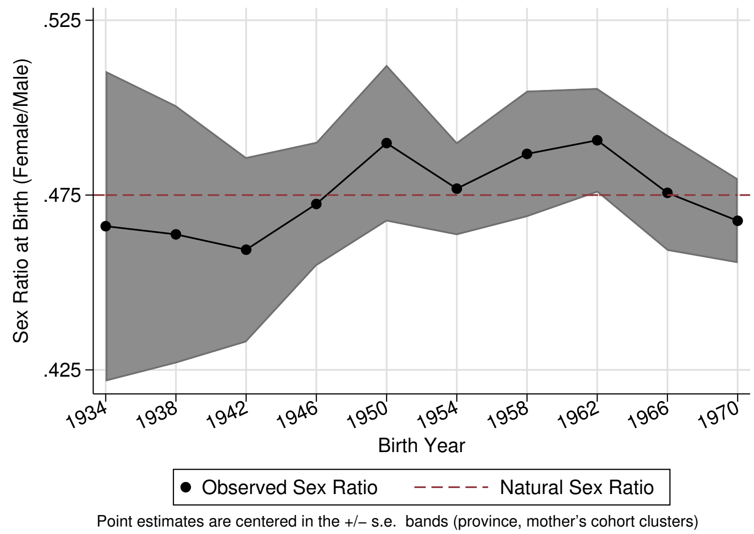
	Girls		Boys	
	Ultrasound Data Available			
Leads	-0.009		0.011	
	(0.010)		(0.009)	
Lags	-0.070	-0.072	0.003	-0.000
	(0.014)	(0.014)	(0.013)	(0.012)
N	3,457	3,457	3,557	3,557
R^2	0.862	0.862	0.863	0.863
	Ultrasound Data Not Available			
Leads	-0.007		0.007	
	0.009		0.008	
Lags	-0.065	-0.077	0.016	0.005
	(0.016)	(0.016)	(0.012)	(0.012)
N	4,203	4,203	4,342	4,342
R^2	0.846	0.846	0.863	0.862

Note: This table presents estimates analogous to those in Table 4 for girls and boys for the subsamples for which data on ultrasound availability is available (top) and not available (bottom). **Source:** Author's calculations, created by linking information from Scharping (2003), Ebenstein (2010b), and the China Health and Retirement Longitudinal Study (National School of Development, 2009-2017).

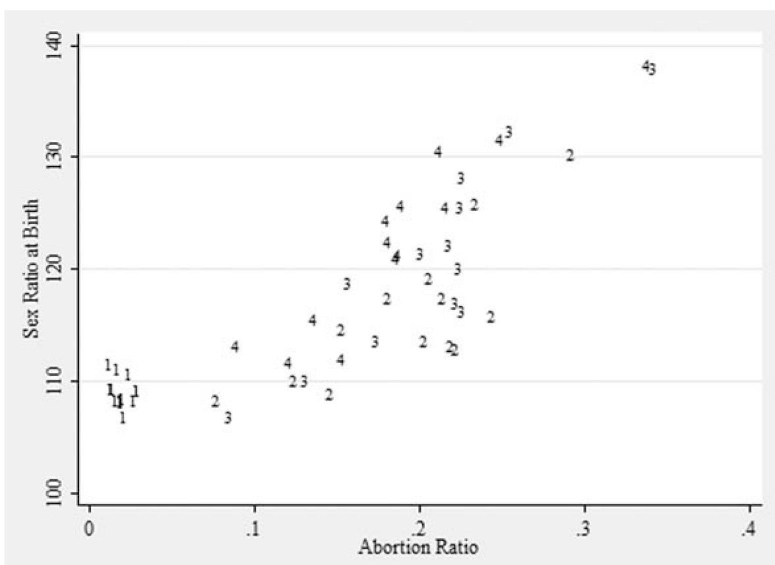
Sex Balance at First Birth. Figure A.4 documents sex balance at birth, both in the sample that I analyze and in the sample analyzed in Chen, Li, and Meng (2013). The natural sex ratio is biased towards males, approximately $\frac{\text{females}}{\text{males}} = 0.475$ or 108 males per 100 females.

Figure A.4: Sex Balance at First Birth

(a) This Paper



(b) Chen, Li, and Meng (2013)



Note: This figure displays the relationship between the sex ratio at birth and the abortion ratio. Sex ratio at birth is defined as the number of male births per 100 female births. Abortion ratio is defined as the proportion of pregnancies ending in abortion. The data are aggregated across pregnancy years (1978-90) by pregnancy order cells—1, 2, and 3 denote 1st, 2nd and 3rd pregnancies; 4 indicates 4th and above. **Source:** Reproduced exactly from Chen, Li, and Meng (2013).

10 Bibliography

Almond, Douglas, Hongbin Li, and Shuang Zhang. 2013. "Land Reform and Sex Selection in China." NBER WP 19153, National Bureau of Economic Research.

Arnold, Fred and Liu Zhaoxiang. 1986. "Sex Preference, Fertility, and Family Planning in China." *Population and Development Review* :221–246.

