DISCUSSION PAPER SERIES

DP17973

SPILLOVERS IN CHILDBEARING DECISIONS AND FERTILITY TRANSITIONS: EVIDENCE FROM CHINA

Pauline Rossi and Yun Xiao

DEVELOPMENT ECONOMICS

ISSN 0265-8003

SPILLOVERS IN CHILDBEARING DECISIONS AND FERTILITY TRANSITIONS: EVIDENCE FROM CHINA

Pauline Rossi and Yun Xiao

Discussion Paper DP17973
Published 10 March 2023
Submitted 22 February 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Development Economics

SPILLOVERS IN CHILDBEARING DECISIONS AND FERTILITY TRANSITIONS: EVIDENCE FROM CHINA

Abstract

This article uses China's family planning policies to quantify and explain spillovers in fertility decisions. We test whether ethnic minorities decreased their fertility in response to the policies, although only the majority ethnic group, the Han Chinese, were subject to birth quotas. We exploit the policy rollout and variation in pre-policy age-specific fertility levels to construct a measure of the negative shock to Han fertility. Combining this measure with variation in the local share of Han, we estimate that a woman gives birth to 0.63 fewer children if the average completed fertility among her peers is exogenously reduced by one child. The fertility response of minorities is driven by cultural proximity with the Han and by higher educational investments, suggesting that spillovers operate through both social and economic channels. These results provide evidence that social multipliers can accelerate fertility transitions.

JEL Classification: C36, D1, J11, J13, O15, O53

Keywords: Fertility, Family planning, China, Spillovers, Peer effects

Pauline Rossi - pauline.rossi@ensae.fr
CREST, Ecole Polytechnique and CEPR

Yun Xiao - yun.xiao@economics.gu.se
University of Gothenburg

Spillovers in Childbearing Decisions and Fertility Transitions: Evidence from China

Pauline ROSSI and Yun XIAO *

October 21, 2022

Abstract

This article uses China's family planning policies to quantify and explain spillovers in fertility decisions. We test whether ethnic minorities decreased their fertility in response to the policies, although only the majority ethnic group, the Han Chinese, were subject to birth quotas. We exploit the policy rollout and variation in pre-policy age-specific fertility levels to construct a measure of the negative shock to Han fertility. Combining this measure with variation in the local share of Han, we estimate that a woman gives birth to 0.63 fewer children if the average completed fertility among her peers is exogenously reduced by one child. The fertility response of minorities is driven by cultural proximity with the Han and by higher educational investments, suggesting that spillovers operate through both social and economic channels. These results provide evidence that social multipliers can accelerate fertility transitions.

Keywords: Fertility, Family planning, China, Spillovers, Peer Effects, Partial population experiment.

JEL Codes: C36, D1, J11, J13, O15, O53.

*Rossi (corresponding author): CREST, Ecole Polytechnique, 5 avenue Henry Le Chatelier 91120 Palaiseau France, pauline.rossi@ensae.fr. Xiao: University of Gothenburg. We thank Romain Wacziarg and two referees for their careful reading of the mansucript and their constructive comments. We are grateful to Hessel Oosterbeek for his numerous suggestions. For their helpful comments, we thank Jason Abaluck, Joseph Altonji, Jere Behrman, Hanming Fang, Alessandro Iaria, Maulik Jagnani, Wendy Janssens, Robert Jensen, Lei Li, Li Li, Kaivan Munshi, Jan van Ours, Mark Rosenzweig, Martin Rotemberg, Yanos Zylberberg, as well as participants to seminars at Bristol, UvA, Crest, Yale, ECNU, Columbia, ShanghaiTech, Geneva, CU Denver, College de France, Max-Planck Institute, TSE, WE_ARE webinar, Namur, Alicante, Frisch Center and Pennsylvania, and several conferences: International Symposium on Contemporary Labor Economics, EALE, EEA, ES, NEUDC, and ASSA.

1 Introduction

There is a long-standing debate on the fundamental causes of fertility transitions among social scientists (Lee, 2015). One prominent view is that fertility decline is related to economic development through changes in structural factors affecting the costs and benefits of having children (Notestein, 1953; Becker, 1991). This view has been challenged by scholars who argue that, historically, fertility decline starts at very different stages of development (Coale and Watkins, 1986; Bongaarts and Watkins, 1996). They propose that we think of small families as a cultural innovation that is gradually transmitted through social interactions. There are theoretical attempts to reconcile both views by combining the rational choice approach with social factors (Kohler, 2000; Durlauf and Walker, 2001; Munshi and Myaux, 2006). The key insight from these models is that the interplay between incentives and norms is crucial for understanding the differences in historical transitions. Economic pre-conditions may be amplified or undermined by social interactions, so that the same change in one structural factor may trigger fertility decline in one social context and not in the other. Understanding when and why this happens is an important question that we address in the context of China.

China's family planning policies provide a unique setting to study to what extent and through which mechanisms couples' fertility decisions are influenced by other couples. Birth quotas were introduced in the 1970s with the "Later, Longer, Fewer" campaign (LLF hereafter), followed in 1979 by the stricter One-Child Policy (OCP hereafter). In contrast to the Han people, ethnic minority groups were exempt due to their small share in the total population and political considerations (Scharping, 2013, p.150). These exemptions create an interesting empirical design: whether minorities respond to family planning policies and whether this response depends on the share of Han in the population is informative for peer effects.

Our main dataset is the 1% random sample of the 1990 Population Census. It provides information on the completed fertility of a representative sample of women born in 1926–1945. These women were already married and were in their reproductive years during the 70s. Our sample consists of over 725,000 Han women and 55,000 minority women, which allows us to have very precise estimates and to run several robustness checks. Moreover, we are able to explore heterogeneity across different ethnic groups to shed light on potential mechanisms.

We start with a reduced-form approach to estimate the causal impact of the policies. We implement a difference-in-differences strategy, exploiting the fact that, among Han women, exposure to the policy varies by cohort and by province. We create a measure of policy exposure, capturing the reduction in fertility required to comply with the policies. This measure assumes that Han women would otherwise experience the age-specific fertility rates as of 1969, just before the LLF campaign could effectively prevent births. The measure captures the maximum reduction that we can expect for a given cohort in a given province if the policy were perfectly implemented and fertility rates had remained constant in the absence of the policy. It varies across provinces for two reasons: first, the campaign was launched at different points in time between 1969 and 1975; and second, initial fertility levels were different. Our identification strategy hinges more on the first source of variation than on the second.

We first show that birth quotas indeed translated into a reduction in Han fertility. Then, we find that the higher the policy exposure of Han peers, the higher the decrease in the fertility of an ethnic minority woman. This response is stronger when the local Han share is higher and is absent in places where Han do not live. We conduct several robustness tests suggesting that minorities react to a change in Han behavior and not to the birth quotas per se. The results are robust to alternative measures of policy exposure and to heterogeneous policy effects. Moreover, we argue that they are not confounded by spurious trends nor by concomitant shocks using placebo tests and event studies.

We further explore the reasons why spillovers arise. We are able to test two channels: a social one, emphasizing the concept of conformism; and an economic one, operating through the quality-quantity trade-off – when parents have fewer children they spend more resources per capita, and this may have general equilibrium effects. We formulate the necessary conditions for each channel to work and discuss when these conditions are more likely to hold. We take advantage of the fact that different ethnic groups are more or less integrated into the Han society. Our hypothesis is that the social channel should be stronger for groups culturally closer to the Han, and that the economic channel should be stronger for groups competing with the Han in the labor market. We find evidence supporting both channels. We consider and rule out alternative explanations related to changes in female opportunity cost of time and social learning about birth control.

Finally, we argue that the reduced-form estimates have a structural interpretation. Using our setting as a partial-population experiment, we are able to identify the endogenous peer effect parameter in a standard linear-in-means empirical model (Moffitt, 2001; Lalive and Cattaneo, 2009; Bobonis and Finan, 2009; Brown and Laschever, 2012). In contrast, research designs based on the (quasi-)random allocation of peers cannot separately identify *endogenous* and *exogenous* peer effects, i.e. whether individuals are influenced by the behavior or by the characteristics of their peers. Separating both effects is crucial to understand fertility transitions since only the endogenous component gives rise to social multipliers. Our idea is to use the policy-driven change in Han fertility combined with the share of Han in the reference group as an instrument for the group average fertility. We define the reference group as all women living in the same prefecture, holding the same *hukou*, and belonging to the same age cohort. We find that a woman gives birth to 0.64 fewer children if the average completed fertility among her peers is reduced by one child. The structure of the model allows us to decompose the total policy effect into a direct effect and an indirect effect generated by spillovers. Our estimates suggest that the indirect component accounts for roughly two thirds of the total effect.

Our paper relates to the empirical literature on the effects of China's population policies on fertility and family outcomes. Most studies focus on the OCP (Li et al., 2005, 2011; Ebenstein, 2010; Huang and Zhou, 2016; Huang et al., 2016; Li and Zhang, 2017; Liu, 2014; Qian, 2009; Wang and Zhang, 2018).[1] They find effects on fertility, but the magnitude is not so large because fertility was already low when the OCP was implemented in the 1980s. In fact, a major part of China's fertility decline happened in the 1970s when the LLF campaign was launched. Although historians and demographers have long argued that the Chinese fertility transition was triggered by the LLF campaign (Scharping, 2013, p.312), economists have only recently started looking into this period (Chen and Huang, 2020; Chen and Fang, 2021; Babiarz et al., 2018). We use the same identification strategy as Chen and Fang (2021) and find consistent results: we estimate that the LLF campaign led to a reduction in the completed fertility of Han couples by 0.8 births, accounting for 60% of the decline between the 1926 and 1945 cohorts. Our contribution is to show that population policies also have sizable spillover effects on the minority population. Spillovers are generally ignored in the existing literature. Some studies, for instance Li et al. (2005), use ethnic minorities as a control group, assuming that they were not affected by population policies. Our results imply that these analyses underestimate the causal policy effect.

Our paper also contributes to the smaller literature providing causal evidence of peer effects in fertility decisions. Most articles study the timing of births and do not deal with fertility transitions or social multipliers.[2] One recent paper by Spolaore and Wacziarg (2021) studies the

[1] See Zhang (2017) and Wang et al. (2017) for extensive reviews of this literature.

[2] For instance, Lyngstad and Prskawetz (2010), Ciliberto et al. (2016), Hensvik and Nilsson (2010) investigate whether childbirths among peers (neighbors, colleagues) influence women's likelihood of becoming pregnant in low fertility societies.

diffusion of small families in the 19th century in Europe from a macro perspective. They show that the decline in fertility, starting in France, propagated first to regions that were closer to France in terms of language and then to the rest of the continent. We come to the same conclusion in the case of China: the linguistic distance to the innovator (here, the Han people) explains the adoption of new fertility behaviors. Li and Zhang (2009) is the only paper estimating peer effects in fertility decisions in China. They exploit the exemptions to ethnic minorities during the OCP and get a point estimate of the peer effect parameter ranging between 0.5 and 0.9 across different specifications. Compared to the LLF campaign, the design has three drawbacks. First, the OCP was introduced in 1979 in all provinces, so they cannot exploit any rollout. Their identification relies on spatial variation in the policy implementation, proxied by the amount of fines imposed on unauthorized births. As pointed out by Zhang (2017), the issue is that fines may reflect local fertility demand. Second, the OCP took place at a time of structural transformation and strong economic growth. Third, exemptions to minority groups were gradually removed after 1984. Our contribution is threefold: (i) we find a stable and precise estimate of 0.64 using a larger sample and a time period with more exogenous variation; (ii) we quantify the role of spillovers in the Chinese fertility transition; and (iii) we dig into the mechanisms and conclude that the diffusion of small families results not only from social interactions but also from market interactions through the equilibrium price of a child.

Our results are important for understanding the past and predicting the future. First, they confirm the general view that population policies played a key role in the Chinese fertility decline. Moreover, they explain how these policies could lead to a sizable behavioral change: the policy impact was magnified by a social multiplier. Second, our framework provides guidance to think about new policies. In 2016, faced with the challenges posed by an aging population, the Chinese government officially allowed all couples to have two children. However, the expected baby boom did not happen. One explanation consistent with our findings is that China moved from a high fertility equilibrium to a low fertility equilibrium in response to the introduction of birth quotas, and is now in a low fertility trap. The way out is to actively subsidize second births among a subgroup that is important enough to trigger the switch to a two-child equilibrium.

The rest of the article is organized as follows. The next section summarizes the historical context. Section 3 describes the data, key variables, and descriptive statistics. In Section 4, we present the reduced-form results and we discuss the mechanisms in Section 5. Section 6 explains the semi-structural approach and Section 7 concludes with policy implications.

2 Context

2.1 Family planning policies in China

In the early 1970s, the "Later, Longer, and Fewer" (LLF, "*wan, xi, shao*" in Chinese) campaign was launched to promote later marriage, longer birth intervals, and fewer children.[3] In 1971, the State Council released the document [71]51, "Report on Implementing Birth Control Policies" ("*guanyu zuohao jihua shengyu gongzuo de baogao*"), setting targets for population growth rates and instructions on implementing family planning policies. These policies required women to not marry before age 23 and men before 25, instructed couples with one child to wait for at least three years to have another child, and said couples should have no more than two children. In practice, not all provinces started implementing the policies in the same year, and we exploit the rollout between 1969 and 1975 for identification. Following Chen and Fang (2021), we assume that the LLF campaign was launched in a province in the year when the provincial leading group in charge of organizing the implementation was formed.[4] The variation is shown on the map in Figure 1a. In 1979, birth control became stricter everywhere with the introduction of the OCP, under which couples were only allowed to have one child. The policy was relaxed in 1981 to allow rural couples to have a second birth if the firstborn was a daughter. Several exemptions were later introduced province by province, but we do not exploit such variations because they may reflect the local demand for children.

[Insert Figure 1 here]

During the LLF campaign, local organizations were in charge of birth control at the local level. These organizations were enterprises and neighborhood committees in cities, special cadres of the party, and barefoot doctors in villages. Each local authority had its own way of making sure that people complied with the birth limit. Some places required couples to ask for a third-child permit and only granted it if the couple was eligible for an exemption. Others imposed forced abortions, sterilizations, and IUD insertions on the women who were not eligible to have another child. According to Scharping (2013, p.312), "the crucial mechanism for effecting the drastic fertility decline still seems to have been the penetration of the state power into almost

[3]China's history of family planning can be dated back to 1954 when the first family planning campaign was launched in some provinces. A second urban-oriented family planning campaign was initiated in 1962. But these campaigns were short-lived and had limited influence (Scharping, 2013, p.46-49).

[4]Instead, Babiarz et al. (2018) use the first mention of birth planning regulation in provincial public health archives (*Weishengzhi*) as the starting year of the LLF campaign in each province. Our results remain the same if we follow Babiarz et al. (2018) instead of Chen and Fang (2021).

every aspect of life [...] cadres still yielded almost limitless power for denying means of subsistence to anyone disobeying their commands." In the 1970s, the state controlled the allocation of jobs and housing in cities and the allocation of land and food in villages, and strongly restricted any migration. In the wake of the OCP, a huge bureaucracy was created to plan, monitor, and evaluate the implementation of birth quotas in a more systematic way. Couples who violated the policy faced fines as high as several years of household income, loss of state-provided employment, and exclusion from free public education for their additional children.

Importantly for our purpose, the LLF campaign targeted only the majority ethnic group, the Han Chinese. Between provinces, the main determinant of the rollout is the size of the Minority population: birth quotas were introduced later in provinces where many Minority Chinese lived.[5] Within provinces, exemptions were granted to all minority ethnic groups until 1984 (Scharping, 2013, p.150). Authorities were afraid that family planning policies would push these groups to break away from the central government. Given their small weight in the total population (less than 8% in the 1970s), it was decided that the benefit of imposing strict control over minority groups was not worth the potential cost. After 1984, birth control was gradually extended to large minority groups. Therefore, we focus on the period between 1969 and 1984 and on cohorts of women of reproductive age during these years. The official guidelines regarding minorities were clear and local cadres were not yet subjected to sanctions if population targets were not met, so they had no incentives to put pressure on exempted groups. Still, historians cannot definitely prove that exemptions were actually granted to all minority couples during the whole period.[6] That is why our empirical strategy allows for a direct effect of the policy on minorities and only exploits the heterogeneous response by local Han share to identify spillovers. We leverage the fact that this type of preferential treatment was granted to all individuals from ethnic minority groups irrespective of their place of residence, unlike place-based development policies that benefited both minority Chinese and Han Chinese in areas dominated by minorities (Zang, 2015, p.71).

[5]When we regress the year of implementation on provincial characteristics including size of the Minority population, size of the Han population, fertility rate, sex ratio, GDP per capita, urbanization rate, the structure of the economy, geographical and cultural distance to Beijing, only the size of the Minority population predicts the timing.

[6]Documenting the precise implementation of these exemptions is difficult partly because social sciences were abolished in China between 1952 and 1979. For instance, there is no study exploring how the LLF campaign was perceived by minority groups at that time.

2.2 Ethnic groups in China

China is a multi-ethnic society with a majority group, the Han Chinese, and 55 officially recognized minority groups. According to the 2010 Population Census, the combined population of minority groups amounts to 113 million and there is a large heterogeneity in size ranging from under 4,000 people for the Tatar to almost 17 million people for the Zhuang. The minority groups are mainly located in the southwest, northwest, and northeast of China, while the Han tend to live in the southeast and central part of China. However, there is some local variation that plays an important role in our identification strategy. Figure 1b displays the spatial distribution of the Han in 1990 at the prefecture level. Of the 347 prefectures, 34 have less than 40% of Han and 43 have between 40% and 80% of Han. Together, these 77 prefectures account for 12% of the total population and for 77% of the minority population. In Figure A.1, we plot the distribution of the Han share in our sample of minority women. We have data points on the whole support, including at zero. Therefore, we rely on observations, not on extrapolations, when we estimate the impact of population policies in places where only minorities live.

Ethnicity is part of the administrative identity. It cannot be changed except under very rare circumstances, which we discuss in Section 6. In particular, we rule out the concern that minority women in our sample were in fact classified as Han in the 1970s and, hence, were directly affected by the family planning policies. Inter-ethnic marriage is limited. In our sample, only 15% of minority women were married to a Han Chinese. We assume that these couples were exempted from birth quotas but our results are not driven by this assumption.[7]

An interesting feature for our purpose is that ethnic groups are distinct from each other in terms of language, residence, and occupation. These differences allow us to predict which groups are more likely to be affected by their Han peers through social and/or economic channels. First, we create an indicator of cultural integration into the Han society based on linguistic distance and residential segregation. Each group has its own oral language but some of them use the Han script while others have their own written language. This distinction reflects the history of incorporation into China and the group members' ability to preserve their own culture. Location choices are also informative: some groups mostly reside in their own autonomous region whereas others are widespread and live closer to Han communities. Second, we build a measure

[7] There is no information on how the LLF policy was implemented for Han-minority couples. Under OCP, Huang and Zhou (2016) report that there were differences by provinces, some exempting these couples from strict birth control and others not. Exemptions were more likely to be granted in provinces with more minorities because the local governments did not want to come across mass resistance and complaints. If we restrict our analysis to these provinces or if we drop Han-minority couples, our results do not change.

of labor market competition using occupational choices. Some groups traditionally specialize in particular professions like cattle breeder or craftsman. Others engage in occupations dominated by the Han that typically require more education (e.g., teacher), so the degree of competition is quite different. Appendix A.1 explains in detail how we measure the linguistic distance and the residential segregation between the Han and a given ethnic group, and how we quantify the level of Han dominance and education requirement for a given occupation. Our data-driven classification, shown in Table 1, is consistent with the sociology literature (see, for example, the differentiation indexes in terms of residence, education, occupation, and industry in Poston and Gu (1987)). We exploit it in Section 5 to explore the channels underlying fertility spillovers.

[Insert Table 1 here]

3 Data

3.1 Sample and fertility data

Our main analysis uses the 1% sample of the 1990 Population Census (hereafter the 1990 Census), which provides the relevant information for the cohorts of interest: age, sex, place of residence, *hukou* status,[8] ethnic identity, and the number of children ever born for women aged 15 to 64. We restrict our sample to women aged 45 to 64, who have reached the end of their reproductive life, to avoid censoring issues.[9] These cohorts, born in the years 1926 to 1945, were between 39 and 58 years old in 1984 and hence too old to be directly affected by the relaxation of minority exemptions. Note that they were between 24 and 43 years old in 1969, which limits the possibility that our findings are driven by the potential effects of the policies on marriage market dynamics. Indeed, among these cohorts, over 90% of women were already married by age 24 and divorces were extremely rare, below 1% (Scharping, 2013, p.239-240). The unit of our empirical analysis is a woman because fertility is observed at the female level; but in the discussion we use the words "woman" and "couple" interchangeably since marriage is universal and stable. Summary statistics for our main samples are shown in Table 2. We observe 57,570 minority women and 726,907 Han women. The vast majority of the minority sample hold a rural *hukou* and belong to one of the 14 largest minority groups.

