# Children or work? The impact of fertility on old-age labour supply *

Ziwei Rao†

October 28, 2021

**Abstract**

This study provides empirical evidence of the causal impact of fertility outcomes on old-age labor supply by employing an instrumental variable method. The results show that having more children and relatively more sons prevents parents from strenuous works at senior ages. Rural parents with one more child (son) significantly reduce their probability of working at post-retirement age by 17.7 (19.8) percentage points. The impact of children is heterogeneous and is especially strong among the more vulnerable groups with worse health and little pension benefits. Further analysis of child material support finds a positive impact of family sizes on total child support and complements the labor supply findings. Results in this paper suggest that population policies that aim at reducing fertility might jeopardize the elderly well-being by compelling old parents in bad health to continue working.
***JEL Classification:*** *I14, I18, J13, J26*
***Keywords:*** *fertility, elderly labor supply, old-age support, health, pension*

*I am deeply grateful to my supervisors Rob Alessie and Max Groneck, for their patient guidance and valuable comments. I thank participants at the brownbag seminar and the PhD conference at University of Groningen, the annual conferences of the European Association of Labour Economists (EALE), the German Economic Association (VfS), and the Pension Day of Network for Studies on Pensions, Aging and Retirement (Netspar) for their insightful discussions. I also thank the China Center for Economic Research, the National School of Development of Peking University for providing the data.

†Ziwei Rao, Department of Economics, Econometrics and Finance, University of Groningen, The Netherlands (z.rao@rug.nl).

# 1 Introduction

Many elderly from developing countries keep working until they are physically incapable due to insufficient social security benefits. It is also often infeasible for them to accumulate sufficient savings or assets given their lifetime poverty. Consequently, own labor income and children's support become the only sources of old-age income for these elderly. The proportion of older people who rely primarily on either one of the two sources amounts to 70 percent among the Chinese 60+ population. Meanwhile, working at old age is prevalent in China, especially among rural residents who engage in strenuous farm work. The labor force participation rate reaches 17% among rural elderly who aged 75 and above, in contrast to the life expectancy of 74 years in 2010 according to the World Bank.[1]

In order to examine if children can prevent potentially vulnerable old parents from working into a senior age, this study investigates the impact of child quantity and child sex composition on parents' old-age labor supply decisions. This paper adopts a two-stage least square method to derive causal inferences and uses data from the China Health and Retirement Longitudinal Study (CHARLS). Family planning policies in the early 1970s and the sex of the firstborn child are exploited as instruments for family fertility outcomes. Heterogeneity across differential health statuses and pension benefit levels of the parent are further investigated to show how the importance of children varies according to parental needs.

Children play an undoubtedly crucial role in providing old-age security to elder parents in China. Children, especially sons given the patrilineal tradition, have long been considered as responsible support providers (Whyte, 2003). It is well reflected by the 2013 China Household Finance Survey, where more than 50% of respondents explicitly chose "for old age support" as one reason for having children. Children's duty of supporting and caring for their parents is also written into the Chinese constitution and law. Apart from economic reasons, filial piety, a well-accepted essential virtue in the culture of many Asian countries, also requires children to provide material support and instrumental care to parents in a manner that conveys respect (Sung, 1998).

Although the elderly rely heavily on transfers from descendants, the traditional family support pattern is experiencing adjustments under demographic challenges. Dramatic shrinkage in family sizes has been witnessed during the past decades, starting from the early 1970s when a series of population control policies were introduced. The total fertility rate dropped from 5.7 births per woman in 1970 to 1.7 in 2018. In the meantime, the population is aging rapidly. The share of the population aged 60 and older is projected to rise to 34.6% in 2050. As a result of decreased birth rates and longer life expectancy, parents require longer supports from fewer potential caregivers.

Theoretically, parents would adopt self-insuring strategies to sustain the desired standard of living in older age when the adequacy of child support is potentially threatened. People could increase asset holdings by themselves through the life course or postpone entry into retirement. Abundant literature links the fertility to savings (Caldwell, 1978; Boldrin & Jones, 2002; Tertilt, 2005), and the one-child policy has been argued to contribute to China's high saving rates although empirical results are mixed (Song, Coupé, & Reed, 2020; Ge, Yang, & Zhang, 2018; Modigliani & Cao, 2004; Huang, Lei, & Sun, 2016; Rosenzweig & Zhang, 2014; Banerjee, Meng, Porzio, & Qian, 2014).

[1]Labor force participation and old-age support information are obtained from the 2010 Chinese population census.