Baochang, Gu, Wang Feng, Guo Zhigang, and Zhang Erli. 2007. "China's Local and National Fertility Policies at the End of the Twentieth Century." *Population and Development Review* 33 (1):129–148.

Becker, Gary S and H Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children." *Journal of Political Economy* 81 (2, Part 2):S279–S288.

Cai, Yongshun. 2003. "Collective Ownership or Cadres' Ownership? The Non-Agricultural Use of Farmland in China." *China Quarterly* 175:662–680.

Carolina Population Center and the National Institute for Nutrition and Health. 2009. "China Health and Nutrition Survey (CHNS)." Website, http://www.cpc.unc.edu/projects/china. URL `http://www.cpc.unc.edu/projects/china`.

Chan, Kam Wing. 2001. "Recent Migration in China: Patterns, Trends, and Policies." *Asian Perspective* :127–155.

Chen, Yuyu, Hongbin Li, and Lingsheng Meng. 2013. "Prenatal Sex Selection and Missing Girls in China: Evidence from the Diffusion of Diagnostic Ultrasound." *Journal of Human Resources* 48 (1):36–70.

China Data Center. 2017. "China Statistical Yearbooks Database." Website, http://chinadataonline.org/asp/contact.asp. URL `http://chinadataonline.org/asp/contact.asp`.

Clark, Shelley. 2000. "Son Preference and Sex Composition of Children: Evidence from India." *Demography* 37 (1):95–108.

De Silva, Tiloka and Sylvana Tenreyro. 2017. "Population Control Policies and Fertility Convergence." *Journal of Economic Perspectives* 31 (4):205–228.

Ebenstein, Avraham. 2010a. "Fine Rates on the One-Child Policy." Website, https://scholars.huji.ac.il/avrahamebenstein/links. URL `https://scholars.huji.ac.il/avrahamebenstein/links`.

———. 2010b. "The "Missing Girls" of China and the Unintended Consequences of the One-Child Policy." *Journal of Human Resources* 45 (1):87–115.

Fan, Jie, Thomas Heberer, and Wolfgang Taubmann. 2015. *Rural China: Economic and Social Change in the Late Twentieth Century: Economic and Social Change in the Late Twentieth Century*. Routledge.

Feng, Wang and Yong Cai. 2010. "Did China's One-child Policy Prevent 400 Million Births in the Last 30 Years?" *Zhongguo Gaige* 7:85–88.

Fong, Vanessa L. 2004. *Only Hope: Coming of Age under China's One-Child Policy*. Stanford University Press.

Fuest, Clemens, Andreas Peichl, and Sebastian Siegloch. 2018. "Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany." *American Economic Review* 108 (2):393–418.

Galor, Oded. 2012. "The Demographic Transition: Causes and Consequences." *Cliometrica* 6 (1):1–28.

Gaudin, Sylvestre. 2011. "Son Preference in Indian Families: Absolute Versus Relative Wealth Effects." *Demography* 48 (1).

Gracie, Carrie. 2015. "China to End One-Child Policy and Allow Two." Website, http://www.bbc.com/news/world-asia-34665539. URL `http://www.bbc.com/news/world-asia-34665539`.

Greenhalgh, Susan, Zhu Chuzhu, and Li Nan. 1994. "Restraining Population Growth in Three Chinese Villages, 1988-93." *Population and Development Review* :365–395.

Greenhalgh, Susan and Edwin A Winckler. 2005. *Governing China's Population: from Leninist to Neoliberal Biopolitics*. Stanford University Press.

Hardee-Cleaveland, Karen and Judith Banister. 1988. "Fertility Policy and Implementation in China, 1986-88." *Population and Development Review* :245–286.

Heckman, James J. and Richard Robb. 1985. "Using Longitudinal Data to Estimate Age, Period and Cohort Effects in Earnings Equations." In *Cohort Analysis in Social Research: Beyond the Identification Problem*, edited by William M. Mason and Stephen E. Fienberg. New York: Springer-Verlag, 137–150.

Hesketh, Therese, Li Lu, and Zhu Wei Xing. 2005. "The Effect of China's One-Child Family Policy after 25 Years." *New England Journal of Medicine* 353 (11):1171–1176. PMID: 16162890.

Horvitz, D. G. and D. J. Thompson. 1952. "A Generalization of Sampling Without Replacement from a Finite Universe." *Journal of the American Statistical Association* 47 (260):663–685.

Hotz, V Joseph, Jacob Alex Klerman, and Robert J Willis. 1997. "The Economics of Fertility in Developed Countries." *Handbook of Population and Family Economics* 1:275–347.

Huang, Wei. 2016. "Fertility Restrictions and Life Cycle Outcomes: Evidence from the One-Child Policy in China." Unpublished Manuscript, Harvard University.

Huang, Wei, Xiaoyan Lei, and Yaohui Zhao. 2013. "One-Child Policy and the Rise of Man-Made Twins." *Review of Economics and Statistics* 98 (3).