[8] *Hukou* is an official record that includes, among other characteristics, either a rural status or an urban status. This essential dimension determines the eligibility for state welfare programs. Transfer from urban to rural *hukou* or the other way around is highly restricted (Chan, 2009).

[9] The age-specific fertility rate for women aged 45 and older is extremely low (Coale and Chen, 1987).

[Insert Table 2 here]

In addition, we use the 1% sample of the 1982 census (hereafter the 1982 Census) and the 20% sample of the 2005 One-percent Population Survey (hereafter the 2005 Mini-census) to look at aggregate levels and trends over a long time period.[10] We also use older cohorts surveyed in the 1982 Census to conduct a placebo test and an event study supporting the common trends assumption. Other important sources of data are (i) Coale and Chen (1987) who provide the provincial age-specific fertility rates in 1969 necessary for our measure of policy exposure;[11] and (ii) the provincial socioeconomic characteristics in 1950–1969 from the National Bureau of Statistics of China (2010) that we use to construct some control variables. Note that provincial data before 1969 is missing for Hainan, Chongqing and Tibet so we exclude these provinces. Finally, we exclude migrant women from our sample. They account for only 1.2% of observations because internal migration was under strict control at that time.

Our fertility outcome – completed fertility – is the key indicator in the literature on fertility transitions. Another potential outcome could be birth intervals since the Han-targeting policies put restrictions on spacing as well. However, the census does not provide the birthdates of all children so we cannot estimate spillovers in the timing of births. We also discuss the sex ratio at birth, for which we do not find any evidence of a first stage generated by the LLF campaign. Lastly, one may be concerned about the quality of fertility data in China. High frequency birth statistics started being produced once the OCP bureaucracy was in place, and these numbers were potentially subject to manipulation given the high political stakes. But our analysis relies on census data and focuses on the early period, when monitoring was still loose. According to historians, censuses are the best source of data and neither local officials nor households had strong incentives to underreport births in the 1970s (Scharping, 2013, p.206).

3.2 Measure of policy exposure

Our identification strategy requires isolating the change in Han fertility driven by family planning policies. We construct a measure of policy exposure reflecting the idea that the expected reduction in fertility depends on a woman's age when the LLF and OCP quotas are implemented. Young, childless women are fully exposed whereas older women who already completed their

[10] In the econometric analysis, we cannot use the 2005 Mini-census, which is only a 0.2% sample of the total population, because the number of observations per reference group is too small. The 1982 Census provides fertility data for cohorts 1918-1925. These cohorts had very little exposure to family planning policies and we find the same results if we include them in the analysis.

[11] Coale and Chen (1987) tabulated the rates based on data from the China's One per Thousand Sample Fertility Survey collected by the State Family Planning Commission in 1982.

fertility are not exposed at all. For age groups in-between, one option would be to use a linear function of age to capture their partial exposure. But fertility rates are not constant between 15 and 50 years old. Therefore, following Chen and Fang (2021), we rely on the age-specific fertility rates observed just before the policy in 1969 to get a more precise relationship between age and the intensity of exposure. For each age group, we predict (i) how many additional children would be born in the absence of family planning policies, assuming that a woman would experience the 1969 age-specific fertility rates in her lifetime; and (ii) how many of these births were authorized by the LLF or OCP policy. Our measure of policy exposure is the difference between the two, which we call expected reduction (ER). It indicates how many births would not happen due to the policies if quotas were perfectly enforced and if fertility levels had otherwise remained constant over time.

For a given cohort of Han women, the policy exposure varies across provinces because the implementation of the LLF campaign did not start everywhere in the same year (see Figure 1a). In addition, pre-policy fertility rates differed by province and by *hukou* status. Figure 1c and Figure 1d plot the spatial distribution of the total fertility rates in 1969 in rural and urban areas, respectively. In Appendix A.3, we show that the identification of spillovers does not hinge on this second source of spatial variation: we find the same estimate if we use the national fertility rates for everyone.[12]

Table 3 illustrates how we constructed the expected reduction triggered by the LLF policy (ER LLF) for women in different cohort-province-*hukou* groups. A comparison of columns (1) and (2) indicates how the measure varies across cohorts in the Jiangsu province, where the LLF policy started in 1973. Women born in 1945 were 28 years old; the average woman had experienced age-specific fertility rates between ages 15 and 28, cumulating to 2.3 children already born. She was expected to experience later the rates between ages 29 and 49, cumulating to 2.7 children to be born. None of these additional births would be authorized by the policy because the number of children already born was higher than the policy quota. For this cohort, our measure ER LLF is therefore equal to 2.7. If we consider the 1930 cohort instead, the excepted reduction is almost zero because fertility rates are very low between ages 44 and 49. Columns (3) and (4) show that, for the same cohorts, the measure is different in Xinjiang because (i) the

[12]Our preferred measure of policy exposure uses *hukou*-specific provincial fertility rates to be consistent with our econometric specification, which is a linear model in levels with homogeneous parameters. Since the policy imposes the same quota on all Han Chinese, the impact of the policy in terms of number of children mechanically depends on initial fertility levels. We incorporate this heterogeneity into the exposure variable and interpret the homogeneous treatment effect as the fraction of the expected reduction that is actually implemented. We discuss potential heterogeneous treatment effects in Appendix A.3.

policy started later and (ii) 1969 fertility rates were higher. Lastly, columns (5) and (6) illustrate that within the same province-cohort group, ER LLF differs by *hukou* status since initial fertility levels are usually lower in cities. In the same way, we construct a measure of exposure to OCP (ER OCP). It is noteworthy that, since the vast majority of women in our sample already had two children in 1979, ER OCP is close to zero with little variation and hence contributes very little to identification. We define the total policy exposure ER as the sum of ER LLF and ER OCP. As reported in Table 2, ER is, on average, equal to 1.5 births and ranges from almost zero for the oldest cohort (1926) to 3.4 for the youngest cohort (1945). Appendix A.2 provides more details about the construction of ER; in particular, we discuss (rare) cases in which ER is lower than the number of children to be born because the quota had not been reached.

[Insert Table 3 here]

3.3 Descriptive evidence

Figure 2a shows that the expected reduction in fertility for the Han is reflected in the evolution of actual fertility. The graph plots the average ER and the average completed fertility of Han women cohort by cohort. Vertical red lines indicate which cohorts belong to our main sample; in addition, we use data from the 1982 Census and 2005 Mini-census to extend the time window. Han women born before 1930 gave birth to slightly more than five children. Afterwards, fertility started to decrease, especially quickly between the cohorts of 1930 and 1945, and reaches two children for women born in 1960. The start and speed of the fertility decline coincided with changes in ER, increasing strongly during LLF and more slowly during OCP.

[Insert Figure 2 here]

Although minorities were not targeted by family planning policies, their fertility also decreased during the period. It could be due to spillovers and/or because fertility decline is in fact driven by other forces affecting both Han and minority women. To disentangle these two explanations, we compare minority women living in places with different shares of Han. Figure 2b plots ER against fertility outcomes for minorities, looking separately at prefectures where the share of Han Chinese is above and below 60%. The decline in the fertility of minorities starts earlier and is faster in high Han-share prefectures. As a result, minority women in "high" prefectures have, on average, one fewer child than women in "low" prefectures among cohorts born after 1945. In contrast, for older cohorts born before 1930, fertility was higher in prefectures dominated by Han Chinese.

Another fertility outcome of interest is the gender composition of minorities' offspring. Sex selection has been documented as one of the unintended consequences of China's OCP (Ebenstein, 2010). However, we cannot use our design to investigate spillovers in sex selection decisions because, in line with the recent findings of Chen and Huang (2020), we find no evidence of the LLF campaign having an impact on sex ratios at birth among Han Chinese. Figures A.2a and A.2b plot the share of male births by the mothers' cohort, for Han and minorities, respectively. Among Han women, sex ratios remained constant between the cohorts of 1926 and 1945. It was only after 1945, for cohorts more exposed to OCP, that sex ratios became more imbalanced. Turning to minority women, there is no change in levels and no divergence between prefectures with high and low shares of Han during the period 1926-1945. We conclude that sex selection does not play an important role in our context.

Figures 2a and 2b provide suggestive evidence that family planning policies triggered a fertility decline among Han women that spilled over onto minorities. Next, we formalize the econometric specification and the identification assumptions needed to reach conclusive evidence.

4 Reduced-form approach

4.1 Empirical strategy

To identify the causal effect of family planning policies, we implement a difference-in-differences strategy. Specifically, we consider the following equation:

$$y_i = \rho \text{ER}_{rpc(i)} + \beta X_i + \lambda Z_{rpc(i)} + \eta_{c(i)} + \mu_{rd(i)} + t_{rp(i)} c + \varepsilon_i \tag{1}$$

where y_i is the completed fertility of woman i holding *hukou* r, living in prefecture d within province p, and belonging to cohort c. $\eta_{c(i)}$ is a set of cohort fixed effects capturing changes common to all places and $\mu_{rd(i)}$ is a set of prefecture-*hukou* fixed effects capturing time-invariant characteristics associated with both exposure and fertility. We also include a set of province-*hukou*-specific time trends $t_{rp(i)}c$ to capture any linear trend in outcome-determining variables.[13] X_i is a set of exogenous characteristics (ethnic group and educational level) measured at the indi-

[13] ER does not vary linearly with age so we can exploit deviations from a linear trend for identification. In Appendix A.3, we provide support for the inclusion of provincial linear trends by showing that specifications without trends overestimate the coefficients of interest.

vidual level and $Z_{rpc(i)}$ is a vector of provincial characteristics controlling for potentially different changes in socioeconomic development (log of gross regional product per capita, population density, number of secondary school teachers per capita, number of hospital beds per capita, and number of health workers per capita).[14] Standard errors are clustered at the province-*hukou*-cohort level, at which ER_{rpc} varies.

We use equation 1 to estimate the *average* policy effect on the fertility of Han Chinese and Minority Chinese, separately. ρ captures both the direct effect and the endogenous peer effect. To distinguish between the two, we test whether the response of Minority Chinese varies with the local share of Han (s^H) by estimating the following equation:

$$y_i = \rho_0 \mathrm{ER}_{rpc(i)} + \phi s^H_{rd(i)} \mathrm{ER}_{rpc(i)} + \beta X_i + \lambda Z_{rpc(i)} + \eta_{c(i)} + \mu_{rd(i)} + t_{rp(i)} c + \varepsilon_i \qquad (2)$$

Compared to equation 1, equation 2 has three advantages. First, we can test for the existence of a direct effect of the policies on minorities. ρ_0 captures the policy impact in places without Han women, i.e., where spillovers cannot be at play. If minorities were in fact not exempted from birth quotas, if they were convinced by the propaganda in favor of small families, or if better access to contraceptives increased women's control over births, then ρ_0 should be negative. Second, we can test whether the causal effect is proportional to the share of Han. As we show in section 6, ϕ is informative about the endogenous peer effect parameter in a standard linear-in-means model. If $\phi \neq 0$, this is sufficient to prove the existence of spillovers. Third, we can measure the Han share at the prefecture-*hukou* level, exploiting thus a spatial source of variation that is more local than the province-*hukou* level. In this specification, we cluster standard errors at the prefecture-*hukou*-cohort level, at which $s^H_{rd}\mathrm{ER}_{rpc}$ varies. Note that the main effect of s^H_{rd} is absorbed by the fixed effects μ_{rd}.[15]

Identification does not rely on the assumption that the timing of family planning policies is random. Identification relies on a common trends assumption: if birth quotas had not been implemented, the changes in completed fertility between two cohorts would have been the same for all prefecture-hukou groups, conditional on the covariates. In other words, the identification of ρ fails if, in the absence of the policy, the timing of the decline in fertility would have been correlated with Figure 1a and the magnitude of the decline would have been correlated with

[14]These controls are measured when cohort c turned age 24 (between 1950 and 1969) to capture changes in fertility driven by socioeconomic development before LLF. We allow these effects to differ by pre-policy fertility level, by adding the interaction of provincial characteristics and 1969 fertility rates.

[15]The Han share, measured in 1990, is very stable over time because migration is limited and ethnic identities rarely change.

Figure 1c in rural areas and correlated with Figure 1d in urban areas. For the identification of ϕ to fail, we need to add the condition that, within a province, the decline in the fertility of minorities would have been correlated with Figure 1b. The intuition underlying our identification strategy is that Chinese fertility policies generate a very specific pattern in terms of (i) when the decline should start and (ii) how large the decline should be. This pattern is artificial and hence unlikely to arise endogenously if fertility was not constrained. This is how we can distinguish the impact of the policy from other changes in socio-economic drivers of fertility.

We will provide support for the identification assumption in two ways. First, we discuss potential concomitant shocks like the Great Chinese Famine, the Cultural Revolution, and the Send-down Movement. We show that our estimates are not confounded by these other events. Second, we show that pre-trends are parallel using (i) a placebo test on older cohorts and (ii) an event-study analysis.

4.2 Estimation results

As a preliminary step, we estimate equation 1 on the Han sample to confirm that the policies indeed had a negative effect on their fertility. Results are reported in column (1) of Table 4. We get an estimate of $\rho = -0.24$ significantly different from zero at 1%. For an expected reduction of one birth, we estimate an actual reduction of 0.24 births among Han Chinese. The remaining part was either not enforced or was driven by other factors. Since average ER for each cohort varies from zero to 3.3 in the sample, the estimated ρ implies that Han women reduced their fertility by about $0.24 * 3.3 = 0.8$ births between cohorts 1926 and 1945 in response to family planning policies.

[Insert Table 4 here]

Turning to the minority sample, we find a negative average effect in column (2). The coefficient is large and imprecisely estimated, suggesting that the average effect may mask substantial heterogeneity. In column (3), we allow the effect to vary with the Han share as per equation 2. The estimated ρ_0 is small and not significantly different from zero, thus providing no evidence of a direct policy effect. ϕ is significantly different from zero at 1%. The magnitude of the point estimate implies that, when Han women are expected to reduce their fertility by one child, a minority woman gives birth to 0.2 fewer children if the share of Han in her locality is close to 100%; if instead the share is 50%, she reduces her completed fertility by only 0.1 births. Given the average Han share in our minority sample (47%) and the observed variation in ER, we estimate that, through spillover effects, family planning policies indirectly contributed to a reduction of $0.2 \times 0.47 \times 3.4 = 0.3$ births in the completed fertility of minorities between cohorts 1926 and 1945.

Furthermore, we can address the concern that the potential direct effect might exist only in places with a large share of Han, for reasons unrelated to spillovers. In column (4), we control for the interaction of ER and the share of Han at the province-*hukou*-cohort level. Our estimates remain very stable, suggesting that only the local (prefecture-level) Han share matters. This is in line with the spillover interpretation and rules out the hypothesis that ϕ captures differences in policy implementation between coastal and inland provinces. Then, we consider heterogeneity by socio-economic development, which is correlated with s^H. We allow the direct effect to vary by *hukou* type, literacy, and education at the individual level in column (5) and at the prefecture-*hukou* level in column (6). The estimate of ϕ remains sizable and significant at 1%, indicating that even very conservative specifications provide evidence of spillovers. Lastly, we address the concern that, within a province, authorities might target places with many Han Chinese in order to maximize the impact of resources allocated to family planning. To construct a proxy of how much effort would be devoted to each prefecture-*hukou* unit within a province, we rank them according to the size of the Han population. If ϕ captured the targeting of family planning activities rather than local spillovers, the coefficient should become zero when we control for the rank interacted with ER. We show in column (7) that this is not the case. The local share of Han Chinese still predicts a reduction in the fertility of minorities whereas the number of Han Chinese relative to other localities in the province does not.

4.3 Support for the common trends assumption

Potential confounding shocks

We provide support for the common trends assumption by examining concomitant shocks. We consider three important events that could have potentially affected the fertility of minorities born between 1926–1945: the Great Chinese Famine (1959–1961), the Cultural Revolution (1966–1976), and the send-down movement (1962–1979). Exposure to these events varies across cohorts – they were at different stages of their reproductive lives – and across provinces and *hukou* types – they were hit more or less strongly. We check that the coefficients on ER and s^HER do not capture the impact of these shocks.

We construct a measure of exposure to each event, taking into account (i) how many births would have happened in the absence of the event during the relevant time period, and (ii) the local intensity of the event. We allow impacts on fertility to differ by *hukou* status. To quantify the intensity of each event, we build upon the literature. Regarding the great Chinese famine, we follow Chen and Yang (2016) and construct a *cohort loss* index at the prefecture level capturing the effect of the famine on surviving births. This index measures the percentage deviation from a prefecture-specific linear trend in cohort size during the years 1958–1963. As for the send-down movement, we use data from Gu (1997, p.302) on the number of sent-down youths (SDY) that settled in a given province in a given year, per 1,000 inhabitants. Over 90% of the urban SDYs were sent to rural areas of their own province (Gu, 1997, p.304). Therefore, our measure captures two sides of the same coin: experiencing the inflow of young people for rural women versus the outflow of young people for urban women. Lastly, we approximate the strength of the Cultural Revolution using the number of victims of political events recorded in local gazetteers gathered by Walder and Lu (2017). We aggregate the data at the province-year level and scale it by provincial population size. In column (8) of Table 4, we add the exposure to these potential confounding shocks to the baseline regression. The estimated ρ_0 remains small and insignificant, and the estimated ϕ barely changes in magnitude and remains significant at 1%.[16]

Pre-trends

We complement the findings on concomitant shocks with an analysis of pre-trends. We show that the evolution of completed fertility among older, hence unexposed, cohorts was not systematically related to the local share of Han nor to the future exposure of younger cohorts.

The first piece of evidence is a placebo test. We focus on the unexposed cohorts observed in the 1982 census (1918–1929) and we assign the policy exposure of younger cohorts (1934–1945) to them.[17] We use this placebo ER in equations 1 and 2 and report the results in Table A.3. The coefficient on the placebo exposure is not significant and, if anything, is positive. The coefficient on the placebo interaction term is a precise zero. Therefore, in the baseline analysis, our variable s^HER is not correlated with pre-trends in the fertility of minorities that would vary between prefecture-*hukou* groups.

The second piece of evidence comes from an event-study analysis. Using the sample of minority women born between 1918 and 1945, we compute for each cohort how old they were when the LLF campaign was launched in their own province. We then regress completed fertility

[16] Table A.2 provides more details on the construction of the control variables and their correlation with fertility.

[17] Note that the baseline sample consists of 20 cohorts (1926–1945) but we only observe 12 unexposed cohorts for the placebo test. To have a perfect point of comparison, we restrict the sample to the 12 youngest cohorts (1934–1945) in the baseline specification; estimates are the same as with the whole sample.

on the interactions between age group dummies and the share of Han. Figure 3 plots the coefficients on each interaction term and the corresponding 95% confidence interval.[18] Each coefficient represents the partial correlation between the local share of Han and the average fertility of minorities for women of different ages when the LLF campaign was launched. To relate these correlations to the policy exposure of Han women, we plot the average ER for each age group on the same graph. The figure shows that as long as Han women were too old to be exposed to the policies, the fertility of minorities does not correlate with the local Han share. A significant correlation appears when the Han peers were younger than 40 at the start of LLF; it becomes stronger as Han women were younger and hence more exposed to the LLF campaign. An event study of the Han sample reveals that 40 years old is precisely the age threshold below which Han women started to reduce their fertility; the magnitude of the reduction goes hand in hand with our measure of policy exposure.[19] In short, it is only when the LLF started affecting the Han women that fertility started diverging between minority women who had high and low shares of Han among their peers.

[Insert Figure 3 here]

5 Mechanisms

Fertility spillovers may operate through two main channels: general equilibrium effects and social interactions. For each channel, we proceed in three steps. First, we discuss the conditions of existence. Second, we identify the ethnic groups for whom these conditions are likely to be met using the classification described in Section 2.2. Third, we test whether the fertility response to family planning policies is stronger among these groups.

5.1 An economic channel

The quantity-quality (Q-Q) trade-off model of Becker (1991) predicts that, when the fertility of a Han couple is exogenously reduced, parents will invest more in each of their remaining children. The resulting increase in the average child quality may, in turn, raise the perceived returns to investments in quality if parents view the future as a tournament in which weaker children are

[18] Notes below Figure 3 provide more details on the sample and the econometric specification.