However, research on the impact of family sizes on staying in the labor force at old age is scarce, despite that old-age labor supply and child support are the two major pillars of support for many elderly. Most labor studies focus on the impact of fertility on labor supply decisions of young or middle-aged parents, especially mothers, instead of exploring old-age labor supply behaviors(Angrist & Evans, 1998; Agüero & Marks, 2011; He & Zhu, 2016; Guo, Li, Yi, & Zhang, 2018).[2]

This paper thus aims to fill the gap by explaining the high elderly labor participation and providing the first thorough empirical analysis on the causal impact of fertility on old-age labor supply decisions. Understanding the role of children in preventing old and sick parents from everlasting work provides crucial implications to many aging societies with immature public security systems. Many countries have experienced stronger or weaker family planning policies and a sharp drop in fertility during the past decades, e.g., Vietnam, South Korea, Thailand, Mexico, Columbia. The unprepared generation impacted the most by the family size shrinkage might need more public support to complement the still important family-supporting system.

Results from the two-stage least squares regressions find that having more children, especially sons, significantly reduces rural parents' incidence of participating in the labor market in old age. The probability of working decreases by 17.7 percentage points (pp) if the total number of alive children increases by one among rural parents at post-retirement ages. Having more sons relative to daughters also decreases rural parents' chance of working. For instance, in the extreme case where families change from having all daughters to having all sons (son ratio increases by one), ceteris paribus, the rural elderly are 23.8 pp more likely to drop out of the labor market.

The pattern above is heterogeneous depending on parents' needs. The impact of child quantity is more substantial among parents aged between 65 and 74, with worse health conditions and little pension benefits. Having one more child decreases the probability of working by 14.7 pp among parents with 3+ (I)ADL limitations, which more than doubled from 6.2 pp among parents with no (I)ADL. Among parents with no pension at all, the incidence of working is decreased by 24.1 pp when there is one additional child in the family.

Studying the mechanism of child support, I find supporting evidence that parents with one more child are 6 pp more likely to receive any material transfer from children, and they receive 28% higher amount conditional on receiving. Further analysis of lifetime labor supply suggests greater strains might result from longer working hours, even overwork, throughout the lifetime in addition to prolonged working life in old age.

The paper is organized as follows: Section 2 discusses the institutional background in China regarding the old-age support sources and the evolution of population policies. Section 3 describes the data and variables used in the study, and presents stylized descriptive statistics. The empirical strategy, including fertility instruments and model specification, is introduced in Section 4. Section 5 presents the main results and explores heterogeneity. Section 6 discusses the potential mechanism and provides a set of sensitivity analyses. Section 7 concludes.

[2] Oliveira (2016) is the only exception who uses first born twins to examine the probability of working past retirement age as one of the measures of parental wellbeing. However, old-age labor supply decisions are not investigated in detail, and the heterogeneity is not discussed.

# 2 Institutional Background

## 2.1 Old-age Support in China

Though they often offer inadequate social security, public pension schemes provide the broadest coverage compared to other annuity programs in China. Various public pension schemes target different population and workforce groups, and their financing, contribution, and benefit rules vary substantially. In general, a participation history of 15 years or more is required to be entitled to pension benefits, and eligible beneficiaries start to receive pension benefits after reaching mandatory retirement ages, regardless of their work status. [3]

The various schemes can be generally classified into employee pension and resident pension, where the former targets public and private sector employees, and the latter targets rural and urban residents without a formal non-agricultural job. Compared to employee pension schemes, resident pension schemes were introduced much later, and they are much less generous in terms of benefit levels. Only recently, the public pension has been expanded to cover the vast rural population in agriculture.[4]

Despite recent expansion in public pension coverage, there is significant heterogeneity in pension generosity across pension schemes and regions. In 2017, the average monthly pension of BOAI is 2,870 RMB (1000 RMB yuan equals approx. 150USD), which amounts to an average replacement ratio of 46%. In contrast, most of the RPS participants receive a flat-rate benefit from the basic pension, whose monthly average is less than 5% of the BOAI pension level, and the benefit amounts to only 7-20% of income per capita, depending on the region (Fang & Feng, 2018). Although pension benefits supplement old-age income, the low pension benefits make resident pension schemes inadequate in providing income security, especially for the vast rural elderly.

To overcome the lack of public support, the Chinese elderly, especially those in rural areas, either work into old age or rely on intergenerational child support. As shown in Table 1, at the national level, 29% of the elderly at post-retirement age primarily rely on their own labor income for old-age support, and 41% rely on family support. Many urban elderly enjoy pension benefits, whereas much more rural counterparts obtain support through their labor income. It is also evident that when labor supply decreases as they step into the 70s, they turn to family support. For instance, the proportion of rural elderly who primarily rely on their own labor income decreases from 59.8% to 17.3%, whereas those who rely mainly on family support increase from 31.0% to 69.3% correspondingly.