Institute of Social Science Survey. 2010. "China Family Panel Studies (CFPS)." Website, http://www.isss.edu.cn/cfps/EN/. URL `http://www.isss.edu.cn/cfps/EN/`.

Keyi, Sheng. 2015. "Still No Dignity for Chinese Women." Website, http://www.nytimes.com/2015/11/10/opinion/. URL `http://www.nytimes.com/2015/11/10/opinion/`.

Li Jinfeng. 1992. *Jihua Shengyu Guanli*. The Administration of Birth Planning.

Lu, Xiaobo and Elizabeth J Perry. 1997. *Danwei: The Changing Chinese Workplace in Historical and Comparative Perspective*. Me Sharpe.

McElroy, Marjorie and Dennis Tao Yang. 2000. "Carrots and Sticks: Fertility Effects of China's Population Policies." *American Economic Review* 90 (2):389–392.

Meng, Xin, Nancy Qian, and Pierre Yared. 2015. "The Institutional Causes of China's Great Famine, 1959–1961." *Review of Economic Studies* 82 (4):1568–1611.

Minnesota Population Center. 2017. "Integrated Public Use Microdata Series, International Varsion 6.5 [dataset]." Website, http://doi.org/10.18128/D020.V6.5. URL `http://doi.org/10.18128/D020.V6.5`.

National School of Development. 2009-2017. "China Health and Retirement Longitudinal Study (CHARLS)." Website, http://charls.pku.edu.cn/en. URL `http://charls.pku.edu.cn/en`.

Naughton, Barry. 2007. *The Chinese Economy: Transitions and Growth*. MIT Press.

Postiglione, Gerard A. 2015. *Education and Social Change in China: Inequality in a Market Economy*. Routledge.

Powell, Tabtiha M. 2012. *The Negative Impact of the One-Child Policy on the Chinese Society as it Related to Parental Support and the Aging of the Population*. Georgetown University, Thesis.

Qian, Nancy. 2008. "Missing Women and the Price of Tea in China: The Effect of Sex-specific Earnings on Sex Imbalance." *The Quarterly Journal of Economics* 123 (3):1251–1285.

———. 2009. "Quantity-Quality and the One-Child Policy: The Only-Child Disadvantage in School Enrollment in Rural China." NBER WP 14973, National Bureau of Economic Research.

Scharping, Thomas. 2003. *Birth Control in China 1949-2000: Population policy and Demographic Development*. London, Routledge.

Sen, Amartya. 1991. "More than 100 Million Women are Missing." Website, http://www.nybooks.com/articles/1990/12/20/more-than-100-million-women-are-missing/. URL `http://www.nytimes.com/2015/11/02/opinion/`.

———. 2015. "Women's Progress Outdid China's One-Child-Policy." Website, http://www.nytimes.com/2015/11/02/opinion/. URL `http://www.nytimes.com/2015/11/02/opinion/`.

Suárez Serrato, Juan Carlos and Owen Zidar. 2016. "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms." *American Economic Review* 106 (9):2582–2624.

Sun, Xin, Travis J Warner, Dali L Yang, and Mingxing Liu. 2013. "Patterns of Authority and Governance in Rural China: Who's in Charge? Why?" *Journal of Contemporary China* 22 (83):733–754.

Tang, Wenfang and William L Parish. 2000. *Chinese Urban Life under Reform: The Changing Social Contract*. Cambridge University Press.

The Economist. 2014. "Curbing Climate Change: The Deepest Cuts." Website, https://www.economist.com/news/briefing/. URL `https://www.economist.com/news/briefing/`.

The World Bank. 2017. "Fertility Rate, Total (Birth per Woman)." Website, http://data.worldbank.org/indicator/. URL `http://data.worldbank.org/indicator/`.

Tian, Bingxing. 2003. "Zhongguo Diyi Zhengjian (China's No. 1 Document)." *Huizhou, China: Guangdong Renmin Chubanshe* .

White, Tyrene. 2006. *China's Longest Campaign: Birth Planning in the People's Republic, 1949-2005*. Cornell University Press.

Whyte, Martin King, Wang Feng, and Yong Cai. 2015. "Challenging Myths about China's One-Child Policy." *China Journal* (74):144–159.

Whyte, Martin King and William L Parish. 1985. *Urban Life in Contemporary China*. University of Chicago Press.

Xu, Dixin. 1984. "China's Population Policy and Family Planning Work." *Population and Economy* 2:8–9.

Zhang, Junsen. 2017. "The Evolution of China's One-Child Policy and its Effects on Family Outcomes." *Journal of Economic Perspectives* 31 (1):141–159.

Zhao, Yaohui, John Strauss, Albert Park, and Yan Sun. 2009. "China Health and Retirement Longitudinal Study: Pilot, User's Guide." *National School of Development, Peking University* .

Zhu, Jonathan JH and Enhai Wang. 2005. "Diffusion, Use, and Effect of the Internet in China." *Communications of the ACM* 48 (4):49–53.