[19] The Han event study is presented in Figure A.3. We regress completed fertility on age group dummies and plot the estimates and 95% confidence intervals. The omitted category is the oldest age group (50-51 y.o.). The figure shows that the trend is flat between ages 40-41 and ages 50-51. Fertility starts decreasing among women aged under 40 when the LLF was launched. Coefficients become larger as the expected reduction in fertility ER increases.

left behind. This may prompt minority parents to raise their quality investments at the expense of quantity. This intuition has been recently formalized by Kim et al. (2021): they propose a model with status externalities and endogenous fertility and use it to explain how low birth rates and expensive education reinforce each other nowadays in South Korea. In our context, such a channel can work under two conditions: (i) Han parents respond to family planning policies by increasing quality investments; and (ii) minority parents decide to keep up with the Han. We focus on one dimension of quality – education – because, using data from the 1990 census, we can directly test when both conditions hold for educational investments.

We construct a sample of children aged between 16 and 25 years old, who lived with their parents in 1990. Co-residence is necessary to observe the mother's exposure to family planning policies. The age restriction reflects the trade-off between sample selection and censoring: older children leave the household upon marriage while younger children have not yet completed compulsory schooling (9 years of education).[20] Our measure of education is a dummy indicating whether the child attended school beyond the compulsory level, i.e., completed at least some senior high school. To look at the impact of policies on education, we estimate a variant of equations 1 and 2 for child i whose mother was born in year c.

We test the first condition, related to Han behavior, in Table 5 Panel A. To build confidence in the children sample, we start by replicating the fertility result by looking at the number of siblings ever born in column (1). Increasing the mother's expected reduction in fertility by one reduces the child's number of siblings by 0.17, which is not significantly different from the baseline Han estimate. Turning to the education outcome, in column (2), we find that the same variation leads to a 1.5 percentage points increase (10% of the mean) in the probability of attending senior high school. Under the assumption that family planning policies affect educational outcomes only through the effect on the number of siblings, our estimates imply that having one additional sibling reduces the probability of attending senior high school by $\frac{1.5}{0.17} = 9$ percentage points.[21] In terms of heterogeneity, Li and Zhang (2017) argue that the Q-Q trade-off is more salient when fathers are employed in the non-agricultural sector. Following their lead, we classify paternal occupations by educational requirement (see details below Table 5). Column (3) shows that

[20] 30% of 16 to 25 year-olds were not registered in the same household as their parents in 1990. If we restrict the sample to children between 16 and 23 years old, this proportion drops to 10% and we find similar but less precise estimates.

[21] Other studies have provided evidence of a Q-Q trade-off in China (Rosenzweig and Zhang, 2009; Liu, 2014; Li and Zhang, 2017). Li and Zhang (2017) instrument the number of siblings by the intensity of OCP enforcement and find that an extra sibling reduces junior high school attendance (grade seven to nine) by 11-13 percentage points. Rosenzweig and Zhang (2009) use the incidence of twinning under OCP as an exogenous fertility shock in rural families. They find that an extra sibling causes a reduction in schooling by 0.23 to 0.65 years.

the educational response to birth quotas is driven by fathers working in an education-intensive occupation.

[Insert Table 5 here]

We turn to the minority sample in Panel B to test the second condition. When we consider all children, we confirm that the average policy exposure in the mother's reference group leads to a reduction in the number of siblings (column 1) but this does not translate into more education (column 2). This is not surprising since most minority Chinese work in occupations requiring low levels of education, and we know from Panel A that their Han counterparts did not increase educational investment in their children. We hypothesize that only some ethnic groups, those commonly engaging in education-intensive and Han-dominated occupations, are likely to respond positively because they expect their children to compete with increasingly qualified Han. We provide support for the hypothesis in column (3) by showing that these groups respond differently (see details below Table 5). The coefficient on the interaction term is significant at 5% and the magnitude is large and comparable to the Han sample. By contrast, exposure has no impact on education among children from ethnic groups who weakly compete with the Han. In column (4), we add a second dimension of heterogeneity, cultural integration, which plays a key role in the analysis of the social channel. We find that ethnic groups culturally closer to the Han do not respond more than the others in terms of education. This suggests that competition motives explain the educational response of minorities better than imitation motives.[22]

We conclude that the conditions of existence of the educational channel are only met among ethnic groups facing a skilled competition from the Han in the labor market. If such a channel mediates spillovers, we should observe a stronger fertility response among these groups. In Table 6, we report the reduced-form estimates for different categories of minority women. The first column indicates that the coefficient is twice as large for ethnic groups competing with the Han as for other groups. These results provide suggestive evidence supporting the educational channel. Nonetheless, the fertility response is large and significant even in groups that did not increase their educational investments. For them, the economic channel may still be at play through changes in other forms of investments, like physical capital, financial capital, and health.

[22]Another piece of evidence supporting competition motives is that the educational response of minority groups competing with the Han is driven by groups of small population size. This is in line with the idea that the smallest minority groups have incentives to focus on the quality dimension since they cannot compete with other groups in terms of size (cf. Bezin et al. (2018) for a model of strategic fertility and educational decisions in deeply divided societies).

Unfortunately, the census provides very limited information on these dimensions. An alternative explanation is that social interactions also have an important role.

[Insert Table 6 here]

5.2 A social channel

Conformism is often modeled in economics as a utility cost generated by the distance between own behavior and average group behavior (Patacchini and Zenou, 2012).[23] This concept cannot explain spillovers from Han women onto minority women if our empirical groups are irrelevant from a social perspective. In other words, we expect minority women to be socially influenced by Han women in the same prefecture-*hukou*-cohort group if and only if they interact with each other and share the same fertility norms. Munshi and Myaux (2006) propose that couples learn about changes in the reproductive equilibrium through their social interactions; they show that, in Bangladesh, changes occur independently across religious groups in the same village because interactions and norms do not cross religious boundaries. In the same vein, we aim to identify ethnic groups in which women, (i) have knowledge of, and (ii) are willing to conform to Han fertility behavior.

Following Spolaore and Wacziarg (2021), we argue that linguistic distance is a key determinant of the diffusion of small families. First, individuals who share a common language face lower barriers to learning about new behaviors. Second, they are more likely to obey the same authorities that legitimate and enforce the rules sustaining the new reproductive equilibrium. However, the linguistic proximity might be a necessary but insufficient condition if minorities are cut off from the Han world. For instance, Daudin et al. (2020) argue that internal migration was a key driver of the convergence in regional fertility rates within France. This mechanism was not at play in our context, where migration was strictly limited. Given the lack of information transmission technologies, physical distance presumably hindered the learning and coordination process. Therefore, our hypothesis is that ethnic minorities using their own written language or residing in their own autonomous communities are unlikely to include Han Chinese in their social groups. Contrary to the educational channel, we are unable to test this hypothesis because, to our knowledge, there is no information on the ethnic composition of social networks in China in the 1970s.

[23] Note that this definition does not take a stand on the origins of the utility cost. In our context, we lack evidence to determine whether the motives to conform are intrinsic (deep individual preferences) or extrinsic (social sanctions).

In short, we expect the social influence of Han couples to be concentrated on ethnic groups that are culturally integrated, in terms of language and residence. We test whether the spillover effect of Han-targeting policies is higher among these groups in Table 6, column (2). We find a large and significant reduced-form estimate for them, whereas the coefficient is a precise zero for groups that are less integrated. Such a difference suggests that the fertility response of ethnic minorities is indeed partly mediated by social interactions. If we disaggregate the two components of the social channel (common language and residential integration), we find that each component is necessary but not sufficient to generate spillovers.[24]

Since ethnic minorities affected by the social channel are not the same as those affected by the economic channel, we can look at the interplay in column (3). Interestingly, we find no spillovers at all onto minorities characterized by a weak labor market competition and a weak cultural integration. The economic channel without the social channel is sufficient to generate large, significant spillovers, and vice versa. Finally, when both channels are present, the coefficient is close to the sum of the coefficients estimated when each channel plays individually. We conclude that economic and social forces are both important and act independently of each other. This is consistent with historical evidence from France (de la Croix and Perrin, 2018) and South Korea (Myong et al., 2021) showing that, although economic incentives explain a substantial part of the fertility and education transitions, additional explanatory power can be gained by incorporating cultural norms.

5.3 Alternative explanations

There are alternative mechanisms through which the economic and social channels may operate: changes in female opportunity cost of time and social learning about birth control. We explain why they are unlikely to drive spillovers in our context.

Starting with general equilibrium effects, the cost of a child depends not only on parental investments in quality but also on mothers' opportunity cost of time (Becker, 1991). If Chinese women indeed face a trade-off between production and reproduction, spillovers may be mediated by changes in female labor force participation.[25] We cannot directly test this mechanism be-

[24]We classify ethnic groups into six categories defined by a dummy for using the Han script and the terciles of the residential integration distribution. Table A.4 in Appendix presents the reduced-form estimate for each category. The coefficient is large and significant only for groups using the Han script *and* high enough in the residential integration distribution.

[25]For instance, suppose that the reduction in Han fertility frees up women's time to work. The large inflow of Han women in the labor market may drive adult wages down if males and females are somewhat substitutes. The decrease in wage then induces minority women to start working if men do not earn enough to feed the family. As

cause we do not have retrospective information on female labor force participation in the 1970s. However, we know from the literature that the causal impact of fertility on female labor supply is weak in China. Studies using twins (Guo et al., 2018) and variation in OCP enforcement (Zhang, 2017) both conclude that changes in fertility do not affect the labor supply of mothers. One explanation is that childcare duties fall partly to grandparents, particularly in rural China. Therefore, it seems implausible that Han-targeting family planning policies caused a reduction in minority fertility through substantial changes in mothers' opportunity cost of time.

Turning to social interactions, the literature distinguishes between social influence and social learning (Kohler, 2001, p.61). The mechanism we have emphasized so far, conformism, captures normative influences of Han Chinese on the fertility behaviors of minority women. We believe that conformism was particularly strong in the uncertain and brutal political context of China in the 1970s. An alternative mechanism could be that minority women learned about Han women's positive experiences with birth control, which would induce them to adopt the same methods. This mechanism requires the presence of unmet needs for birth control among Chinese women and that the LLF campaign provided reliable methods to meet these needs. Both conditions seem unlikely to hold in our context. Historians document that, in private, people were reluctant to use birth control methods (Scharping, 2013, p.107-108). The failure rates of condoms and pills were abnormally high, indicating couples' lack of cooperation. Therefore, the LLF campaign promoted long-lasting methods that could be enforced by the authorities: sterilization, abortion, and IUD insertion. They account for the vast majority of contraceptive methods used in 1981 whereas temporary methods account for less than 10% (Scharping, 2013, p.110). These surgeries were potentially traumatic, especially due to the medical injuries caused by unskilled paramedics. Another argument supporting the idea that contraception played a limited role is that the use of modern contraceptives among minorities remained very low: Poston (1986) documents that, in 1982, the share of Minority Chinese in the population is the best predictor of contraception prevalence in a province, with a negative correlation of -0.78 between both variables. For these reasons, we believe that the main social process at play was not learning but the pressure to conform.

a consequence, minority women reallocate their time away from reproduction.

6 Semi-structural approach

In this section, we show that the reduced-form results are consistent with the standard linear-in-means model of social interactions. We further argue that our empirical setting can be used as a partial population experiment to quantify the role of spillovers in the fertility transition in China.

6.1 A partial-population experiment

Following Manski (1993), we assume that the fertility decision of a woman is affected by the average fertility in her reference group.[26] We define the reference group as women holding the same type of *hukou* r, living in the same prefecture d within province p, and belonging to the same cohort c. This is the most relevant unit for which we have plausibly exogenous variation and enough observations per unit. In the minority sample, we observe 5,380 reference groups with approximately 50 women per group. We discuss alternative reference groups in Appendix A.3. The structural equation is:

$$y_i = \alpha + \beta X_i + \gamma \bar{X}_{rdc(i)} + \lambda V_{rdc(i)} + \theta \bar{y}_{rdc(i)} + \varepsilon_i \tag{3}$$

where the parameter of interest θ captures the spillovers in fertility decisions, namely the marginal change in the completed fertility of woman i when the average completed fertility in her reference group varies. For clarity of exposition, we use the same notations as in the reduced-form equations and denote $V_{rdc(i)}$ the vector including all contextual variables specific to the group: the time-varying provincial characteristics $Z_{rpc(i)}$, the set of cohort dummies $\eta_{c(i)}$, the set of prefecture-*hukou* dummies $\mu_{rd(i)}$, and the set of province-*hukou*-specific time trends $t_{rp(i)}c$. X_i are exogenous characteristics of the woman and $\bar{X}_{rdc(i)}$ are the mean exogenous characteristics of women in the reference group. Manski (1993) shows that one cannot separately identify γ and θ by estimating equation (3) using ordinary least squares (OLS), due to potential endogeneity issues and the simultaneity of individuals' actions.

To identify θ, we exploit a partial-population experiment (Moffitt, 2001). The idea is to construct an exogenous treatment affecting the behavior of only a share of individuals within each

[26] It is possible to provide micro-foundations for this functional form by assuming that people play a non-cooperative game with a specific utility function. The linear-in-means best-response is typically obtained by including a utility cost proportional to deviations from the average behavior in the group (Patacchini and Zenou, 2012). In addition, we postulate that the average group behavior may also enter the price function through general equilibrium effects, as discussed in Section 5.

group. In our case, we use the exemptions granted to ethnic minorities during family planning policies. We add the measure of policy exposure ER to model (3), assuming that ER $= 0$ for all minority women. Denoting outcomes and the characteristics of the Han by superscript H and those of ethnic minorities by M, we get the following augmented models for the two groups:[27]

$$y_i^M = \alpha + \beta X_i^M + \gamma \bar{X}_{rdc(i)} + \lambda V_{rdc(i)} + \theta \bar{y}_{rdc(i)} + \varepsilon_i^M \tag{4}$$

$$y_i^H = \alpha + \beta X_i^H + \gamma \bar{X}_{rdc(i)} + \lambda V_{rdc(i)} + \theta \bar{y}_{rdc(i)} + \delta \mathrm{ER}_{rpc(i)} + \varepsilon_i^H \tag{5}$$

where δ measures the share of the expected reduction in fertility that is actually realized. $\delta = -1$ means that all the extra births above the quota would have happened in the absence of the policy and did not due to the implementation of the policy. If enforcement was in fact not perfect, or if fertility decreased over time for other reasons, then $\delta \in (-1, 0]$.

Under the assumption that ER is exogenous conditional on the covariates, we can combine equations (5) and (4) to express the equilibrium outcome in the reference group as:

$$\bar{y}_{rdc} = \frac{\alpha}{1-\theta} + \frac{\beta+\gamma}{1-\theta}\bar{X}_{rdc} + \frac{\lambda}{1-\theta}V_{rdc} + \frac{\delta}{1-\theta}s_{rd}^H \mathrm{ER}_{rpc} \tag{6}$$

where s_{rd}^H is the share of Han women in group rd.[28] Intuitively, the average group fertility is affected by family planning policies if there are Han women in the group (s^H), they are exposed to the policies (ER), they respond (δ), and their response is amplified by spillovers (θ).

Plugging equation (6) into equation (4), we get the reduced-form equation for minorities:

$$y_i^M = \frac{\alpha}{1-\theta} + \beta X_i^M + \frac{\theta\beta+\gamma}{1-\theta}\bar{X}_{rdc(i)} + \frac{\lambda}{1-\theta}V_{rdc(i)} + \frac{\theta\delta}{1-\theta}s_{rd(i)}^H \mathrm{ER}_{rpc(i)} + \varepsilon_i^M \tag{7}$$

The structural model provides three insights when analyzing equation (2). First, $\rho_0 = 0$, i.e. there is no direct effect of the policies on minorities. Second, $\phi = \dfrac{\theta\delta}{1-\theta}$ has a structural interpretation. Under the common trend assumption, the reduced-form approach allows us to test whether $\theta \times \delta = 0$. This is sufficient to prove the existence of a direct effect and a spillover effect. Third, to be perfectly consistent with the structural model, we should control for group

[27]To simplify the notations in this section, we assume that all parameters are the same in the Han equation and in the Minority equation. However, in the empirical analysis, we relax this assumption. As long as θ is the same for both groups, allowing other parameters to vary by group does not affect the identification of spillovers. The key assumption is that θ is homogeneous; we discuss the case of heterogeneous θ in Appendix A.4.

[28]In principle, we can measure the Han share at the rdc level. However, in practice, we have around 50 women per group. Since the local Han share is very stable over time for the cohorts 1926-1945, we prefer to aggregate all cohorts in the same rd group to mitigate measurement errors. The point estimate of θ is unchanged if we use s_{rdc}^H instead of s_{rd}^H in the IV specification.

average characteristics $\bar{X}_{rdc}$ in the reduced-form specification. In Table 7 column (1), we estimate equation 7. We find an estimated $\phi = -0.19$ significant at 1%, very close to the estimate in Table 4 column (3) where we did not impose that $\rho_0 = 0$ and did not control for $\bar{X}_{rdc}$.

6.2 Identification of the spillover effect θ

To go one step further and identify θ, we turn to an instrumental variable approach. We estimate the following system of equations:

$$\bar{y}^{(-i)} = \psi s^H_{rd(i)}\mathrm{ER}_{rpc(i)} + \tilde{\beta}X^M_i + \tilde{\gamma}\bar{X}_{rdc(i)} + \tilde{\lambda}V_{rdc(i)} + \tilde{\varepsilon}_i \tag{8}$$

$$y^M_i = \theta\bar{y}^{(-i)} + \beta^* X^M_i + \gamma^*\bar{X}_{rdc(i)} + \lambda^* V_{rdc(i)} + \varepsilon^*_i \tag{9}$$

The endogenous regressor $\bar{y}^{(-i)}$ is the leave-me-out mean fertility (i.e., the average fertility in the reference group excluding individual i). The instrumental variable is $s^H_{rd(i)}\mathrm{ER}_{rpc(i)}$. Standard errors are clustered at the reference-group level.

The IV estimator of θ is consistent under two conditions. First, there is a robust partial correlation between the instrumental variable and the endogenous regressor, i.e., $\tilde{\psi} = \dfrac{\delta}{1-\theta} \neq 0$. In other words, Han women do respond to family planning policies. The first-stage regression provides a test for this condition. Second, the instrumental variable and the error term in the second-stage equation are uncorrelated, i.e., $E[s^H_{rd(i)}\mathrm{ER}_{rpc(i)}\varepsilon^*_i] = 0$. This condition is satisfied if, in addition to the common trends assumption discussed earlier, the exclusion restriction holds: average group exposure only affects minority women's fertility behavior through its effect on average group fertility.

Estimation results

We estimate equations (8) and (9) in columns (2) and (3) of Table 7. The first-stage regression reveals a strong partial correlation between $s^H\mathrm{ER}$ and the group average fertility: $\tilde{\psi} = \dfrac{\delta}{1-\theta}$ is significantly different from zero at 1% and the F-statistic is higher than 100. The IV regression yields an estimate of $\theta = 0.63$ with a 95% confidence interval of $[0.43; 0.83]$. The point estimate implies that a minority woman reduces her completed fertility by 0.63 births when the average group fertility is exogenously reduced by one. Using OCP and several specifications, Li and Zhang (2009) find point estimates between 0.5 and 0.9. Our analysis suggests that the higher part of their range tends to overestimate peer effects. It can be either because certain minorities were directly affected by birth restrictions after 1984 or due to confounding factors like structural reforms affecting the local demand for children.

[Insert Table 7 here]

Support for the exclusion restriction

The exclusion restriction could potentially be violated if minorities were *directly* affected by family planning policies in different ways depending on where they lived. We show that this is not the case by allowing for potentially heterogenous direct effects, in the same vein as the robustness checks for the reduced-form specification. Table A.1 Panel B confirms that θ remains in the range [0.46; 0.63] and significant at 1% when we control for ER as well as the interactions of ER with the provincial share of Han, with indicators of socio-economic development at the individual and group levels, and with the rank of the locality in terms of Han population. As predicted by the structural model, the main effect of ER is never significantly different from 0. Therefore, our IV estimate is not driven by differential policy enforcement between places with high and low shares of Han.

The last potential threat is that, although ethnic identity is in principle time-invariant, a reclassification happened in the 1980s during which 12 million people switched from either unofficial minorities or Han to official minorities.[29] At the scale of China, this is not salient: the correlation of ethnicities by prefecture-birth year between the 1982 and 1990 censuses is 0.96. Still, some women classified as minorities in the 1990 census were likely to be identified as Han beforehand and hence subject to LLF. To make sure that our main estimates are not driven by these women, we take advantage of the fact that reclassification was concentrated in some prefectures and some ethnic groups (Manchu, Tujia, Miao, Dong, Yilao and Qiang). In Table A.5, we exclude 12 prefectures with large changes in the share of Han between 1982 and 1990 in column (1) and we exclude the above-mentioned ethnic groups in column (2). The IV estimate is stable when excluding prefectures; it decreases a little but remains sizable and very precise when excluding ethnic groups. The decrease is consistent with the spillover interpretation: ethnic groups that are less prone to reclassification are culturally less close to the Han and hence less affected by the social channel.