Apart from economic reasons, Confucian filial piety and social norms embedded in the culture shape inter-generational relations. Filial piety (xiao) is considered an essential virtue in Chinese and other Asian countries' cultures. Xiao, as an unconditional obligation of children, commands children to be obedient and respectful, and it is demonstrated

[3]The BOAI and PEP retirement eligibility age is 60 for males, 55 for female white-collar workers, and 50 for female blue-collar workers. The pension eligibility age for the NRP, URP, and RPS is 60 for both males and females. The PEP does not require any contribution from public employees. Residents who were already 60 when the NRP and URP were introduced could receive flat-rate basic pension benefits without contribution if they opt to enroll.

[4]Established in the 1950s, Public Employee Pension (PEP) applies to public sector employees, and the Basic Old Age Insurance (BOAI) applies to urban private sector employees. New Rural Resident Pension (NRP) and Urban Resident Pension (URP) were introduced in 2009 and 2011 as voluntary schemes for rural and urban residents without a formal non-agricultural job. The former two were merged into one employee pension scheme BOAI in 2015, and the latter two were merged into a uniform Resident Pension Scheme (RPS) in 2014.

Table 1: **Primary source of support 60+**

| Age | Labor income | Pension | Family support | Others |
|---|---|---|---|---|
| *Rural* | | | | |
| 60-69 | 59,8% | 4,4% | 31,0% | 4,8% |
| 70+ | 17,3% | 4,8% | 69,3% | 8,7% |
| *Urban* | | | | |
| 60-69 | 19,6% | 49,7% | 25,8% | 5,0% |
| 70+ | 4,4% | 50,7% | 38,5% | 6,3% |
| Total | 29,1% | 24,1% | 40,7% | 6,1% |

*Source:* National Bureau of Statistics of China (NBSC) 2010 census. Respondents could choose one primary source from labor income, pension, family support, unemployment insurance, basic living allowance, property income and others. Latter categories are further merged into one.

when children reciprocate parents' care (Ho, 1994; Kwan, 2000). In practice, it requires children to provide material support and instrumental care to parents in a manner that conveys respect (Sung, 1998).

Children's duty of supporting and caring for their parents is also written into the Chinese constitution and law. Given the patriarchal social system and the patrilineal tradition, many parents consider children, especially sons, as responsible old-age support providers. For instance, rural families without sons are more likely to participate in pension programs and to save more for old age support (Ebenstein & Leung, 2010).

## 2.2 Family Planning Policies in China

Before introducing OCP in 1979, family planning policies were already in place and had evolved for decades. The development of China's official party's family planning experienced back-and-forth between pro-natal and anti-natal phases. The debate in general shifted to consensus on population control during the middle 1950s to curb the high fertility rates to facilitate economic growth. However, political instability and major social events, including Great Famine between 1959 and 1961, Great Leap Forward during 1955 to 1960, and Cultural Revolution starting from 1966, jointly disrupted family planning movements.

Only at the beginning of the 1970s, anti-natal family planning movements were first formalized into a series of policies and expanded to rural areas at the national level.[5] This series of policies were later incorporated into one slogan – "Later, longer and fewer" (LLF), as they promoted later marriage, longer birth intervals and fewer births. Later marriage denotes delaying marriage to age 23 for women and 25 for men; longer birth intervals require birth spacing to be longer than three years; and fewer births urge urban couples to have no more than two children and rural couples no more than three (Greenhalgh, 2008; White, 2009). The family planning policy gradually intensified during the decade and eventually evolved into the one-child policy at the end of the 1970s.

[5] Earlier birth-planning campaigns were small, and they targeted at urban residents with weaker restrictions and enforcement (White, 2009; Scharping, 2013; Wang, 2016).

The most dramatic decline in total fertility rate (TFR) occurred during the early 1970s, which coincides with the implementation period of LLF. As shown in Figure A1, during the LLF rolling-out period, the TFR decreased from around six children per woman in 1970 to below three children in 1978. After 1979, the TFR continued declining whereas at a lower rate.