Heterogeneous spillovers

[29]Source: http://www.gov.cn/test/2005-07/26/content_17366_2.htm (government website). Jia and Persson (2020) argue that identity choice is driven in part by material motives and that policies favoring minorities created incentives to avoid the Han ethnicity. They consider the case of Han-minority couples choosing the ethnicity for their children at birth, but their point may also apply to the rare circumstances in which changing ethnicity is possible.

An important assumption in the linear-in-means model is that all individuals respond similarly to a change in the group average outcome. In Appendix A.4, we relax this assumption in two ways. First, we consider the possibility that Han Chinese and minority Chinese respond differently to the same change in the group average. The main result is that our IV strategy gives a consistent estimator of the parameter in the minority equation. In the presence of heterogeneous spillovers, we should therefore specify in our conclusion that *minority* women reduce their fertility by 0.63 children in response to a decrease of one child in the group average.

Second, we discuss the case when the same individual responds differently to changes in the outcomes of other group members depending on the identity of these members. We explain how the identification of inter-ethnicity spillovers (between Han and minorities as a whole) and intra-ethnicity spillovers (within Han or within minorities as a whole) can be achieved in our partial-population experiment setting. The intuition is that minority Chinese are only affected by the policy through spillovers, which are proportional to the Han share, while the Han Chinese are also subject to a direct policy effect independent of the population structure. Formally, we use the structural equations to come up with a set of empirical IVs. The heterogeneous model does not reject the hypothesis that the inter- and intra-ethnicity parameters are the same and that they are equal to 0.63, providing no clear evidence against the homogeneity assumption. Still in magnitude, the intra-ethnicity parameter is larger, suggesting that individuals tend to put more weight on group members of their own ethnicity, in line with the social channel.

6.3 Identification of the direct effect δ

Following the method proposed by Lalive and Cattaneo (2009), we use the Han Chinese to get an estimate of the direct effect δ. Intuitively, to disentangle the direct policy effect and the indirect spillover effect, we compare the Han's response and the minorities' response, holding social interactions constant. Pooling the Han sample and the Minority sample, we instrument $\bar{y}$ with $s^H\text{ER}$ in the following equation:

$$y_i = \alpha + \beta X_i + \gamma \bar{X}_{rdc(i)} + \lambda V_{rdc(i)} + \theta \bar{y}_{rdc(i)} + \delta \mathbb{1}\{Han_i\}\text{ER}_{rpc(i)} + \varepsilon_i \tag{10}$$

where $\mathbb{1}\{Han_i\}$ is a dummy equal to one if the woman is Han Chinese. The coefficient on the interaction between ER and the Han dummy gives an estimate of δ. Moreover, we can test whether θ remains stable when including Han Chinese in the sample. An alternative strategy to identify δ is to look at within-group variation. Taking the group average of equation 10 and, before solving for $\bar{y}$, subtracting this expression from equation 10, we get:

$$y_i - \bar{y}_{rdc(i)} = \beta(X_i - \bar{X}_{rdc(i)}) + \delta(\mathbb{1}\{Han_i\} - s^H_{rd(i)})\mathrm{ER}_{rpc(i)} + \varepsilon_i \tag{11}$$

Results are reported in Table 7, columns (4) and (5).[30] The IV strategies yields an estimate of $\delta = -0.08$, and the within strategy yields an estimate estimate of $\delta = -0.07$, both significant at 1%. We also confirm that θ is around 0.6. These results support the assumption that the spillover parameter is the same for Han Chinese and minority Chinese.[31] We conclude that about 7% to 8% of the expected reduction in Han fertility actually took place as a direct result of the quotas. This accounts for approximately one third of the total effect on Han Chinese ($\rho = 0.24$ in column (1) of Table 4).

7 Discussion and conclusion

Our results help to explain the fast fertility decline during the 1970s in China. The total fertility rate (TFR) decreased from 5.7 children per woman in 1969 to 2.7 in 1978. Historians and demographers forged the concept of "induced fertility transition" to reflect the idea that population policies played a key role in this decline, particularly the LLF campaign (Zhang, 2017). Our results confirm this view. Using estimates of the Han regression, we find that the average policy effect is a reduction by 0.8 births, accounting for 60% of the decline in Han completed fertility between the cohort of 1926 (4.91 births) and the cohort of 1945 (3.58 births). Estimates in the literature range from 0.85 to 1.5 births (Chen and Huang, 2020; Chen and Fang, 2021; Babiarz et al., 2018).[32] They are larger than ours because they include younger cohorts in their sample and hence capture effects on marriage and on the onset of fertility. The structural model allows us to go one step further and separate the direct effect and the indirect effect. The total effect is

[30] For the sake of parsimony, we write equations 10 and 11 with a homogenous λ. We relax this assumption in the estimation by allowing fixed effects and time trends to be different between Han and minority Chinese. We can also allow β and γ to be Han-specific and Minority-specific and it does not change the estimates of interest.

[31] There are two other ways of estimating δ. First, using only the Minority sample and subtracting the reduced-form coefficient from the first-stage coefficient, we get $\delta = \frac{\delta}{1-\theta} - \frac{\theta\delta}{1-\theta} = -0.3 - (-0.2) = -0.10$. This number belongs to the confidence interval of the estimate of δ computed with the IV strategy on the pooled sample, providing another validity check of the model. Second, using only the Han sample, we estimate the reduced-form equation; δ is the coefficient on ER. We show in Appendix A.5 that this method leads to unstable and imprecise estimates.

[32] Note that in relative terms, our estimate of 60% of the decline is not directly comparable to estimates in the literature because they use the change in period fertility (TFR) while we use the change in cohort fertility (completed fertility) as a reference point.

approximately three times as large as the direct effect parameter δ, suggesting that two thirds of the policy impact on Han is driven by spillovers. Turning to the minority regression, we estimate that the policy caused a 0.3 reduction in the number of births, accounting for 40% of the decline in the completed fertility of minorities. This effect is entirely driven by spillovers. We conclude that, in the Chinese context, spillovers strongly magnified the policy impact. Economic and social interactions between couples generate a social multiplier, which helps us to understand how the LLF campaign could trigger the fertility transition.

In our framework, two conditions are necessary to generate strong spillovers. First, the policy should have a large effect on the population directly targeted; second, people should have strong economic or social incentives to keep up with others. Indeed, the reduced-form estimate of spillover effects is determined by the product of δ and θ; it is therefore close to zero if either one or the other is close to zero. In terms of external validity, we expect much weaker spillovers in the following scenarios: (i) a similarly stringent policy is enforced in a context with looser social ties, and (ii) a more lenient policy is enforced in a similar social context. Historically, both conditions were met in South Korea, Singapore, Iran, and to a smaller extent in Indonesia and India. These countries experienced fast fertility declines following strict family planning programs in contexts marked by strong political ideologies. Our approach to quantify the social multiplier is potentially applicable there.

Existing research on diffusion processes focuses on spontaneous fertility transitions, in which population policies have little to no role. We believe that studying transitions induced by policies is important because they took place in the world's most populated countries and substantially contributed to the global fertility transition. Moreover, both types of transitions differ in two main ways. First, in spontaneous transitions, people gradually learn about each other and the process takes time (Munshi and Myaux, 2006; Spolaore and Wacziarg, 2021; Blanc, 2021). Instead, we consider the case when a policy, serving as a coordination device, leads to a rapid change. Second, the literature on spontaneous transitions highlights the role of migrants, arguing that they foster the fertility transition in their communities of origin by transmitting ideas and information that prevail in their communities of destination. For instance, Daudin et al. (2020) show that migration flows to and from Paris explain the spread of low-fertility behaviors in France in the second half of the nineteenth century. Fargues (2007) and Beine et al. (2013) document that the same mechanism is currently at play at the international level between developed and developing countries. In our paper, we provide evidence that fertility preferences can be transmitted horizontally between cultural groups even in the absence of migration.

Which lessons can we learn? We argue that our framework may shed light on future fertility rates in China. Despite the abolishment of the OCP in 2016, birth rates kept falling and reached their lowest level in 2019 (National Bureau of Statistics, 2019). This is a worry for the government who is concerned about the deterioration of dependency ratios. A common explanation is that children are so costly that couples cannot afford more than one. Our results suggest that the reverse causal link also exists: couples spend a lot per child partly because they only have one. It may well be the case that all couples would be better off splitting the same amount of resources between two children, but they fail to coordinate. The government can get out of this low fertility trap by subsidizing second births. The main insight is that the subsidy does not have to be universal. The rise in fertility is predicted to endogenously propagate from the subsidy recipients to groups that are economically or culturally related.

To sum up, this paper provides empirical evidence of large spillovers in fertility choices in China. Other people influence a couple's childbearing decisions through conformism and educational investments. These findings are consistent with both the diffusionist view and the structuralist view of fertility transitions. A reduction in fertility spreads from some individuals to others, presumably because preferences change and also because the price of a child changes. The interplay between economic and social factors should, therefore, be taken into account when designing and evaluating family planning policies. Our results suggest that the fertility decline in China was induced to a sizable extent by population policies and amplified by a social multiplier; current "lowest-low" fertility levels are likely to be self-sustaining even in the absence of birth quotas.

References

Arduini, T., Patacchini, E., and Rainone, E. (2020). Treatment effects with heterogeneous externalities. *Journal of Business & Economic Statistics*, 38(4):826–838.

Babiarz, K. S., Ma, P., Miller, G., and Song, S. (2018). The limits (and human costs) of population policy: Fertility decline and sex selection in China under Mao. Working Paper 25130, National Bureau of Economic Research.

Becker, G. S. (1991). *A Treatise on the Family*. Harvard University Press.

Beine, M., Docquier, F., and Schiff, M. (2013). International migration, transfer of norms and home country fertility. *The Canadian Journal of Economics / Revue canadienne d'Economique*, 46(4):1406–1430.

Bezin, E., Chabé-Ferret, B., and de la Croix, D. (2018). Strategic fertility, education choices and conflicts in deeply divided societies. LIDAM Discussion Papers IRES 2018011, Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES).

Blanc, G. (2021). The cultural origins of the demographic transition in France.

Bobonis, G. J. and Finan, F. (2009). Neighborhood peer effects in secondary school enrollment decisions. *Review of Economics and Statistics*, 91(4):695–716.

Bongaarts, J. and Watkins, S. C. (1996). Social interactions and contemporary fertility transitions. *Population and Development Review*, 22(4):639–682.

Brown, K. M. and Laschever, R. A. (2012). When they're sixty-four: Peer effects and the timing of retirement. *American Economic Journal: Applied Economics*, 4(3):90–115.

Chan, K. W. (2009). The Chinese hukou system at 50. *Eurasian Geography and Economics*, 50(2):197–221.

Chen, Y. and Fang, H. (2021). The long-term consequences of China's "Later, Longer, Fewer" campaign in old age. *Journal of Development Economics*, 151:102664.

Chen, Y. and Huang, Y. (2020). The power of the government: China's family planning leading group and the fertility decline since 1970. *Demographic Research*, 42:985–1038.

Chen, Y. and Yang, D. (2016). Historical traumas and the roots of political distrust: Political inference from the Great Chinese Famine. Unpublished.

Ciliberto, F., Miller, A. R., Nielsen, H. S., and Simonsen, M. (2016). Playing the fertility game at work: An equilibrium model of peer effects. *International Economic Review*, 57(3):827–856.

Coale, A. J. and Chen, S. L. (1987). *Basic Data on Fertility in the Provinces of China, 1940-82*. East-West Population Institute, East-West Center.

Coale, A. J. and Watkins, S. C. (1986). *The Decline of Fertility in Europe*. Princeton University Press.

Daudin, G., Franck, R., and Rapoport, H. (2020). Can internal migration foster the convergence in regional fertility rates? Evidence from nineteenth century France. *Economic Journal*, forthcoming.

de Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

de la Croix, D. and Perrin, F. (2018). How far can economic incentives explain the French fertility and education transition? *European Economic Review*, 108:221–245.

Durlauf, S. N. and Walker, J. R. (2001). Social interactions and fertility transitions. In *Diffusion Processes and Fertility Transition: Selected Perspectives*. National Academy Press.

Ebenstein, A. (2010). The "missing girls" of China and the unintended consequences of the One Child Policy. *Journal of Human Resources*, 45(1):87–115.

Fargues, P. (2007). The demographic benefit of international migration: Hypothesis and application to the Middle Eastern and North African contexts. In Ozden, C. and Schiff, M., editors, *International Migration, Economic Development & Policy*, chapter 5. World Bank, Washington, DC.

Gu, H., editor (1997). *The History of China's Send-down Movement of Educated Youths (Zhong Guo Zhi Shi Qing Nian Shang Shan Xia Xiang Shi Mo)*. China Procuratorial Press.

Guo, R., Li, H., Yi, J., and Zhang, J. (2018). Fertility, household structure, and parental labor supply: Evidence from China. *Journal of Comparative Economics*, 46(1):145 – 156.

Hensvik, L. and Nilsson, P. (2010). Businesses, buddies and babies: Social ties and fertility at work. Working Paper Series 2010:9, IFAU - Institute for Evaluation of Labour Market and Education Policy.

Huang, W., Lei, X., and Zhao, Y. (2016). One-Child Policy and the rise of man-made twins. *Review of Economics and Statistics*, 98(3):467–476.

Huang, W. and Zhou, Y. (2016). One-Child Policy, marriage distortion and welfare loss. Unpublished.

Jia, R. and Persson, T. (2020). Choosing ethnicity: The interplay between individual and social motives. *Journal of the European Economic Association*, 19(2):1203–1248.

Kim, S., Tertilt, M., and Yum, M. (2021). Status externalities and low birth rates in Korea. Working paper.

Kohler, H.-P. (2000). Fertility decline as a coordination problem. *Journal of Devlopment Economics*, 63:231–263.

Kohler, H.-P. (2001). *Fertility and Social Interactions: An Economic Perspective*. Oxford University Press.

Lalive, R. and Cattaneo, M. A. (2009). Social interactions and schooling decisions. *Review of Economics and Statistics*, 91(3):457–477.

Lee, R. (2015). Becker and the demographic transition. *Journal of Demographic Economics*, 81(1):67–74.

Li, B. and Zhang, H. (2017). Does population control lead to better child quality? Evidence from China's One-Child Policy enforcement. *Journal of Comparative Economics*, 45(2):246 – 260.

Li, H., Yi, J., and Zhang, J. (2011). Estimating the effect of the One-Child Policy on the sex ratio imbalance in China: Identification based on the difference-in-differences. *Demography*, 48(4):1535–1557.

Li, H. and Zhang, J. (2009). Testing the external effect of household behavior: The case of the demand for children. *Journal of Human Resources*, 44(4):890–915.

Li, H., Zhang, J., and Zhu, Y. (2005). The effect of the one-child policy on fertility in China: Identification based on the differences-in-differences. Discussion Papers 00019, Chinese University of Hong Kong, Department of Economics.

Liu, H. (2014). The quality–quantity trade-off: Evidence from the relaxation of China's one-child policy. *Journal of Population Economics*, 27(2):565–602.

Lyngstad, T. and Prskawetz, A. (2010). Do siblings' fertility decisions influence each other? *Demography*, 47:923–34.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542.

Moffitt, R. A. (2001). Policy interventions, low-level equilibria, and social interactions. *Social dynamics*, 4(45–82):6–17.

Munshi, K. and Myaux, J. (2006). Social norms and the fertility transition. *Journal of Development Economics*, 80(1):1 – 38.

Myong, S., Park, J., and Yi, J. (2021). Social norms and fertility. *Journal of the European Economic Association*, 19(5):2429–2466.

National Bureau of Statistics of China (2010). *China Compendium of Statistics 1949-2008*. China Statistics Press.

Notestein, F. W. (1953). Economic problems of population change. *Proceedings of the Eigth International Conference of Agricultural Economists*, pages 13–31.

Patacchini, E. and Zenou, Y. (2012). Juvenile delinquency and conformism. *Journal of Law, Economics, and Organization*, 28(1):1–31.

Poston, D. L. (1986). Patterns of contraceptive use in China. *Studies in Family Planning*, 17(5):217–227.

Poston, D. L. and Gu, B. (1987). Socioeconomic development, family planning, and fertility in China. *Demography*, 24(4):531–551.

Qian, N. (2009). Quantity-quality and the One Child Policy: The only-child disadvantage in school enrollment in rural China. Working Paper 14973, National Bureau of Economic Research.

Rosenzweig, M. R. and Zhang, J. (2009). Do population control policies induce more human capital investment? Twins, birth weight and China's "One-Child" Policy. *Review of Economic Studies*, 76(3):1149–1174.

Scharping, T. (2013). *Birth Control in China 1949-2000: Population Policy and Demographic Development*. Routledge.

Spolaore, E. and Wacziarg, R. (2021). Fertility and Modernity. *The Economic Journal*. Forthcoming.

Walder, A. G. and Lu, Q. (2017). China political events dataset, 1966-1971. `https://stanford.app.box.com/s/1p228gewy2pjd3817ksq9kd4d6cz3jy8`. Accessed: 2019-03-23.

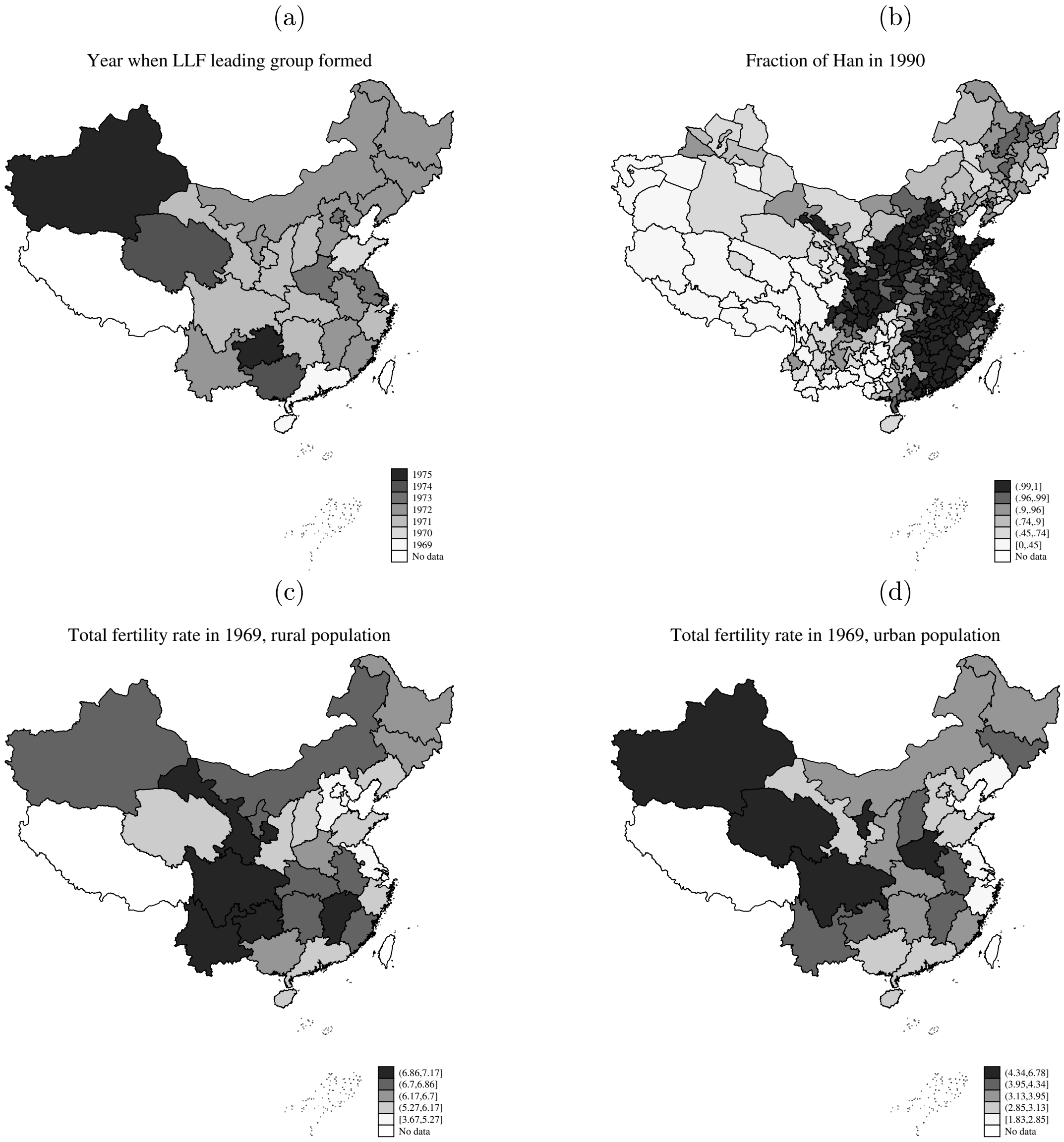
Wang, F., Zhao, L., and Zhao, Z. (2017). China's family planning policies and their labor market consequences. *Journal of Population Economics*, 30(1):31–68.

Wang, X. and Zhang, J. (2018). Beyond the quantity–quality tradeoff: Population control policy and human capital investment. *Journal of Development Economics*, 135:222 – 234.

Zang, X. (2015). *Ethnicity in China: A Critical Introduction*. Cambridge: Polity Press.

Zhang, J. (2017). The evolution of China's One-Child Policy and its effects on family outcomes. *Journal of Economic Perspectives*, 31(1):141–60.

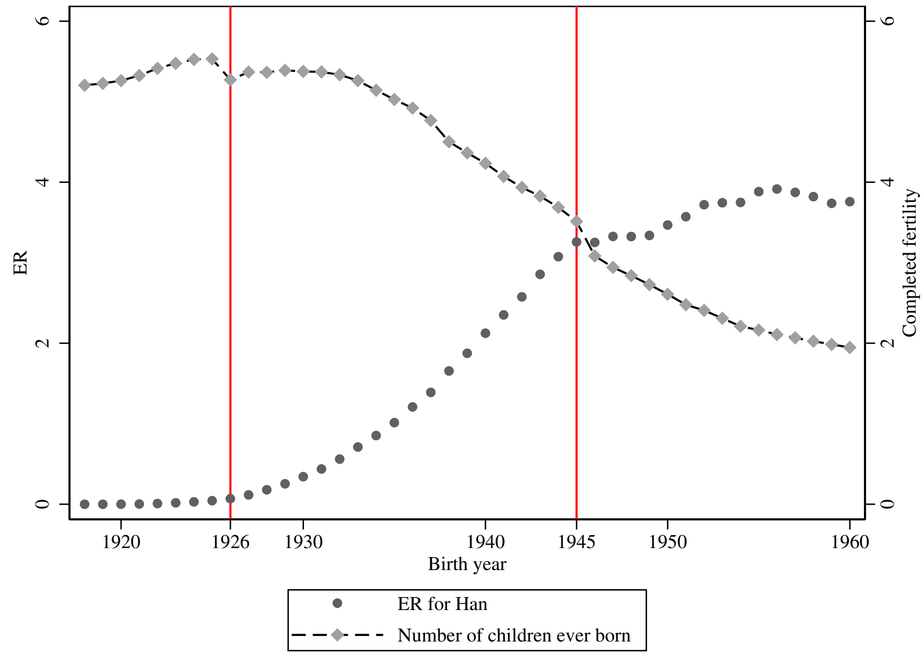
Figure 1: Sources of variation in treatment exposure



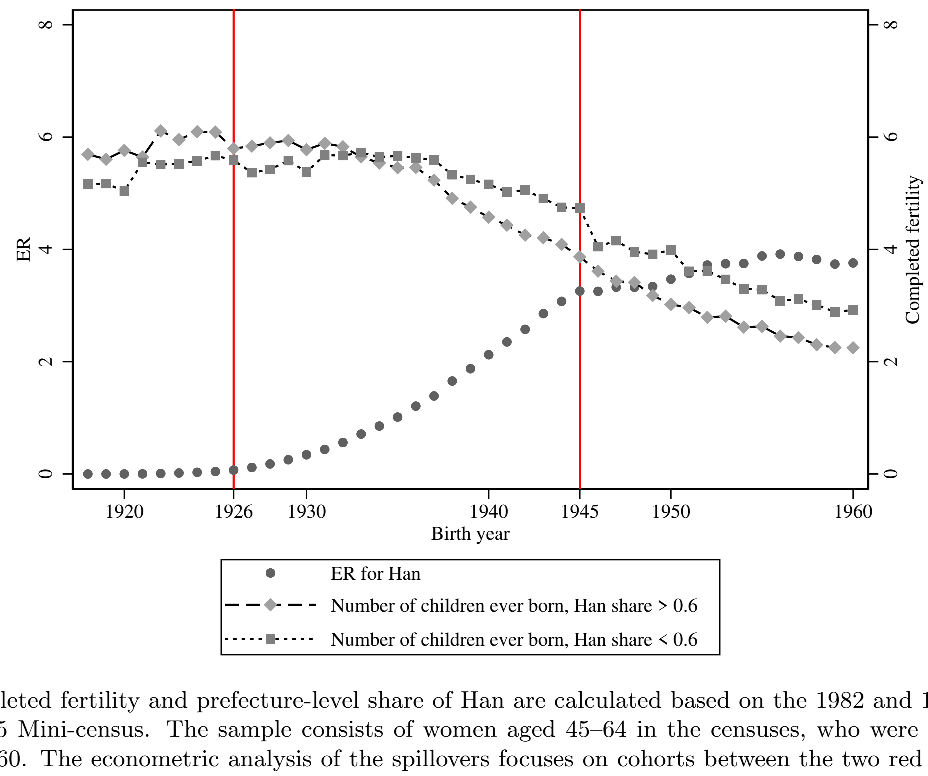
Note: Figure (a) plots the establishment year of provincial fertility leading groups taken from Chen and Fang (2021). Darker shades correspond to later dates when the leading groups were formed.
Figure (b) plots the share of Han residents by prefecture. Authors' own calculation based on the 1% sample of the 1990 Census. Darker shades correspond to higher share of Han population in the prefecture.
Figures (c) and (d) show province-level total fertility rates in 1969 by *hukou* type. Fertility data are compiled by Coale and Chen (1987) using retrospective data of fertility history collected by China's One per Thousand Sample Fertility Survey in 1982. Darker shades correspond to higher fertility rates.

Figure 2: Completed fertility and expected fertility reduction by ethnicity and by local Han share

(a) Han Chinese

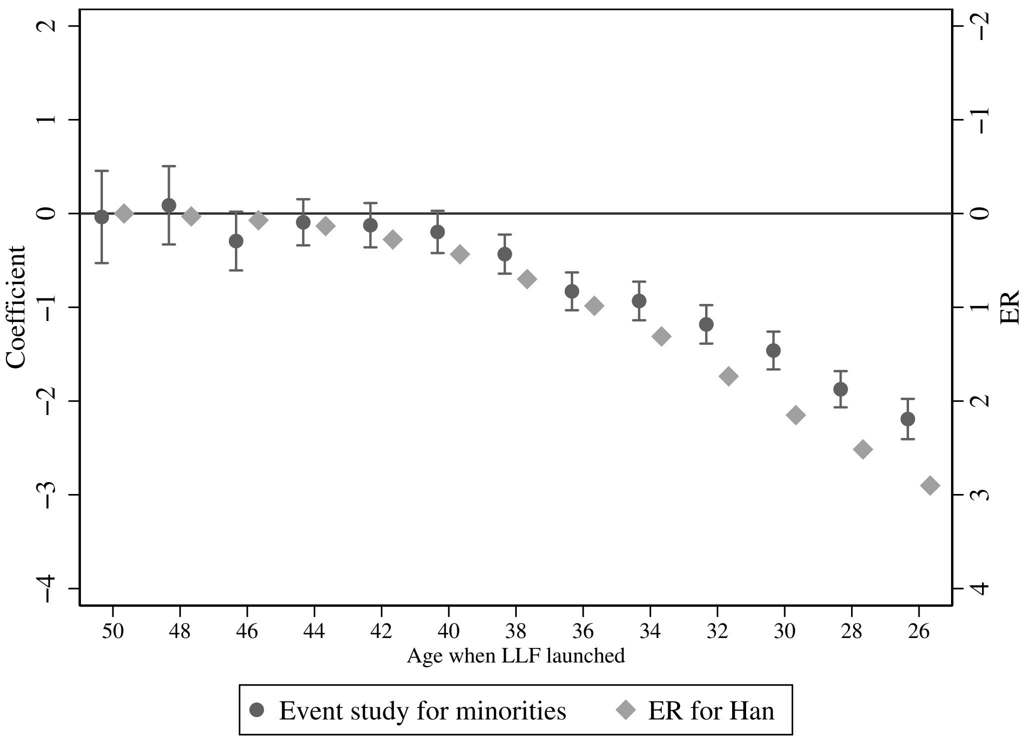


(b) Minority Chinese



Note: Completed fertility and prefecture-level share of Han are calculated based on the 1982 and 1990 Censuses and the 2005 Mini-census. The sample consists of women aged 45–64 in the censuses, who were born between 1918 and 1960. The econometric analysis of the spillovers focuses on cohorts between the two red vertical lines. In Figure (a), black dots represent the measure of policy exposure (ER) for each cohort of Han Chinese. Gray diamonds represent the evolution of completed fertility for Han women. In Figure (b), black dots represent the measure of policy exposure (ER) for each cohort of Han Chinese. Light (resp. dark) gray diamonds represent the evolution of completed fertility for minority women living in prefectures with more than 60% (resp. less than 60%) of the population being Han. We chose the 60% cut-off to have roughly the same number of prefectures in both categories.

Figure 3: Event study estimates of minority women



Note:
1. We use a sample of minority women aged between 26 and 51 years old when LLF was launched. Individual observations come from the 1% sample of the 1982 census (for cohorts 1918–1937) and 1990 census (for cohorts 1926–1945). We collapsed the data at the prefecture-*hukou*-cohort level (rdc).
2. We estimate the following event study model:

$$\bar{y}^M_{rdpc} = \sum_{j=13}^{25} \xi_j s^H_{rd} \mathbb{1}\{2j \leq (c - LLF_p) \leq 2j+1\} + \epsilon^M_{rdc}.$$

$(c - LLF_p)$ is the age of cohort c when LLF was launched in province p. We created bins of two years to have enough observations per age group. s^H_{rd} is the share of Han in prefecture d holding *hukou* r. $\bar{y}^M_{rdpc}$ is the average completed fertility of minority women in group rdc.
3. The plot shows the event study estimates ξ_j together with the 95% confidence intervals. The coefficient for age group $2j$ on the graph corresponds to ξ_j and measures the partial correlation between the share of Han and completed fertility for minority women aged $2j$ or $2j+1$ when LLF was launched.
4. Light gray diamonds represent the measure of policy exposure (ER) for the average Han woman in each age group (note that the y-axis is reversed). Standard event studies display a vertical line separating the before and after periods. Our design is richer since we can predict the continuous intensity of exposure. The before period is formed by women aged 46 and more when LLF was launched ($ER = 0$). In the after period, exposure increases slowly until age 40 and more quickly for younger women.

Table 1: Classification of main ethnic groups

		Cultural integration	
		Strong	Weak
Labor market competition	Strong	Man, Tujia, Li	Mongol, Korea
	Weak	Hui, Miao, Dong, Yao, Hani, She, Lisu, Va, Sui, Daur, Blang, Maonan	Zhuang, Tibetan, Uyghur, Yi, Buyei, Kazak, Bai, Dai, Kirgiz, Dongxiang, Jingpo, Xibe, Naxi

Note: We define the labor market competition as strong if more than 50% of the ethnic group's members aged between 25 and 44 are employed in occupations (i) dominated by the Han, meaning that the share of Han is higher than the median (65%), and (ii) requiring a high level of education. We define the cultural integration as strong if group members (i) use the Han script and (ii) are not residentially segregated, meaning that less than 80% of group members live in official autonomous regions or prefectures. Only groups with more than 100 observations are included in this table; they account for 97% of the total minority population. See details in Appendix A.1.

Table 2: Summary Statistics

	Minority sample		Han sample	
	Mean	SD	Mean	SD
Completed fertility	5.075	2.380	4.561	1.985
Completed fertility for 1926 cohort	5.205	2.595	4.913	2.393
Completed fertility for 1945 cohort	4.485	2.116	3.580	1.403
Reference group average fertility ($\bar{y}^{(-i)}$)	5.062	0.970	4.562	0.976
Expected reduction for Han women (ER)	1.569	1.224	1.481	1.257
ER under LLF	1.568	1.224	1.480	1.258
ER under OCP	0.001	0.006	0.001	0.007
ER for 1926 cohort	0.063	0.046	0.070	0.060
ER for 1945 cohort	3.393	0.955	3.286	1.199
Han Share (s^H)	0.491	0.313	0.963	0.100
Han Share x ER	0.754	0.823	1.425	1.230
Age in 1990	53.458	5.617	53.802	5.668
Rural *hukou*	0.851	0.357	0.766	0.423
Literate	0.299	0.458	0.390	0.488
Ever attend junior high school	0.079	0.269	0.118	0.323
Ever attend senior high school	0.027	0.161	0.042	0.200
Ever obtain vocational education	0.012	0.107	0.016	0.125
Ever attend college	0.006	0.075	0.010	0.100
Zhuang	0.202	0.401		
Hui	0.104	0.306		
Man	0.091	0.288		
Miao	0.079	0.270		
Uyghur	0.078	0.268		
Tujia	0.074	0.262		
Yi	0.069	0.254		
Mongol	0.044	0.205		
Tibetan	0.035	0.183		
Buyei	0.033	0.178		
Yao	0.028	0.166		
Dong	0.026	0.158		
Korean	0.024	0.154		
Bai	0.024	0.154		
Other ethnic groups	0.089	0.284		
Observations	58887	785479		

Note: Sample: women born between 1926 and 1945 from the 1990 Census. $\bar{y}^{(-i)}$ is the average fertility of all women except woman i in her reference group. Han share (s^H) is the share of Han in a woman's reference group. ER is the expected reduction in fertility to meet birth quotas set by family planning policies.

Table 3: Examples of constructing the measure of exposure to LLF

Panel A.						
Province, *hukou*	Jiangsu, rural		Xinjiang, rural		Xinjiang, urban	
Cohort	1930	1945	1930	1945	1930	1945
Policy year	1973	1973	1975	1975	1975	1975
Age in policy year	43	28	45	30	45	30
Panel B.						
Policy quota	2	2	2	2	2	2
AFR(15-19)	0.026	0.026	0.15	0.15	0.008	0.008
AFR(20-24)	0.248	0.248	0.317	0.317	0.296	0.296
AFR(25-29)	0.296	0.296	0.341	0.341	0.29	0.29
AFR(30-34)	0.205	0.205	0.259	0.259	0.166	0.166
AFR(35-39)	0.142	0.142	0.146	0.146	0.142	0.142
AFR(40-44)	0.070	0.070	0.113	0.113	0.076	0.076
AFR(45-49)	0.008	0.008	0.046	0.046	0.076	0.076
TFR in 1969	4.975	4.975	6.86	6.86	5.27	5.27
Panel C.						
Number of children already born	4.951	2.258	6.63	4.04	4.89	3.136
Number of children to be born	0.024	2.717	0.23	2.82	0.38	2.134
ER LLF	0.024	2.717	0.23	2.82	0.38	2.134

Note: Panel A: policy year is the year when the provincial fertility leading group was formed (Chen and Fang, 2021). Panel B: provincial age-specific fertility rates (AFR) are taken from Coale and Chen (1987). Panel C: we combine the information in Panel A and B to measure the exposure to LLF for each province-*hukou*-cohort group. Specifically, we use the AFR in light gray to predict the average number of children already born when the policy started, and the AFR in dark gray to predict the number of children to be born in the absence of the policy. When the number of children already born is larger than the policy quota (2 births), the expected fertility reduction (ER LLF) is equal to the number of children to be born. Appendix A.2 provides more details on the general formula that we use to construct ER.

Table 4: Direct and spillover effects of family planning policies on completed fertility

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. var.	Completed fertility							
Sample	Han	Minority						
				Potentially heterogenous direct effect				Other events
Exposure (ER)	-0.239***	-0.142	-0.022	-0.021	-0.042	-0.381*	-0.412*	0.064
	(0.042)	(0.158)	(0.157)	(0.217)	(0.218)	(0.230)	(0.233)	(0.168)
Han share × Exposure (s^H_{rd} × ER)			-0.208***	-0.208***	-0.212***	-0.134***	-0.183***	-0.203***
			(0.043)	(0.044)	(0.044)	(0.044)	(0.067)	(0.044)
Baseline controls and fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	No	No	No	s^H_{rpc}ER	s^H_{rpc}ER W_iER	s^H_{rpc}ER W_iER $\bar{W}_{rd}$ER	s^H_{rpc}ER W_iER $\bar{W}_{rd}$ER $rank_{rd}$ER	Famine Send-down Cultural Revolution
Average policy effect, cohort 1945	-0.8 births		-0.3 births					
R^2	0.234	0.173	0.173	0.173	0.173	0.174	0.174	0.173
Number of clusters	1120	1010	5514	5514	5514	5514	5514	5514
Mean dep var	4.561	5.075	5.075	5.075	5.075	5.075	5.075	5.075
Observations	785479	58887	58887	58887	58887	58887	58887	58887

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.

1. Each column represents a separate regression. Robust standard errors in parentheses are clustered at the province-*hukou*-cohort level in columns (1) and (2), and at the prefecture-*hukou*-cohort level in columns (3)-(8).

2. ER is the expected fertility reduction to meet the birth quota set by family planning policies and is measured at the province-*hukou*-cohort level. s^H_{rd} is the share of Han in prefecture d holding *hukou* r.

3. Baseline controls: prefecture-*hukou* fixed effects, cohort fixed effects, province-*hukou*-specific cohort trends, individual-level controls, and province-level controls. Individual-level controls include dummy indicators of educational attainment and ethnic identity. Province-level controls include secondary school teachers per capita, health workers per capita, hospital beds per capita, logarithm of gross regional product per capita, and population density measured at age 25 and the interactions of these variables with provincial total fertility rates in 1969.

5. Additional controls: s^H_{rpc} denotes the share of Han in the province-*hukou*-cohort group. W_i includes dummy indicators of *hukou* status, literacy, and high school attainment. $\bar{W}_{rd}$ includes the share of urban *hukou* holders in the prefecture as well as the literacy and high school graduation rates at the prefecture-*hukou* level. To construct $rank_{rd}$, we order the prefecture-*hukou* units within a province in terms of number of Han Chinese, and group them by quintile. See Table A.1 and Table A.2 for coefficients on the additional controls and for details on how we constructed the exposure to other events (Famine, Send-Down and Cultural Revolution).

4. The analysis is based on a sample of women born between 1926 and 1945 from the 1% sample of the 1990 Chinese census data.

5. The average policy effect for Han women born in 1945 is the product of the average exposure and the estimate from column (1) $(3.3 \times (-0.24) = -0.8)$. The average policy effect for minority women born in 1945 is the product of the average local Han share, the average Han exposure, and the estimate from column (3) $(0.47 \times 3.4 \times (-0.2) = -0.3)$.

Table 5: Quality-Quantity trade-off

	(1)	(2)	(3)	(4)
Dep. var.:	# siblings	Attend senior HS		
Panel A. Han Chinese children				
Mother's Exposure (ER)	-0.168***	0.015***	0.008	
	(0.044)	(0.006)	(0.005)	
Father's occupation education-intensive			0.008***	
× Mother's Exposure			(0.002)	
R^2	0.371	0.295	0.301	
Number of clusters	1120	1120	1120	
Mean dep var	4.053	0.143	0.143	
Observations	824858	824858	824858	
Panel B. Minority Chinese children				
Han share × Mother's exposure (s^H × ER)	-0.236***	-0.002	-0.004	-0.002
	(0.051)	(0.004)	(0.004)	(0.004)
Strong labor market competition with Han			0.011**	0.012**
× Han share × Mother's Exposure			(0.006)	(0.006)
Strong cultural integration with Han				-0.004
× Han share × Mother's Exposure				(0.004)
R^2	0.303	0.295	0.295	0.295
Number of clusters	4393	4393	4393	4393
Mean dep var	4.835	0.092	0.092	0.092
Observations	67665	67665	67665	67665

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.
1. Each column of each panel represents a separate regression. Robust standard errors in parentheses are clustered at the province-*hukou*-cohort level in Panel A and at the prefecture-*hukou*-cohort level in Panel B.
2. The analysis uses a sample of children aged 16-25 who are registered in the same household as their parents in the 1990 Census. We focus on children whose mothers were born between 1926 and 1945 to be consistent with the sample used in Table 4.
3. The dependent variable in column (1) is the total number of siblings ever born. The dependent variable in columns (2) and (3) is a dummy equal to one if the child attended school beyond the compulsory level i.e. completed at least some senior high school (> 9 years of education).
4. Panel A uses the same specification as column (1) of Table 4. Panel B uses the same specification as column (3) of Table 4; in particular we control for ER to capture a potential direct effect of family planning policies on minorities.
5. In column (3) of Panel A, we allow the coefficient on ER to vary by the educational requirement of the father's occupation. We consider that an occupation is education-intensive if the fraction of workers who have ever attended senior high school is above the median (13%) among all occupations. Appendix A.1 provides details on the classification of occupations.
6. In column (3) of Panel B, we allow the coefficient on s^H × ER to vary by the ethnic group's level of competition with Han Chinese in the labor market. In column (4) of Panel B, we further allow the coefficient on s^H × ER to vary by the ethnic group's level of cultural integration with Han Chinese. See notes below Table 1.