Growing literature attributes the fertility decline following OCP to socioeconomic development since the economic reform in 1978 and finds limited impact of OCP on lifetime fertility in the long term (Cai, 2010; Whyte, Feng, & Cai, 2015; Wang, 2016; Wang, Zhao, & Zhao, 2017). Instead, LLF policy is considered to be the main reason for the fertility drop during the 1970s, when economic development was almost stagnant by the time (Mauldin, 1982; Whyte et al., 2015; Zhang, 2017). Recent studies examining the causal impact of LLF on fertility also find supporting evidence (Wang, 2016; Y. Chen & Fang, 2019; Babiarz, Ma, Miller, & Song, 2019; Y. Chen & Huang, 2020).

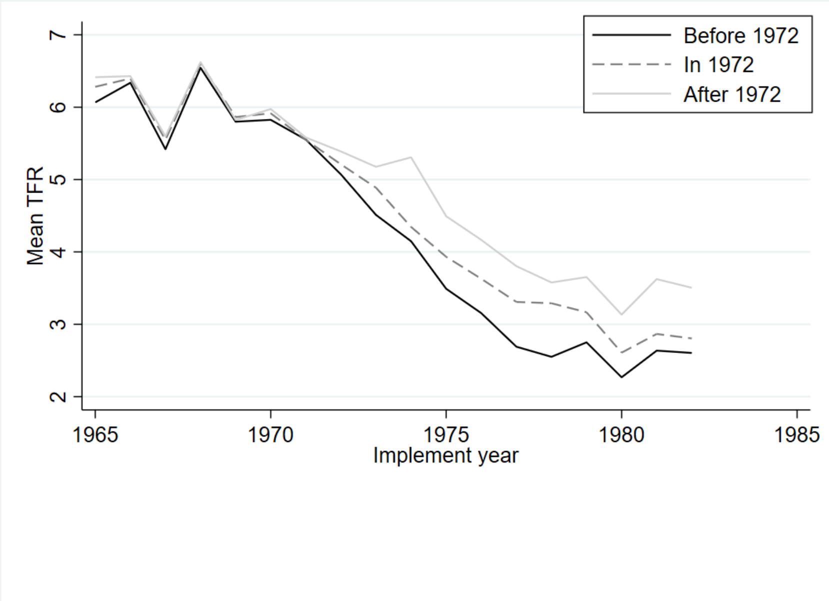
In practice, local governments set up Family Planning Leading Groups to respond to the state's calling for population control and took over the organization and implementation of LLF policies from the central level. Leading groups were established between 1969 and 1975 across provinces to facilitate local implementation of LLF, and the staggered roll-out created variation in the timing of the policy enforcement (Y. Chen & Huang, 2020).

To illustrate how differential policy enforcement time affects the decline of TFR, I categorize provinces into early, reference, and late groups based on their leading group establishment year.[6] The TFR trend of the three groups is displayed in Figure 1, where the decline of TFR follows the establishment of leading groups. Provinces that established the leading groups before 1972 – the early group– witnessed an immediate decrease in the average TFR starting from around 1970. Whereas the average TFR of provinces in the late group started to drop rapidly with a similar rate as the early group did only after 1974. The graph suggests that the differential timing of LLF implementation plausibly explains the cross-province difference in fertility drop.

Two sources of variation of the LLF policy are exploited in this paper, including the variable timing and the differential intensity of enforcement across provinces. The aforementioned varying timing implies that, at the micro-level, women of the same age who live in provinces with earlier enforcement were exposed longer to the LLF policy. Thus their fertility decisions were more affected. Additionally, pre-existing fertility trends contribute to the differential enforcement intensity of the policy. Before the LLF policy was implemented, there was considerable heterogeneity in fertility rates across regions. In 1969, the national TFR was 5.67, while the TFR in Beijing was 3.57 in contrast to a TFR of 6.71 in Jiangxi province. Since the LLF policy advocated no more than two (three) children for urban (rural) families homogeneously to the whole country, families in provinces with higher levels of pre-existing fertility rates were exposed heavier to the policy.

[6] Establishment year data are collected from province population chronicles, see more details and the original documentation in Y. Chen and Huang (2020). The early group consists of 9 provinces that established leading groups in 1969, 1970 or 1971, the reference group contains nine provinces that established leading groups in 1972, and the late group has six provinces that established leading groups from 1973 to 1975. See Table A1 in the appendix for an overview of FPLG establishment years across provinces.

Figure 1: Differential TFR changes with Leading Group establishment




# 3 Data

## 3.1 Data

We use data from the China Health and Retirement Longitudinal Study (CHARLS), an ongoing micro-longitudinal survey nationally representative of the Chinese older population aged 45 and above. CHARLS is the sister data set of Health and Retirement Study (HRS), Survey of Health, Aging and Retirement in Europe (SHARE), etc. CHARLS contains rich information on individual demographic characteristics, work status, retirement, pensions, and the family structure and interpersonal transfers. The main surveys start from 2011, and I pool data from all available waves, including the 2011, 2013, 2015, and 2018 waves. Details of the survey design, sampling procedure, and samples are described in Zhao et al. (2013) and X. Chen, Smith, Strauss, Wang, and Zhao (2017).

The main respondent and its spouse (if present) are interviewed as respondents of the household. The respondents are studied as the parents who make fertility and old-age labor supply decisions. Therefore, this study exploits individual fertility and old-age labor supply decisions of at most two respondents per household per year.

The sample is selected by focusing on only respondents who were born after 1925 and before 1952. The latter restriction ensures that respondents are in their old age because they have passed the normal retirement age (60) in all waves. Since the main interest is the impact of fertility on elderly labor supply, the selection also limits the channel through which family planning policies influence mothers' education as female respondents in the selected sample should have completed education by the time of LLF enforcement.[7] Since there are only few oldest old in the survey and that it is reasonable to study the labor supply of individuals under 90, the sample is restricted to be born after 1925.

Never married and never worked individuals are excluded, given that the main research

[7] The average year of leading team establishment is 1972, thus the average age of the youngest female respondents after selection is 20. Given the very low educational level of the cohort born before 1952, among whom 37% are illiterate, and 93% attained no higher than middle school, it is likely that women who were 20 years old had already completed their education before LLF was implemented.

question is how fertility affects the old-age labor supply. Couples with either spouse being an ethnic minority are dropped as well since family planning policies either exclude minorities at large or pose less strict restrictions on them (H. Li & Zhang, 2007; H. Li, Yi, & Zhang, 2011; Scharping, 2013). Finally, direct-administered municipalities, including Beijing, Tianjin, and Shanghai, are not included in the analysis as big cities could be essentially different from other provinces. They are much smaller in size, wealthier in economics, and more populated with higher educated residents. These cities might have experienced earlier exposure to small-scale family planning movements before LLF, which are difficult to measure and control. The final sample consists of 3,883 households, 6,292 respondents, and 18,453 respondent-year observations.

## 3.2 Variables and Descriptive Statistics

**Labor Supply** Both the extensive margin and the intensive margin of labor supply are investigated. *Whether currently working* is used to measure the retirement decision of the elderly. I define retirement as a cessation of work since the conventional definition of retirement – the reception of pensions – is not appropriate in the developing context. In addition, whether continuing working in old age better reflects the elderly's need to work due to the lack of support. In order to consider gradual retirement, the intensive margin – (the natural logarithms of) *Hours worked per week* – is studied as well. [8]

There are a considerable number of elderly, especially rural elderly, who work into their late 70s. Both rural and urban elderly in the sample gradually stop working as they become older, whereas urban residents retire much earlier, as plotted in Figure 2. Rural elderly are shown to be more likely to be working at all ages, which reflects their lack of pension support compared to urban counterparts. Between age 70 - 74, around 35% of urban elderly keep working, and the number amounts to 60% in rural areas. Even at age 80, almost 40% of rural residents are not retired, in contrast to the life expectancy at birth of 74 years. The high labor force participation in old age corresponds to an earlier description of the retirement pattern of Chinese elderly as "ceaseless toil" – the elderly have to work their entire lives due to lacking sufficient means of support (Davis, 1991).

Among the elderly who keep working, the vast majority are agricultural workers, regardless of their residence. Figure 2 not only reflects the population distribution of the sample but also suggests a physically strenuous and challenging nature of the working elderly's job. Farmers tend to keep working at older ages than individuals with non-agricultural jobs since they receive little or no pension benefits and have less accumulated wealth throughout their lifetime. No differentiation of agricultural and non-agricultural work is made in the empirical analysis, given that only few respondents over 60 are non-agricultural workers.[9]

The elderly gradually exit the labor market along with a reduction in their working hours while aging. At the intensive margin, among those who are currently working, the mean (median) total work hours falls from 47 (49) hours per week at age 60 to 25 (15) hours per week at age 80 and above. In addition, only 1 percent of the working sample have a second job besides their primary job. Thus having a second job is not individually investigated in the regression analysis. I employ the total working hours in the empirical

[8]The unemployed are classified into the retired group, as the proportion of unemployed is extremely small (0.08%), and there is no information on the intensive margin of labor supply among the unemployed.

[9]Urban residents living in counties or towns could participate in agriculture if they migrated from nearby villages and maintained access to agricultural land.

Figure 2: Work status across age by residence type


analysis to take side jobs and potential second jobs into account.