Table 6: Spillover effects, by level of labor market competition and cultural integration

Dep. var.: completed fertility of minorities	(1)	(2)	(3)
Economic channel: labor market competition with Han			
Weak × Han share × Exposure	-0.181***		
	(0.043)		
Strong × Han share × Exposure	-0.363***		
	(0.053)		
Social channel: cultural integration with Han			
Weak × Han share × Exposure		-0.046	
		(0.048)	
Strong × Han share × Exposure		-0.302***	
		(0.044)	
The interplay of both channels			
No channel × Han share × Exposure			-0.038
			(0.051)
Only economic channel × Han share × Exposure			-0.147**
			(0.072)
Only social channel × Han share × Exposure			-0.271***
			(0.045)
Both channels × Han share × Exposure			-0.435***
			(0.053)
R^2	0.173	0.174	0.174
Number of clusters	5514	5514	5514
Observations	58887	58887	58887

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.
1. Each column represents a separate regression. Robust standard errors in parentheses are clustered at the prefecture-*hukou*-cohort level.
2. All columns use the same specification and sample as column (3) of Table 4. Each row represents the coefficient on s^H_{rd}ER interacted with a dummy for each category of ethnic groups: weak vs. strong labor market competition in column (1) and weak vs. strong cultural integration in column (2). In column (3), no channel means weak competition and weak integration (27,962 obs,, mean dep. var.=5.12); only economic channel means strong competition and weak integration (4,052 obs., mean dep. var.=4.82); only social channel means weak competition and strong integration (17,034 obs., mean dep. var.=5.19); both channels means strong competition and strong integration (9,839 obs., mean dep. var.=4.87). Note that we control for ER in all regressions.
3. Table 1 provides details on the classification of ethnic groups by level of labor market competition and cultural integration.

Table 7: Estimating the spillover effect θ and the direct effect δ

	(1)	(2)	(3)	(4)	(5)
Sample	Minority			Han and Minority	
Dep. var.	y^M	$\bar{y}^{(-i)}$	y^M	y_i	$y_i - \bar{y}^{(-i)}$
	RF	FS	IV	IV	Within-group
Han share × Exposure	-0.190***	-0.302***			
	(0.042)	(0.030)			
Group average fertility (θ)			0.630***	0.602***	
			(0.103)	(0.080)	
Han dummy × Exposure (δ)				-0.084***	
				(0.024)	
(Han dummy - Han share) × Exposure (δ)					-0.072***
					(0.013)
R^2	0.174	0.849	0.166	0.231	0.017
Number of clusters	5514	5514	5514	12642	12642
Mean dep var	5.075	5.075	5.075	4.597	4.597
F statistics			102.981	78.050	
Observations	58887	58887	58887	844138	844138
Individual characteristics	Yes	Yes	Yes	Yes	Yes
Group average characteristics	Yes	Yes	Yes	Yes	Yes
Provincial characteristics	Yes	Yes	Yes	Yes	Yes
Prefecture-*hukou* FE	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
Province-*hukou* Trends	Yes	Yes	Yes	Yes	Yes

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.
1. Each column represents a separate regression. Robust standard errors in parentheses are clustered at the prefecture-*hukou*-cohort level.
2. Individual characteristics include dummy indicators of educational attainment and ethnic identity. Shares of women in the same reference group that completed at least some primary/lower secondary/upper secondary/vocational/college education are included separately as well as the shares of women from large ethnic groups (groups with more than 1.5 million population in 1990, including Han). Provincial characteristics include secondary school teachers per capita, health workers per capita, hospital beds per capita, logarithm of gross regional product per capita, and population density measured at age 25 and the interactions of these variables with provincial total fertility rates in 1969.
3. We control for prefecture-*hukou* fixed effects, cohort fixed effects, and province-*hukou* linear trends. In columns (4) and (5), we further allow these fixed effects and trends to be Han- and minority-specific. The results remain the same if we allow the coefficients of other covariates to be Han- and minority-specific too.
4. F statistics are the first-stage F statistics of the instrument $s^H_{rd}ER$.
5. RF = Reduced-Form, FS = First Stage and IV = Instrumental Variable.

For Online Publication

Appendix A.1 Classification of ethnic minority groups

This appendix provides details on the classification of ethnic minority groups into strong vs. weak labor market competition and strong vs. weak cultural integration with the Han Chinese. All the calculations are based on the 1% sample of the 1990 Population Census. To define the labor market categories, we restrict the sample to working age people (between 25 and 44 years old in 1990). We use the definitions below.

- **Labor market competition**: we consider the labor market competition to be strong if more than 50% of group members work in an occupation defined as (i) Han-dominated and (ii) education-intensive. We use the two-digit occupation codes used in the 1990 Census to identify the occupations.
  - *Han-dominated occupation*: in each province, for a given occupation, we compute the fraction of Han Chinese. We define a minority person as working in a Han-dominated occupation if this fraction is higher than the median (65%). This fraction is below 1% in occupations like forestry workers in the Qinghai province and above 90% in occupations like healthcare providers in the Jilin province.
  - *Education-intensive occupation*: in each province, for a given occupation, we compute the fraction of men and the fraction of women who ever attended senior high school. We define a minority man (resp. woman) as working in an education-intensive occupation if this fraction is higher than the median (13% for men, 5% for women). This fraction is below 1% in occupations like farmers in the Guizhou province and above 90% in occupations like teachers in Guangxi province.
- **Cultural integration**: we consider the cultural integration to be strong if a group satisfies two conditions: (i) the group has a history of using the Han Chinese script as the written language, and (ii) the group is not residentially segregated from the Han.
  - *Linguistic distance*: we rely on the following official document accessed on 2019-09-27: `https://www.mct.gov.cn/whbphone/ggfw_phone/whzxzgph/ys/wzysf/201111/t20111121_687996.htm`. The document specifies which written language is used by each minority group. Out of 55 minority groups, 12 do not use the Han script, among which the largest groups are the Uygur, Yi, Mongol, Tibetan, and Korean.

– *Residential segregation*: we define a group as residentially segregated from the Han if more than 80% of group members were living in official minority autonomous regions or prefectures in 1990. Out of 55 minority groups, 12 fall into this category. Among the 14 largest groups, the fraction ranges from 6.1% for Man to 99.9% for Uyghur.

Note that for the analysis of the quantity-quality trade-off among the Han people in Table 5 Panel A, we need to determine whether the father worked in an education-intensive occupation. To be consistent with the definition above, we define a Han father as working in an education-intensive occupation if the fraction of men who ever attended senior high school is above 13%. Our results remain unchanged if we use other thresholds between 10 and 15%. An alternative would be to split the sample between agricultural and non-agricultural sectors, like Li and Zhang (2017). However, we notice that there is substantial variation in terms of average educational attainment across occupations within both sectors. That is why we opted for a data-driven approach to measure the level of education requirement for each occupation.

Appendix A.2 Construction of the policy exposure measure

Table 3 provides examples to illustrate how we constructed the expected reduction in fertility triggered by the LLF (ER LLF) for selected cohort-province-*hukou* groups. The general formula for any Han woman holding a $r = \{urban, rural\}$ *hukou*, living in province p and born in year c is the following:

$$\text{ER LLF}_{rpc} = \begin{cases} 0 \text{ if } TFR_{rp} - \text{LLF quota} \leq 0 \\ \min\{\sum_{a=15}^{49} AFR_{rp}(a) \cdot I[c + a \geq LLF_p, 1990 - c \geq a], \\ \qquad TFR_{rp} - \text{LLF quota}\} \text{ otherwise} \end{cases}$$

where LLF_p is the year when the LLF campaign started in province p; LLF quota $= 2$ is the birth quota set by LLF; TFR_{rp} is the total fertility rate for women holding a *hukou* r in province p in 1969; $AFR_{rp}(a)$ is the annual fertility rate for women aged a holding a *hukou* r in province p in 1969. $I[\cdot]$ is an indicator function that takes the value one if the argument is true and takes value zero otherwise. $c + a \geq LLF_p$ is true if cohort c was exposed to LLF at age a, while $1990 - c \geq a$ is a technical restriction ensuring that cohort c had not turned age a in year 1990, the census year. The sum of $AFR_{rp}(a)$ yields the total number of children to be born after the start of LLF.

Note that ER LLF differs from the measure of policy exposure constructed by Chen and Fang (2021) because they consider the number of children to be born, whereas we consider the number of children to be born beyond the quota. Both measures are identical when the quota is binding, and they differ in two cases. The first case happens when the pre-policy total fertility rate was already below the birth quota set by LLF, e.g. for urban *hukou* holders in Shanghai. The quota never binds and hence the expected reduction is zero for women of all ages. Instead, in Chen and Fang (2021), the exposure would be strictly positive for all age groups who have not completed their fertility yet. The second case is mostly relevant for cohorts born after 1945. Our formula takes into account that the maximum reduction expected in a given province-*hukou* group is the difference between the total fertility rate in 1969 and the policy quota. For example, if the initial fertility is five children, the expected reduction is capped by three because the youngest, childless cohorts are allowed to have two children. Instead, in Chen and Fang (2021), the exposure of youngest cohorts would be equal to five. Their measure is therefore equal to ER LLF + 2 for those cohorts who started having children once the quota was in place. In practice, both measures are similar in our sample of interest (cohorts 1926-1945) because (i) very few places had an initial fertility rate lower than two, and (ii) very few cohorts had fewer than two children when the LLF started. The main advantage of our definition is that the interpretation of the coefficient on ER is straightforward: this is the share of the expected reduction which is actually implemented.

We create another indicator measuring the additional reduction in fertility required to reach the stricter birth quota set by OCP, assuming that the LLF quota has been met. The (additional) exposure to OCP is defined as follows:

$$\text{ER OCP}_{rpc} = \begin{cases} 0 \;\text{ if }\; \widetilde{TFR}_{rp} - \text{OCP quota}_r \leq 0 \\ \min\{\sum_{a=15}^{49} \widetilde{AFR}_{rp}(a) \cdot I[c + a \geq OCP, 1990 - c \geq a], \\ \qquad \widetilde{TFR}_{rp} - \text{OCP quota}_r\} \;\text{ otherwise} \end{cases}$$

where $OCP = 1979$ is the year when the OCP is introduced; OCP quota_r is the birth quota equal to one for urban women and 1.5 for rural women to take into account the exemption for rural couples with a firstborn daughter. We construct measures of initial fertility rates in 1979, denoted by $\widetilde{TFR}_{rp}$ and $\widetilde{AFR}_{rp}(a)$, assuming full compliance with the LLF quota. More precisely, we keep the 1969 age-specific fertility rates up to the age when the cumulated fertility equals two children; afterwards, we assume that fertility rates are zero. Note that for women in our sample, ER OCP is always very close to zero even if we marginally change the assumptions about initial fertility rates. Remember that these women were between 34 and 53 years old in 1979; they had already given birth to their second child in the vast majority of provinces and were therefore not exposed to the switch in the policy quota from two to one.

Appendix A.3 Robustness checks

A Estimations without linear provincial trends

In the main analysis, we include province-*hukou*-specific trends to control for unobserved determinants of fertility that vary linearly with time. If we remove these trends, the coefficient on $s^H \times ER$ increases in magnitude, as reported in Table A.6, column (2). One explanation could be that the linear trends capture part of the identifying variation, in which case our main estimates are too conservative. A competing explanation is that estimations without trends overestimate the parameter due to omitted variables. For instance, the fertility of minorities did not decrease at all in provinces with the lowest shares of Han (Guangxi, Qinghai and Xinjiang) and this could be due to different trends in provincial socio-economic development.[33]

To discriminate between both explanations, we look separately at provinces with low Han shares in columns (3) and (4) and provinces with high Han shares in columns (5) and (6). Specifications with and without trends give very similar estimates, suggesting that there is no omitted variable that varies linearly with time at the provincial level *within each sub-sample*. Importantly, estimates in columns (4) and (6) are smaller in magnitude than estimates obtained with the whole sample without trends (column 2). This proves that omitting provincial trends in the baseline specification would be a mistake because provinces with high and low Han shares follow different paths.

Another interesting finding in Table A.6 is that magnitudes are slightly larger in columns (3) and (4) than in columns (5) and (6), i.e., we estimate larger spillovers when the support of s^H tends to be concentrated in the lower part of the distribution compared to the upper part. This suggests that indirect policy effects might vary non-linearly with the local Han share. To further investigate this question, we relax the linearity assumption in the reduced-form specification by replacing s^H with dummy variables for each bin of 0.1 in Equation 2. We plot the coefficient and confidence interval for each dummy variable in Figure A.5; the omitted category is the first bin.

[33] See Figure A.4a and Figure A.4b, where we plot the evolution of completed fertility and ER for the Han and completed fertility and $s^H ER$ for the minorities, province by province.

We find that the magnitude is weakly increasing with the local share of Han. Point estimates vary quasi-linearly when $s^H \in [0, 0.7)$ and remain stable around -0.2 when $s^H \geq 0.7$. There is a threshold above which an increase in the local share of Han does not lead to a stronger response by minorities.[34]

B Alternative measure of policy exposure

Our baseline measure of policy exposure depends on age-specific fertility rates observed in 1969 at the province-*hukou* level. Ideally, we would like to use fertility rates among the Han but we do not have information by ethnicity. If Han and minority Chinese had systematically different fertility rates in 1969, the province average is a noisy measure of the Han average and the magnitude of the error is correlated with the share of Han in the province.

To circumvent this concern, we construct an alternative exposure variable (ER^{nat}) based on national age-specific fertility rates in 1969. For a given age when LLF was launched, this variable is the same in all provinces so there is no measurement error correlated with provincial characteristics. Here, we exploit only the staggered adoption of LLF across provinces and not the cross-sectional variation in pre-policy fertility levels. Column (7) of Table A.6 report an estimate of $\phi = -0.255$, which is close to the baseline estimates and significant at the 1% level. Therefore, our identification strategy is not invalidated by potential measurement errors in ER and does not crucially rely on the variation in pre-policy fertility across provinces.

C Heterogeneous treatment effects

de Chaisemartin and D'Haultfoeuille (2020) argue that interpreting the treatment effect coefficient in linear regressions with group and period fixed effects is a challenge if treatment effects are heterogeneous. The coefficient is the weighted sum of the average treatment effect (ATE) in each group and period, with weights that may be negative. When negative weights are large and correlated with the heterogeneous treatment effect, the estimated coefficient and all ATEs can have different signs. The scope for negative weights is salient in the case of continuous treatment variables due to the lack of groups where treatment remains stable between two periods.

In our setting, we argue that negative weights are not a serious issue when interpreting the

[34]Deriving implications for the linear-in-means model discussed in Section 6 goes beyond the scope of this paper. We simply note that our baseline coefficient of -0.2 on $s^H ER$ tends to underestimate the indirect policy effect in reference groups not dominated by the Han. If we restrict the sample to the groups with $s^H < 0.7$, for whom the linearity assumption holds, we get an estimate of $\theta = 0.74$ which is not significantly different from our baseline estimate.

sign of the reduced-form coefficient. We design a specification in which the incidence of negative weights is very limited by dichotomizing the exposure measure ER and the Han share s^H. Using the Stata package provided by de Chaisemartin and D'Haultfoeuille (2020), we find that only 10% of weights are negative and their sum is equal to -0.016. As shown in Table A.6 column (8), this new specification yields a significantly negative ϕ, making it implausible that all policy effects are positive.

D Alternative reference groups

Our baseline reference group is the prefecture-*hukou*-cohort level. It may be argued that a woman is influenced not only by her cohort but also by older women when making fertility decisions. Therefore, we extend the definition to include in the reference group women who are one or two years older. As shown in the third column of Table A.5, the IV estimate remains stable.

A second alternative is to use a smaller geographical unit like the county instead of the prefecture. In this case we keep three cohorts instead of one in order to have enough observations per group. We exclude extremely small reference groups because the group average measures are too noisy. Results are reported in the fourth column of Table A.5. We find that $\theta = 0.74$, i.e. stronger spillovers when we use a more local reference group. This is in line with the hypothesis that spillovers are generated by social interactions.[35]

Appendix A.4 Heterogeneous spillovers

A Different parameters for Han Chinese and minority Chinese

We rewrite the model to allow for different θ in the Minority equation and in the Han equation:

$$y_i^M = \alpha + \beta X_i^M + \gamma \bar{X}_{rdc(i)} + \lambda V_{rdc(i)} + \theta^M \bar{y}_{rdc(i)} + \varepsilon_i^M$$
$$y_i^H = \alpha + \beta X_i^H + \gamma \bar{X}_{rdc(i)} + \lambda V_{rdc(i)} + \theta^H \bar{y}_{rdc(i)} + \delta \mathrm{ER}_{rpc(i)} + \varepsilon_i^H$$

The group average fertility is now:

$$\bar{y}_{rdc} = \frac{\alpha}{1-\theta'} + \frac{\beta+\gamma}{1-\theta'}\bar{X}_{rdc} + \frac{\lambda}{1-\theta'}Z_{rdc} + \frac{\delta}{1-\theta'}s_{rd}^H \mathrm{ER}_{rpc}$$

where $\theta' = s_{rd}^H \theta^H + (1 - s_{rd}^H)\theta^M$. If $\theta^H \neq \theta^M$, the effect of the average policy exposure on the average fertility varies across reference groups. As a consequence, the baseline model with

[35] We checked that the difference with the baseline estimate of 0.63 is driven by the new group definition, not by the new sample restrictions.

homogeneous θ is misspecified. However, Lalive and Cattaneo (2009) formally show that the IV strategy gives a consistent estimator of θ^M. The intuition is that the specification error introduces a bias in the estimation of $\psi = \frac{\delta}{1-\theta'}$ in the first-stage (equation 8) and in the estimation of $\phi = \frac{\theta^M \delta}{1-\theta'}$ in the reduced-form (equation 7). The IV estimator identifies θ^M from the ratio of the two and at that stage, the biases cancel out.

To sum up, if we allow Han Chinese and minority Chinese to respond differently to a change in the group average fertility, we are able to identify the minority parameter with the IV strategy reported in Table 7, column (3). We get an estimate of $\theta^M = 0.63$.

B Different parameters for intra- and inter-ethnicity spillovers

To introduce heterogeneous externalities in a partial-population experiment, Arduini et al. (2020) replace $\theta\bar{y}$ with different averages for eligible and non-eligible individuals, allowing each person to weigh differently other group members depending on their eligibility status. The idea is that people sharing the same eligibility status might form a subgroup within the reference group. They propose a set of empirical IVs to separately identify within subgroup spillovers (from individuals with the same status) and between subgroup spillovers (from individuals with the opposite status).

Applying their methodology to our context, where the Han Chinese are "eligible" for birth quotas whereas minority Chinese are not, we rewrite the model as follows:

$$\begin{aligned}
y_i^M =& \alpha + \beta X_i^M + \gamma \bar{X}_{rdc(i)} + \lambda V_{rdc(i)} + \theta^b s^H_{rd(i)} \bar{y}^H_{rdc(i)} + \theta^w (1 - s^H_{rd(i)}) \bar{y}^M_{rdc(i)} + \varepsilon_i^M \\
y_i^H =& \alpha + \beta X_i^H + \gamma \bar{X}_{rdc(i)} + \lambda V_{rdc(i)} + \theta^w s^H_{rd(i)} \bar{y}^H_{rdc(i)} + \theta^b (1 - s^H_{rd(i)}) \bar{y}^M_{rdc(i)} + \delta \mathrm{ER}_{rpc(i)} + \varepsilon_i^H
\end{aligned}$$

where θ^w is the within/intra parameter and θ^b is the between/inter parameter; $\bar{y}^H_{rdc}$ (resp. $\bar{y}^M_{rdc}$) is the average fertility of Han (resp. minority) women in reference group rdc. When the fertility of everyone in the reference group is reduced by one, a minority woman responds by reducing her fertility by $\theta^b s^H + \theta^w(1 - s^H)$. The response is an average of the inter- and intra-ethnicity parameters, weighted by the share of Han.