**Fertility** Family fertility outcomes are measured by a) *The number of living children* and *Son ratio*, or alternatively, b) *The number of living daughters* and *The number of living sons*. The second set of fertility outcomes provides an alternative way to interpret the impact of fertility.

The number of living children better captures the potential old-age support from children than the number of children ever born. Thus it is considered more relevant in determining elderly parents' labor supply. The sex ratio among living children within families provides additional information on the composition of children, which might affect family wealth and parental expectation about elderly support (Ebenstein & Leung, 2010; L. Li & Wu, 2017). Sex selection through either a male-biased fertility stopping rule or postnatal selection is possibly a side effect of the LLF policy, given the prevailing son preference in China (Yamaguchi, 1989; Arnold & Zhaoxiang, 1992; Clark, 2000; Babiarz et al., 2019).[10] Without differentiating the approaches and estimating the impact size of sex selection, I focus on the son ratio among living children (number of living sons/total number of living children) within families.

As the total fertility rate in Chinese families has dropped substantially during the past decades as a pronounced demographic trend, family size is also decreasing rapidly among younger households. Table 2 displays the trend of family size and within family son ratio across birth cohorts of the mother. From the cohort born between 1932 and 1936 onward, the total number of children keeps decreasing, whereas the ratio of sons is relatively stable. It implies that the number of sons and the number of daughters are reducing at a similar rate. The statistics confirm that families do not reduce family sizes by curbing only the number of daughters. It is also noticeable that the positive rural-urban difference in family sizes is present in every cohort. Rural families always have more children, potentially for risk-sharing and old-age support purposes.

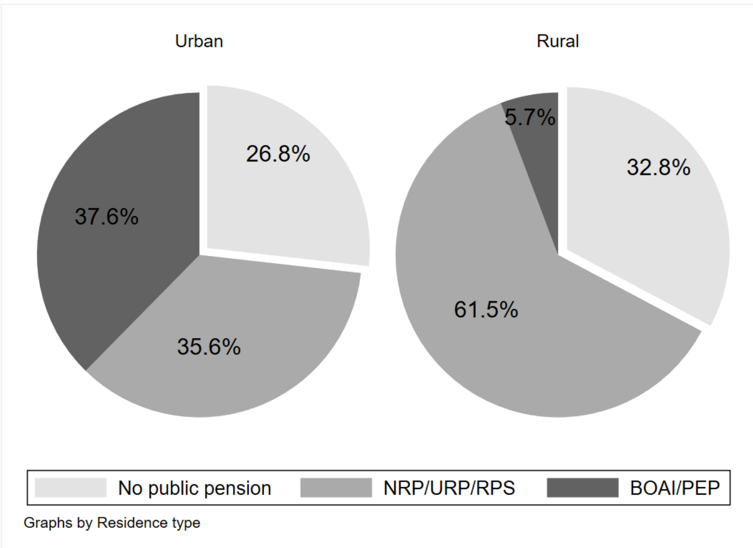
[10]Prenatal selection, i.e., abortion, is not a concern during the study period, as ultrasound technology was not widely available until the mid-1980s (H. Li & Zheng, 2009; Y. Chen, Li, & Meng, 2013).

Table 2: **Fertility outcomes across mother's birth cohort**

| Birth cohort | 1926-1931 | 1932-1936 | 1937-1941 | 1942-1946 | 1947-1951 | Total |
|---|---|---|---|---|---|---|
| *Total Nr. Children* | | | | | | |
| Rural | 4.51 | 4.87 | 4.43 | 3.74 | 3.25 | 3.76 |
| Urban | 3.98 | 4.29 | 3.56 | 3.23 | 2.78 | 3.22 |
| *Son Ratio* | | | | | | |
| Rural | 0.55 | 0.54 | 0.56 | 0.54 | 0.55 | 0.55 |
| Urban | 0.56 | 0.51 | 0.52 | 0.53 | 0.55 | 0.53 |
| OBS | 724 | 1,648 | 3,061 | 4,902 | 8,118 | 18,453 |

**Old-age support** To visualize the lack of pension benefits among both rural and urban residents, Figure 3 displays the proportion of residents with different types of the public pension. I categorize public pension schemes into employee pension, including BOAI and PEP, and non-employed resident pension, including NRP, URP, and RPS. Recall that pension benefits of the former (average replacement ratio of 50%) are much more generous than that of the latter (average replacement ratio lower than 20%) as discussed in the previous section. 32.8% of rural elderly and 26.8% of urban elderly do not participate in nor receive any kind of public pension. Thus not only rural elderly but also many urban counterparts have to rely on their own labor income or family support.[11] In addition, few rural elderly have access to employee pension (5.7%) compared to urban elderly (37.6%), which indicates that most rural pension recipients potentially receive a very low and inadequate level of pension benefits.

Figure 3: **Type of public pension by residence type**




The pension inequality between rural and urban is evident when comparing the amount of pension and subsidy benefits received during the previous year of the interview, as shown in Table 3. The mean and median, conditional on receiving any pension

[11]Note that although some urban residents participate in private pension programs such as annuity programs and commercial pension, the size is small.

or subsidy benefits, are 3,284 yuan and 840 yuan respectively among rural residents, in contrast to the official national poverty line of 2,300 yuan per person per year. However, the mean is four times, and the median is 14 times higher among urban elderly than rural residents.

Unsurprisingly, to compensate for the massive disparity in pension benefits, there are a higher labor participation rate and stronger reliance on inter-generational support among rural elderly. Table 3 shows that more rural elderly receive inter-generational material transfers (81% versus 72%) and remain staying in the labor market (58.4% versus 28.8%). Although the conditional amount of inter-generational transfer received is higher among urban parents, it is likely due to urban children's higher income and urban parents' higher expenditure. There is not much rural-urban difference at the intensive margin of labor supply if the elderly keep working. Overall, own labor supply and inter-generational support substitute pension when there is a lack of pension benefits.

Table 3: **Old-age support sources**

| | Rural | Urban |
|---|---|---|
| *Pension and subsidy* | | |
| Prop. | 70.7% | 79.2% |
| Mean (if >0) | 3,284 | 16,567 |
| Median (if >0) | 840 | 12,840 |
| *Intergenerational transfer* | | |
| Prop. | 81.0% | 72.2% |
| Mean (if >0) | 4,720 | 6,980 |
| Median (if >0) | 2,500 | 3,100 |
| *Labor supply* | | |
| Prop. | 58.4% | 28.8% |
| Hours/week (if >0) | 36.9 | 35.7 |
| More than part time (if >0) | 0.71 | 0.65 |
| OBS | 11,407 | 7,046 |

The importance of children in preventing elder parents, especially parents with worse health and little pension, from everlasting work is strongly supported by lower labor participation rates among parents with larger family sizes. Instead of age and rural-urban differentiation, I use (instrumental) activities of daily living ((I)ADL) limitations and the reception of pension benefits to more accurately reflect the restriction of working capability led by aging and the access to social support, respectively. Table 4 reports the varying labor participation rates across groups.

Parents with larger family sizes are less likely to be working conditional on health status and pension benefits. For instance, among parents with 3+ ADL limitations and no pension, those with more than three children are 10.3 pp (37.3% versus 27.0%) more likely to retire. Furthermore, parents are more likely to retire as their health deteriorates if they have more children. Among elderly parents with a pension, those who develop three and more ADL limitations are 19.1 pp (49.2% versus 30.1%) more likely to retire if they have at most three children, whereas the number is as high as 26.7 (49.6% versus 22.9%) if they have more than three children.

Table 4: **Family sizes and labor supply**

| Family size | ADL | No Pension | Have Pension |
|---|---|---|---|
| No more than 3 Children | $< 3$ | 62.8% | 49.2% |
| | 3+ | 37.3% | 30.1% |
| More than 3 Children | $< 3$ | 54.1% | 49.6% |
| | 3+ | 27.0% | 22.9% |

*Note:* The proportion of respondents who are currently working is reported in the table for each group.

**Control variables** Variables possibly influence elderly labor supply, including gender, hukou, residence type(rural or urban areas), educational attainment, age and age gap, whether widowed, health, wealth, pension, region, and survey year, are controlled.

Hukou denotes whether the individual is classified into a rural or an urban status in the household registration system, and it determines one's social welfare (Wu & Treiman, 2004). I control for both the first hukou and the current hukou, as the former might be more relevant to family planning policy enforcement in early lives whereas the latter determines the current social welfare. Age fixed effects are controlled by a group of age dummies. The age gap refers to the age gap between the oldest and the youngest among couples. Health status restricts the ability to work, while wealth and pension benefits determine the necessity to work. Thus these three are crucial factors to consider when making old-age labor supply decisions. Health is objectively measured by the number of chronic diseases and the number of (I)ADL. Wealth is measured by the value of household assets, including the housing value, durable assets, fixed capital assets, irrigable land, and agricultural asset – livestock and fisheries. Pension measures the amount of annual pension benefits received during the past year. Based on the mother's birth year, five cohort interval dummies (1926-1931, 1932-1936, 1937-1941, 1942-1946, and 1947-1951) are created and included to capture cohort fixed effects. Province and wave fixed effects are applied everywhere.