To identify θ^w and θ^b, we need instruments for the two endogenous variables: $s^H \bar{y}^H$ and $(1 - s^H)\bar{y}^M$. The key intuition is that only $\bar{y}^H$ is *directly* affected by family planning policies; $\bar{y}^M$ is affected through spillovers. Spillovers are proportional to the share of Han while the direct policy effect is not. To see this, consider the following reduced-form equations:

$$\bar{y}^M_{rdc} = \alpha' + \beta'\bar{X}_{rdc} + \lambda' V_{rdc} + \frac{\delta\theta^b}{f(\theta^w,\theta^b,s^H_{rd})} s^H_{rd}\mathrm{ER}_{rpc}$$
$$\bar{y}^H_{rdc} = \alpha' + \beta'\bar{X}_{rdc} + \lambda' V_{rdc} + \frac{\delta}{1-\theta^w s^H_{rd}}\mathrm{ER}_{rpc} + \frac{\delta\theta^{b2}(1-s^H_{rd})}{f(\theta^w,\theta^b,s^H_{rd})(1-\theta^w s^H_{rd})} s^H_{rd}\mathrm{ER}_{rpc}$$

where $f(\theta^w,\theta^b,s^H_{rd}) = 1-\theta^w + s^H_{rd}(1-s^H_{rd})(\theta^{w2}-\theta^{b2})$.

To circumvent the non-linear equations, we rewrite them as infinite sums using the formula $\frac{1}{1-x} = \sum_{k=0}^{+\infty} x^k$ if $|x| < 1$. $\bar{y}^H$ can be expressed as a linear function of ER, $s^H ER$, $(s^H)^2 ER$ etc. and $\bar{y}^M$ can be expressed as a linear function of $s^H ER$, $(s^H)^2 ER$, $(s^H)^3 ER$ etc. The coefficients are combinations of δ, θ^w and θ^b. This intuition is formally developed and tested by Arduini et al. (2020). The key idea is to use the structural equations to come up with a set of empirical IVs exploiting higher order effects of the share of treated individuals. We use the following set: $\{s^H ER, (s^H)^2(1-s^H)ER, (1-s^H)s^H ER\}$ as instruments for $s^H\bar{y}^H$ and $(1-s^H)\bar{y}^{M(-i)}$ in the Minority equation. We include the same fixed effects and controls as in the baseline IV specification.

The results are reported in Table A.7. To be able to measure the average Han fertility and the leave-me-out average Minority fertility in each reference group, we exclude reference groups without Han Chinese and reference groups with only one minority Chinese (around 10% of the sample). We check in column (2) that this restricted sample gives us the same estimate of homogeneous θ as our baseline sample. Column (1) reports the estimates of $\theta^b = 0.63$ and $\theta^w = 0.71$, both significant at 1%. Both coefficients are not significantly different from each other and not significantly different from the homogeneous θ, suggesting that the heterogeneity is not strong enough to ruin the homogeneous model.

Still, an interesting takeaway is that within subgroup spillovers seem to be a bit stronger than between subgroup spillovers. In a given reference group, a single ethnicity often dominates the minority population; for instance, 75% of minority women belong to a reference group in which their own ethnicity accounts for more than half of all minority women. Therefore, another way to summarize the results is to write that minorities tend to respond more to changes in the fertility of other minority members, most of whom belong to the same ethnicity, than to changes in the Han fertility. This is consistent with the social channel described in Section 5: individuals put more weight on group members culturally closer to them.

The last question is how to deal with potentially heterogeneous θ^b within the minority population, depending on the ethnicity. To be fully flexible, we should allow each ethnic group to respond differently to the Han Chinese, to their own ethnic group members, and to other minority group members. We are not aware of any article discussing whether some of these parameters may be identified and the exercise is beyond the scope of our paper. We limit ourselves to noticing that, in the minority reduced-form equation, the coefficient on $s^H ER$ is unambiguously increasing with θ^b and equal to 0 if $\theta^b = 0$. We therefore conjecture that estimating separate reduced-form regressions for each ethnic group is informative about the existence and size of spillovers from the Han Chinese onto the members of this ethnic group.

Appendix A.5 Estimation of δ: robustness

Our initial strategy to estimate δ was to use the reduced-form equation for the Han Chinese: plugging equation (6) into equation (5), we get:

$$y_i^H = \frac{\alpha}{1-\theta} + \beta X_i^H + \frac{\theta\beta+\gamma}{1-\theta}\bar{X}_{rdc(i)} + \frac{\lambda}{1-\theta}V_{rdc(i)} + \frac{\theta\delta}{1-\theta}s^H_{rd(i)}\mathrm{ER}_{rpc(i)} + \delta\mathrm{ER}_{rpc(i)} + \varepsilon_i^H \quad (12)$$

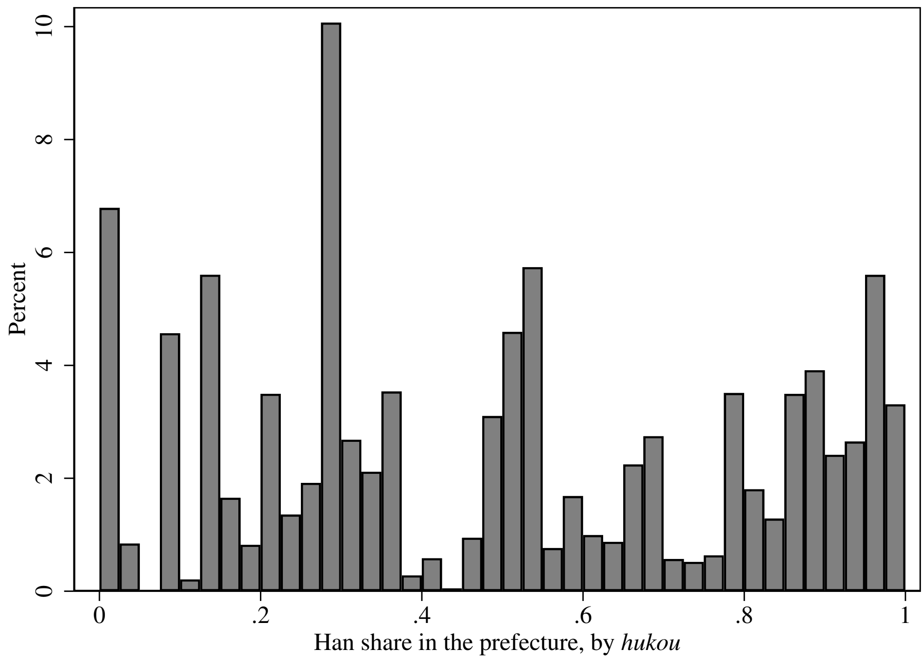
The coefficient on ER is δ. However, this specification leads to unstable and imprecise estimates. This is because, in the Han sample, there is little variable in s^H (mean=0.96, sd=0.10 compared to mean=0.47, sd=0.31 in the Minority sample). In particular, and by construction, there are very few Han observations and many Minority observations in reference groups with low Han shares. In this strategy, we therefore implicitly extrapolate the estimates to a part of the distribution ($s^H \to 0$) where we have very few observations. As a consequence, we overestimate the direct effect δ and underestimate the role of spillovers within the Han group.

This is the reason why we developed two alternative methods to estimate δ: (1) we pool the Han sample and the Minority sample and estimate the IV equation; and (2) we look at within-group variation. In (1), the identification still relies on the limit $s^H \to 0$ but we have more observations since we include the Minority Chinese. In (2), the identification is no longer at the limit; instead, we exploit a transformation where the social interaction term ($\theta\bar{y}$) cancels out. Strategy 2 should be more robust because all reference groups with at least one Minority Chinese and at least one Han Chinese contribute to the identification. However, the main drawback of strategy 2 is that we cannot identify θ. This is why strategy 1 is also important: we can check that the estimate of δ is the same as in strategy 2, and that the estimate of θ is the same as in the IV strategy with the Minority sample only.

Figure A.6 displays the estimates of δ using the IV strategy with the pooled sample (strategy 1), the within strategy (strategy 2) and the reduced-form strategy with the Han sample (naive strategy). We vary the restrictions on the sample by gradually excluding reference groups with the highest Han share. The legend shows the support of s^H in the restricted samples. The estimate on the extreme right of each graph corresponds to the unrestricted sample, where s^H ranges from 0 to 1. In theory, these restrictions should not matter since reference groups with few minorities are not useful to disentangle the direct effect and the indirect effect. As expected, the estimate of δ in strategy 2 is very precise and stable at -0.07. The estimate in strategy 1 is also stable around -0.07 but more imprecise. The estimate in the naive strategy is very imprecise and the magnitude ranges between -0.1 and -0.2 depending on the restrictions. We conclude that we should not take the magnitude of the point estimates in the Han sample at face value.

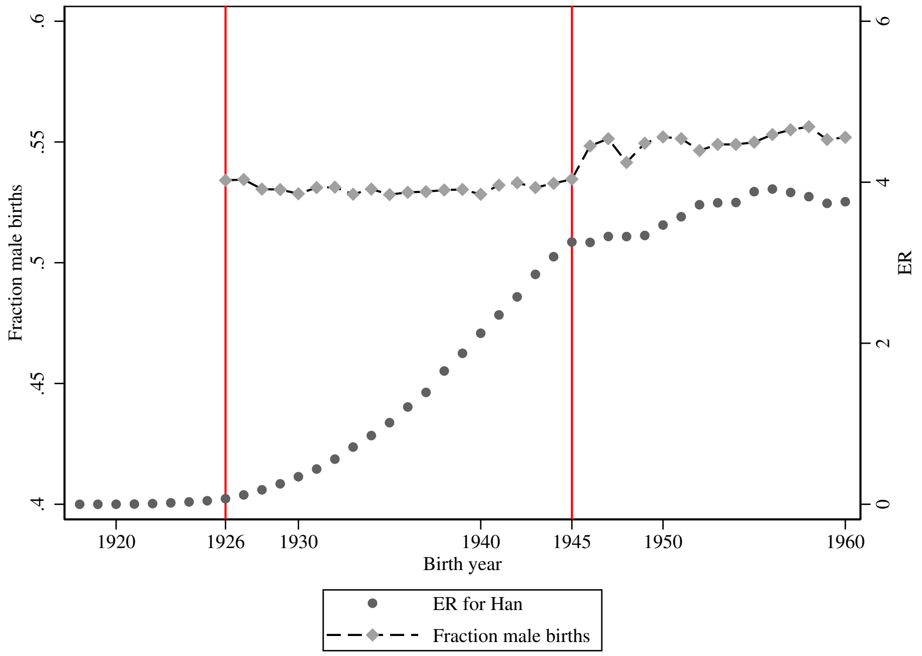
Appendix A.6 Appendix figures

Figure A.1: Distribution of Han share in the reference group of minority women

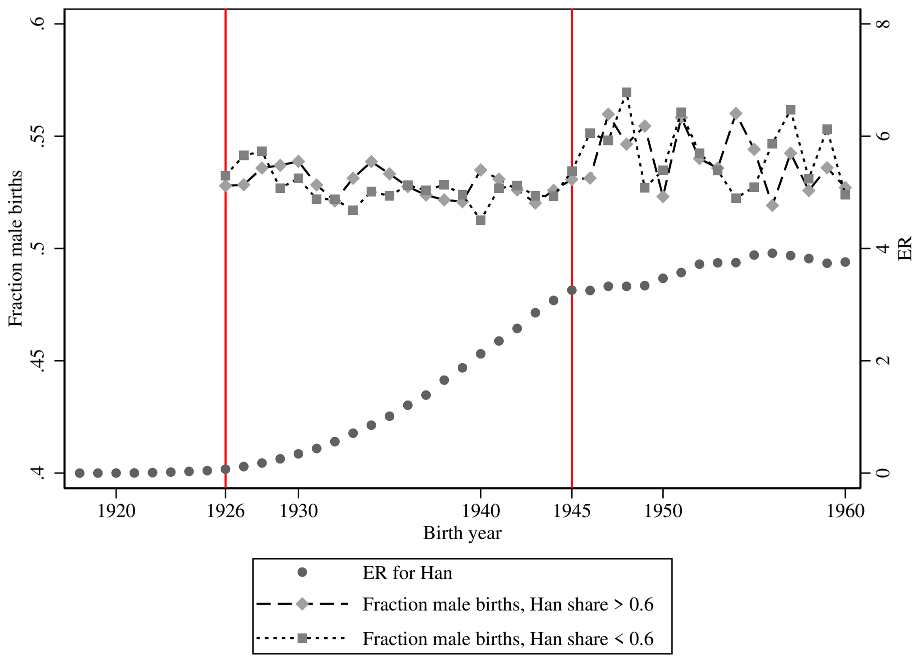


Note:
1. Authors' own calculation based on the 1% sample of the 1990 Census.
2. The graph is a histogram of s^H_{rd} in the Minority sample using bins of 2.5 percentage points.

Figure A.2: Sex ratio at birth and expected fertility reduction by ethnicity and by local Han share



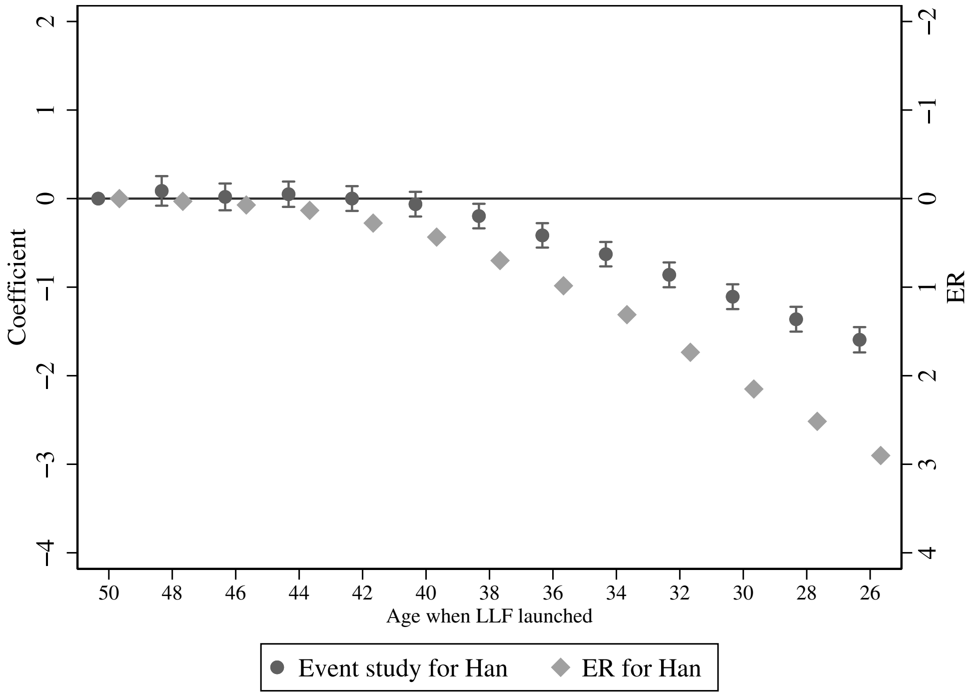
(a) Han



(b) Minority

Note: Sex ratio at birth is calculated based on the 1% sample of the 1982 and 1990 censuses and the 20% of the 2005 mini-census. The sample consists of women aged 45-64 in the censuses, who were born between 1918-1960. Sex ratio data are missing for mothers born before 1926 because the 1982 census does not report the number of male and female births separately. The econometric analysis of the spillovers focuses on cohorts between the two red vertical lines.

Figure A.3: Event study estimates for Han women

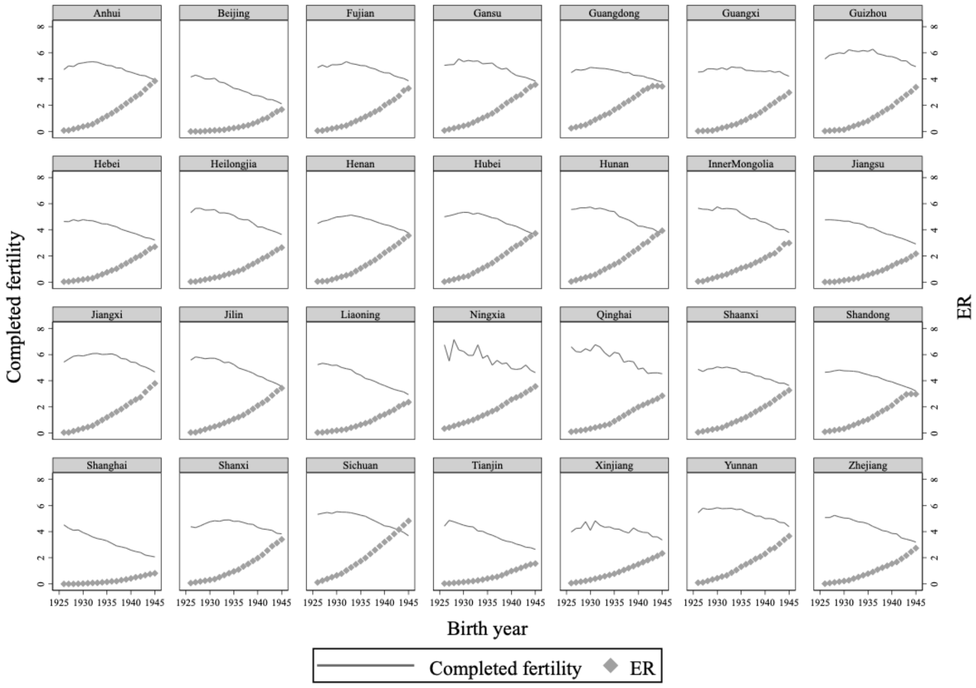


Note:
1. We use a sample of Han women aged between 26 and 51 years old when LLF was launched. Individual observations come from the 1% sample of the 1982 census (for cohorts 1918–1937) and 1990 census (for cohorts 1926–1945). We collapsed the data at the prefecture-*hukou*-cohort (rdc).
2. We estimate the following event study model:

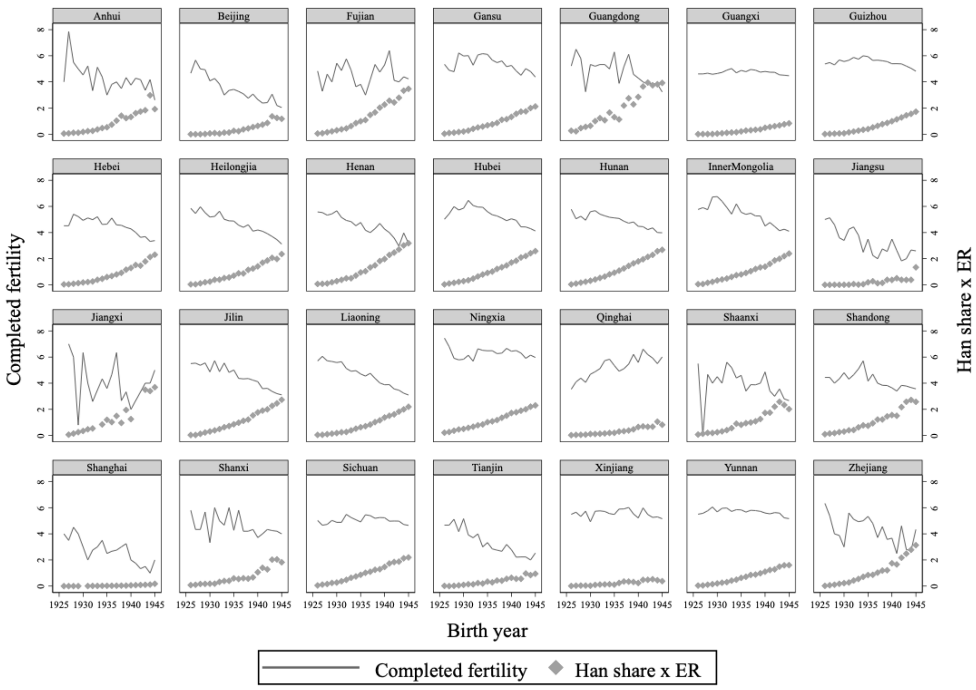
$$\bar{y}^H_{rdpc} = \sum_{j=13}^{25} \xi^*_j \mathbb{1}\{2j \leq (c - LLF_p) \leq 2j+1\} + \epsilon^H_{rdc}.$$

$(c - LLF_p)$ is the age of cohort c when LLF was launched in province p. We created bins of two years to have enough observations per age group. $\bar{y}^H_{rdpc}$ is the average completed fertility of Han women in group rdc.
3. The plot shows the event study estimates ξ^*_j together with the 95% confidence intervals. The omitted category is age group 50-51 (i.e., we impose that $\xi^*_{25} = 0$). The coefficient for age group $2j$ on the graph corresponds to ξ^*_j and measures the difference in average fertility between women aged $2j$-$2j+1$ and women aged 50-51 when LLF was launched.
4. Light gray diamonds represent the measure of policy exposure (ER) for the average Han woman in each age group (note that the y-axis is reversed). Standard event studies display a vertical line separating before and after periods. Our design is richer since we can predict the continuous intensity of exposure. The before period is formed by women aged 46 and over when LLF was launched ($ER = 0$). In the after period, exposure increases slowly until age 40 and more quickly for younger ages.

Figure A.4: Completed fertility and expected fertility reduction by province



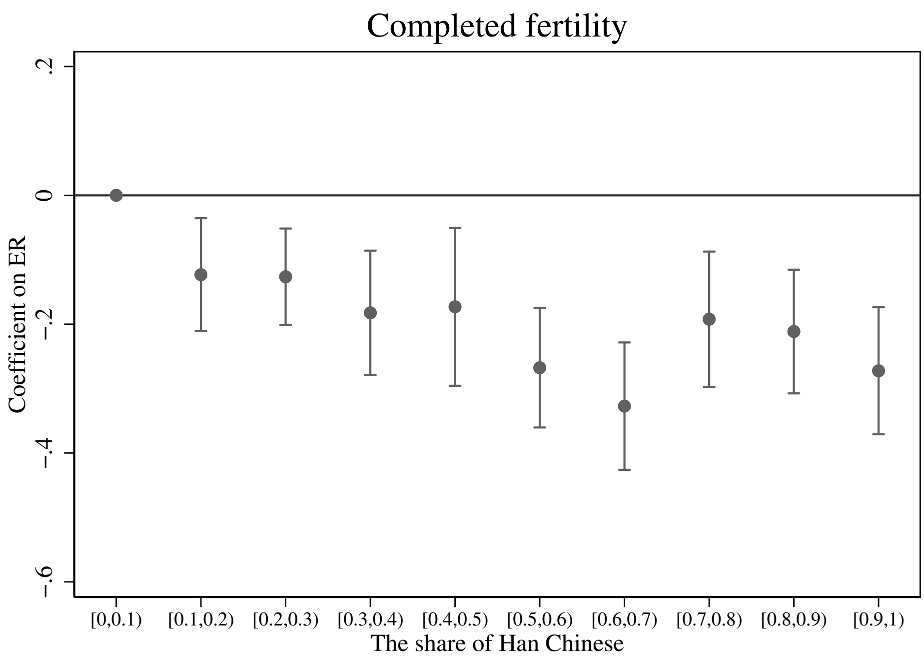
(a) Han



(b) Minority Chinese

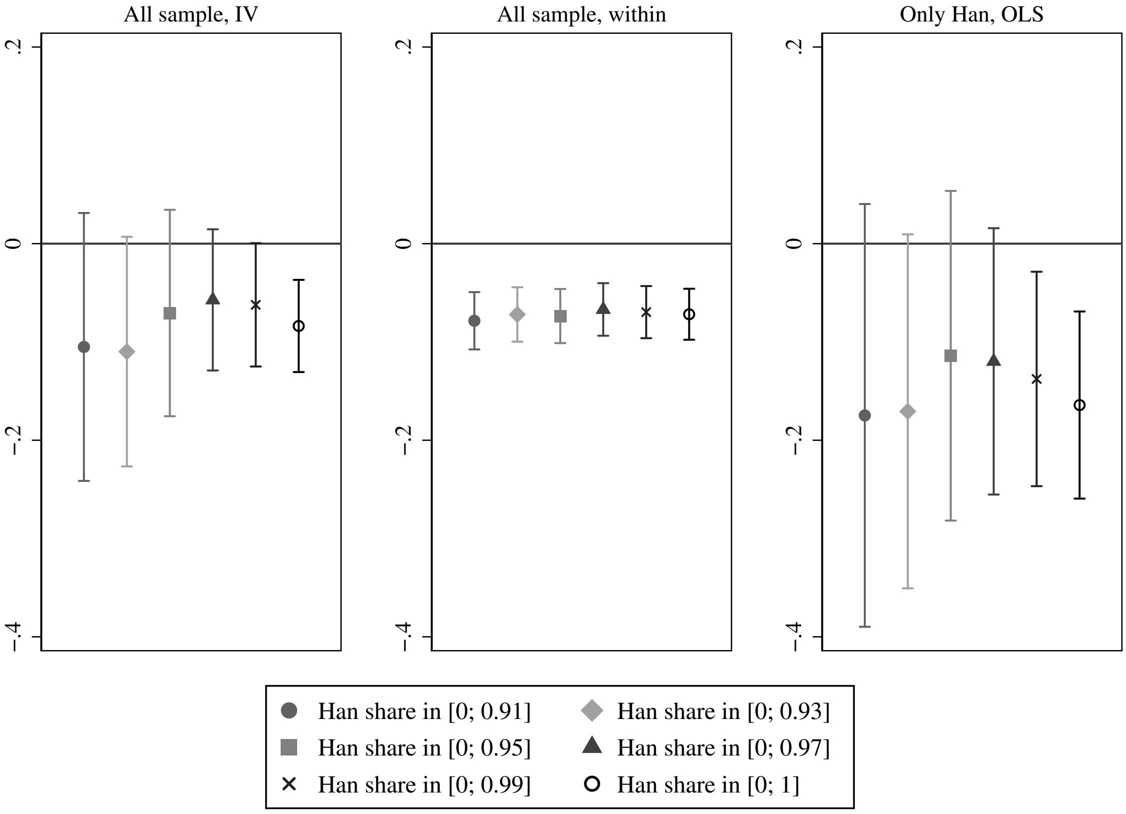
Note: Completed fertility is calculated based on the 1% sample of the 1990 census. Expected fertility reduction are constructed with provincial fertility data from Coale and Chen (1987).

Figure A.5: Reduced-form estimates of the indirect policy effect in a nonlinear specification



Note:
1. We use the same sample as column (3) of Table 4.
2. We replace s^H with dummy variables for each bin of 0.1 in Equation 2. The omitted category is the first bin. The plot shows the coefficients on interaction terms between each dummy and ER, together with the 95% confidence intervals.

Figure A.6: Estimation of δ: robustness to sample restrictions



Note: The three graphs shows the estimates of δ together with the 95% confidence intervals using IV strategy with the pooled sample (left), the within strategy with the pooled sample (middle), and the reduced-form OLS strategy with the Han sample (right). We vary the restrictions on the sample by gradually excluding reference groups with the highest Han share. The legend shows the support of s^H in the restricted samples.

Appendix A.7 Appendix tables

Table A.1: Robustness check: controlling for potentially heterogenous direct effects on minorities

Dep. var.: Completed fertility of minorities	(1) Direct effect	(2) Potentially heterogenous direct effect	(3)	(4)	(5)
Panel A: Reduced-form					
ER	-0.022	-0.021	-0.042	-0.381*	-0.412*
	(0.157)	(0.217)	(0.218)	(0.230)	(0.233)
s^H_{rd} × ER	-0.208***	-0.208***	-0.212***	-0.134***	-0.183***
	(0.043)	(0.044)	(0.044)	(0.044)	(0.067)
s^H_{rpc} × ER		-0.001	-0.080	-0.094	-0.082
		(0.191)	(0.193)	(0.189)	(0.189)
Urban *hukou* × ER			0.679***	0.829***	0.861***
			(0.192)	(0.206)	(0.208)
Illiterate × ER			-0.012	-0.030	-0.031
			(0.019)	(0.019)	(0.019)
High school × ER			0.043	0.050	0.050
			(0.051)	(0.052)	(0.052)
Share urban *hukou* × ER				-0.150**	-0.123*
				(0.066)	(0.071)
Share illiterate × ER				0.549***	0.581***
				(0.134)	(0.140)
Share high school × ER				0.364	0.329
				(0.601)	(0.603)
Second quintile × ER					0.015
					(0.027)
Third quintile × ER					0.047
					(0.034)
Fourth quintile × ER					0.062
					(0.051)
Fifth quintile × ER					-0.001
					(0.060)
Panel B: IV, instrument $s^H_{rd} \times ER$					
Group average fertility	0.629***	0.628***	0.644***	0.500***	0.464***
	(0.104)	(0.106)	(0.105)	(0.140)	(0.142)
ER	-0.016	-0.014	-0.025	-0.236	-0.243
	(0.109)	(0.139)	(0.138)	(0.168)	(0.174)
R^2	0.167	0.167	0.166	0.170	0.171
Mean dep var	5.075	5.075	5.075	5.075	5.075
Number of clusters	5514	5514	5514	5514	5514
F statistics	99.691	94.054	94.819	70.856	54.970
Observations	58887	58887	58887	58887	58887

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.
1. Each column represents a separate regression. Robust standard errors in parentheses are clustered at the prefecture-*hukou*-cohort level.
2. s^H_{rd} is the share of Han in the prefecture-*hukou* group. ER is interacted with s^H_{rpc}, the share of Han in the province-*hukou*-cohort group, in columns (2)-(5); with individual characteristics, including indicators of being a urban *hukou* holder, illiterate, and a high school graduate, in columns (3)-(5); with the shares of urban *hukou* holders, illiterate women and high school graduates at the prefecture-*hukou* level in columns (4) - (5); and with the rank (in quintiles) of the prefecture-*hukou* within the province in terms of number of Han Chinese.
3. In Panel A, column (1) is the same as column (3) of Table 4. Columns (2)-(5) show the coefficients on additional controls mentioned in Table 4 columns (4)-(7).
4. In Panel B, all columns use the same baseline specification and sample as column (3) of Table 7. Coefficients on interaction terms are not reported due to space constraints.

Table A.2: Common trends: controlling for potential confounding shocks

	(1) Dep. var.: completed fertility of minorities
ER	0.064
	(0.168)
s^H_{rd} × ER	-0.203***
	(0.044)
Urban × Famine [a]	-0.016
	(0.027)
Rural × Famine	0.005
	(0.018)
Urban × Send-down [b]	0.150***
	(0.039)
Rural × Send-down	-0.052*
	(0.028)
Urban × Cultural Revolution [c]	0.426
	(0.348)
Rural × Cultural Revolution	0.037
	(0.251)
R^2	0.173
Number of clusters	5514
Mean dep var	5.075
Observations	58887

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%. The sample and econometric specifications are the same as in Table 4 column (8), with the coefficients of additional controls displayed. ER is the expected fertility reduction to meet the birth quota set by family planning policies.

[a] The Famine variable is constructed as follows:

$$\text{Famine}_{rdpc} = \sum_{a=15}^{34} AFR^{1958}(a) \times CohortLoss_{d,c+a} \times I[1958 \leq c+a \leq 1963]$$

Where AFR^{1958} is the national-level age-specific fertility rate. $CohortLoss_{p,c+a}$ is an index measuring the intensity of the famine in year $c+a$ in prefecture d. To construct this index, we follow Chen and Yang (2016) and estimate a rural population linear trend in non-famine times fitting the sizes of cohorts 1952–1954 and 1964–1966. Next, we use the estimated trend to project counterfactual cohort sizes for 1958–1963 cohorts (one year before and two years after the famine to account for anticipation and compensation). The index is equal to one minus the ratio between actual and projected cohort sizes.

[b] The Send-down variable is constructed as follows:

$$\text{Send-down}_{pc} = \sum_{a=15}^{34} AFR^{1958}(a) \times \frac{SDY_{p,c+a}}{Pop_p}$$

Where $SDY_{p,c+a}$ is the number of sent-down youths settled in year $c+a$ in province p. Pop_p is the provincial population in thousands of people just before the cultural revolution (1966). Data on SDY are taken from Gu (1997).

[c] The Cultural Revolution variable is constructed as follows:

$$\text{Cultural Revolution}_{pc} = \sum_{a=15}^{34} AFR^{1958}(a) \times \frac{Victims_{p,c+a}}{Pop_p}$$

Where $Victims_{p,c+a}$ is the number of political victims in year $c+a$ in province p. Data on $Victims$ are aggregated from the number of fatalities in political events recorded by Walder and Lu (2017).

Table A.3: Parallel pre-trends: Placebo test using older cohorts

	(1)	(2)
	Dep. var.: completed fertility of minorities	
Placebo exposure	0.178	0.163
	(0.420)	(0.454)
Han share × Placebo exposure		0.022
		(0.111)
R^2	0.119	0.119
Number of clusters	595	3122
Mean dep var	5.740	5.740
Observations	20370	20370

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.

1. Each column represents a separate regression. Robust standard errors in parentheses are clustered at the prefecture-*hukou*-cohort level.

2. The analysis is based on a sample of minority women born in 1918–1929 from the 1% sample of the 1982 Chinese Census data. We construct the placebo exposure for these women as if they were born 16 years later, so they get the ER of cohorts 1934–1945. To be consistent with the baseline specification, we also construct placebo provincial characteristics at age 25. The share of Han at the prefecture-*hukou* level and individual characteristics are based on the true information of these women.

3. Econometric specifications are the same as in Table 4. Column (1) shows the estimation corresponding to column (2) of Table 4. Column (2) shows the estimation corresponding to column (3) of Table 4.

Table A.4: Spillover effects, by language and residential distances

Dep. var: Completed fertility of minorities	(1)
Other written language $\times T_1 \times$ Han share $\times$ Exposure	-0.053
	(0.216)
Other written language $\times T_2 \times$ Han share $\times$ Exposure	-0.005
	(0.079)
Other written language $\times T_3 \times$ Han share $\times$ Exposure	-0.029
	(0.055)
Han script $\times T_1 \times$ Han share $\times$ Exposure	-0.109
	(0.071)
Han script $\times T_2 \times$ Han share $\times$ Exposure	-0.228***
	(0.050)
Han script $\times T_3 \times$ Han share $\times$ Exposure	-0.364***
	(0.047)
R^2	0.174
Number of clusters	5514
Observations	58887

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.

1. Each column represents a separate regression using the same specification and sample as column (3) of Table 4. Robust standard errors in parentheses are clustered at the prefecture-*hukou*-cohort level.

2. Each row represents the coefficient on s^H_{rd}ER interacted with a dummy for each category of ethnic groups defined by (1) common language and (2) residential integration. The first component is a dummy variable equal to one if the group uses the Han Chinese script as their written language. The second component measures the share of ethnic group members living *outside* official minority autonomous regions or prefectures in 1990. A higher share indicates more residential integration. We classify the groups into six categories: Han script (0/1) $\times$ terciles of the residential integration distribution. The terciles are created for groups using and not using Han script separately, to have similar number of observations in each tercile. The terciles are as follows for groups using own written languages: $T_1 = [0; 0.05], T_2 = (0.05; 0.39], T_3 = (0.39; 1]$. The terciles are as follows for groups using Han script: $T_1 = [0; 0.2], T_2 = (0.2; 0.6], T_3 = (0.6; 1]$. Note that the dichotomous variable in Table 6 puts together groups using own written language or belonging to T_1 as weak cultural integration and groups using Han script and belonging to T_2 or T_3 as strong cultural integration.

Table A.5: IV approach: robustness checks

	(1)	(2)	(3)	(4)
	Excl. units affected by reclassification		Alternative reference groupls: three cohorts in	
Dep. var.: completed fertility of minorities	Prefectures	Ethnic groups	Prefecture-*hukou*	County-*hukou*
Group average fertility ($\bar{y}^{(-i)}$)	0.610***	0.500***	0.629***	0.736**
	(0.108)	(0.115)	(0.138)	(0.340)
R^2	0.169	0.165	0.179	0.222
Number of clusters	5077	4500	5063	8074
F statistics	95.285	88.785	189.937	25.177
Observations	50534	42556	55049	41813

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.
1. Each column represents a separate regression. Robust standard errors in parentheses are clustered at the prefecture-*hukou*-cohort level.
2. All columns use the same specification as column (3) of Table 7.
3. In column (1), 12 prefectures where the share of Han changed by more than 20 percentage points from 1982 to 1990 (in Census data) are excluded from the sample. In column (2), ethnic groups documented as the most affected by the reclassification wave (Manchu, Tujia, Miao, Dong, Yilao and Qiang) are excluded from the sample.
4. A women's reference group is defined as all women, holding the same *hukou*, born in the same year or 1 or 2 years before and living in the same prefecture (in column 3) or county (in column 4). Note that we exclude cohorts 1926-1927 because we have no information on the completed fertility of cohorts born 1 and 2 years before them in the 1990 census. In column (4), we also exclude outliers in terms of group size and keep only reference groups with more than 20 observations.

Table A.6: Reduced-form approach: specification checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep. var.: completed fertility of minorities							
	Estimations with and without linear trends						Alternative measures	
Sample:	All provinces		Low Han provinces		High Han provinces		All	
Han share × Exposure ($s^H \times ER$)	-0.208***	-0.422***	-0.206***	-0.222***	-0.173**	-0.179**		
	(0.043)	(0.035)	(0.053)	(0.052)	(0.077)	(0.072)		
Han share × ER^{nat}							-0.255***	
							(0.051)	
$\mathbb{1}\{$Han share $\geq 0.47\} \times \mathbb{1}\{$ER $\geq 2\}$								-0.281***
								(0.050)
Province-*hukou* linear trend	Yes	No	Yes	No	Yes	No	Yes	Yes
R^2	0.173	0.168	0.105	0.104	0.260	0.255	0.173	0.173
Number of clusters	5514	5514	1392	1392	4122	4122	5514	5514
Mean dep var	5.075	5.075	5.340	5.340	4.704	4.704	5.075	5.075
Observations	58887	58887	34374	34374	24513	24513	58887	58887

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.

1. Each column in each panel represents a separate regression. Robust standard errors in parentheses are clustered at the prefecture-*hukou*-cohort level.
2. We control for ER in all regressions and use the same specification as in column (3) of Table 4, except that in columns (2), (4), and (6) province-*hukou* linear trends are removed.
3. We use the same sample as in column (3) of Table 4 in columns (1), (2), (7), and (8). We split the sample into province-*hukou* groups where the share of Han is below 70% in columns (3)-(4) and above 70% in columns (5)-(6). The share of Han is below 70% in rural Guangxi, Guizhou, Ningxia, Qinghai and Yunnan, as well as in Xinjiang. Note that the Han share used to split the sample is measured at the province-*hukou* level whereas the Han share in the regression (s^H) is measured at the prefecture-*hukou* level.
4. In column (7) , ER^{nat} is an alternative measure of policy exposure based on national (instead of provincial) age-specific fertility rates (AFR) in 1969. ER^{nat} is constructed at the province-cohort level; for a given cohort, the expected reduction to comply with birth quotas varies across provinces only because women were of different ages when LLF was launched and not because women had different initial fertility levels.
5. In column (8), we replace the continuous variable ER by $\mathbb{1}\{ER \geq 2\}$, a dummy equal to one if $ER \geq 2$ and zero otherwise. We also replace the continuous variable s^H by $\mathbb{1}\{s^H \geq 0.49\}$, a dummy equal to one if the local Han share is above the average (49%) and zero otherwise. We use the Stata package provided by de Chaisemartin and D'Haultfoeuille (2020) to assess the importance of negative weights in the case of heterogeneous treatment effects: the share of negative weights is equal to 10% and the sum of negative weights is equal to -0.016.

Table A.7: Estimation of intra- and inter-ethnicity spillovers

Dep. var.: completed fertility of minorities	(1) Heterogenous model	(2) Homogenous model
Han share × Han average fertility ($s^H \bar{y}^H$)	0.627***	
	(0.105)	
Minority share × Minority average fertility ($(1-s^H)\bar{y}^{M(-i)}$)	0.713***	
	(0.143)	
Group average fertility ($\bar{y}^{(-i)}$)		0.617***
		(0.146)
F statistics	22.384	66.368
R^2	0.160	0.161
Number of clusters	3747	3747
Observations	53829	53829

Note: * means significant at 10%, ** significant at 5%, and *** significant at 1%.

1. Each column represents a separate regression. Robust standard errors in parentheses are clustered at the reference group level. The analysis is based on the same sample as Table 7, column (3), except that we exclude reference groups without Han Chinese and reference groups with only one minority Chinese (to be able to compute $\bar{y}^H$ and $\bar{y}^{M(-i)}$).

2. $s^H \bar{y}^H$ is the product of the Han share and the average fertility of Han women in the reference group of minority woman i. $(1-s^H)\bar{y}^{M(-i)}$ is the product of the minority share and the average fertility of minority women, excluding i, in the reference group of minority woman i. $\bar{y}^{(-i)}$ is the average fertility of all women, excluding i, in the reference group of minority woman i.

3. In column (1), we use the set of empirical IVs $\{s^H ER, (s^H)^2(1-s^H)ER, (1-s^H)s^H ER\}$ as instruments for $s^H \bar{y}^H$ and $(1-s^H)\bar{y}^{M(-i)}$ to recover an estimate of the inter-ethnicity parameter θ^b (coefficient on $s^H \bar{y}^H$) and an estimate of the intra-ethnicity parameter θ^w (coefficient on $(1-s^H)\bar{y}^{M(-i)}$). We include the same fixed effects and controls as in Table 7, column (3).

4. In column (2), we use the same specification as in Table 7, column (3): we instrument $\bar{y}^{(-i)}$ with $s^H ER$ to recover an estimate of the homogeneous parameter θ.