# 4 Empirical Strategy

**Fertility Instruments** This paper explores the less studied family planning policy "Later, longer and fewer" (LLF) implemented in the early 1970s to instrument fertility outcomes within families. To address the endogeneity issue of fertility decisions, various variations related to one-child policy (OCP) are frequently exploited in the Chinese context as they provide exogenous variations in fertility (Zhang, 2017). Recent empirical evidence suggests that LLF, instead of OCP, has mainly contributed to the significant decrease in family sizes during the past decades (Wang, 2016; Y. Chen & Fang, 2019; Babiarz et al., 2019; Y. Chen & Huang, 2020). Together with the policy exposure measure, the sex of the firstborn child and their interaction are adopted as instruments.

The sex of the firstborn predicts both the total number of children and the sex ratio within families. Families with son-biased preferences usually discontinue childbearing after achieving a desired number of sons. Thus they are more likely to give fewer births and comply with birth control policies when the eldest child is male.[12] The interaction term thus accounts for the additional policy impact among families with first-born sons. First-born son also raises the ratio of sons within families given no daughter preference at higher disparities.

The sex of the firstborn is plausibly exogenous since ultrasound technology for prenatal selection was in general unavailable at the time (H. Li & Zheng, 2009; Y. Chen et al., 2013), and that couples had little incentive to select the sex of the first birth as they were able to have more than one child.

This paper follows the policy exposure measure from Y. Chen and Fang (2019), which combines the age-specific exposure with the heterogeneous intensity across provinces. Similar age-specific exposure measures to family planning policies can be found in Miller (2010) and Wang (2016). The policy exposure for families with mothers born in year $c$ in province $p$ is measured as following

$$Exp_{pc} = \sum_{a=15}^{49} AFR_p(a) I[c + a \geq T_p] \tag{1}$$

where $AFR_p(a)$ is the provincial age-specific fertility rate in 1969 when the LLF policy has not yet been introduced anywhere, $a$ is age and $T_p$ denotes the establishment year of Family Leading Group in province $p$.[13] $I[.]$ is an indicator function which equals 1 when the statement is true. The policy exposure adds up the provincial age-specific fertility rate (AFR) at all remaining fertile ages of the mother.[14]

Therefore, based on the definition in Equ. 1, the variation of policy exposure comes from: a) differential leading group establishment year across provinces, b) varying initial

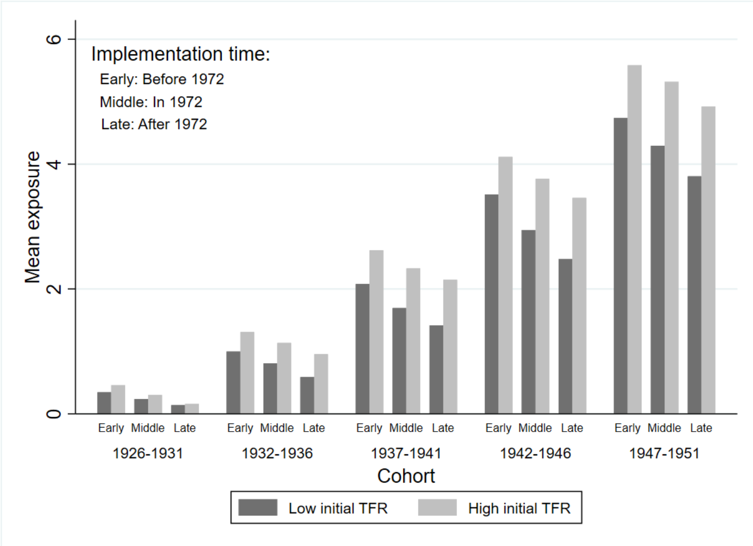
[12]Evidence supporting son preference is abundant. For instance, 94% families in the sample have at least one son, females on average have more siblings than males do, son-biased sex ratios are greater at high disparities, especially the last-born, and various male-biased stopping rules have been adopted (i.e., Babiarz et al., 2019).

[13]The age-specific fertility rate measures the annual number of births to women of a specified age or age group per 1,000 women in that age group. Provincial AFR data is retrieved from (Coale & Chen, 1987) who compute the data based on 1982 One-per-thousand Sample Fertility Survey.

[14]Child-bearing years range from age 15 to age 49, and women who were younger than 15 at the enforcement time are considered as full exposed, whereas women who were already 49 and above are considered to be not affected as their fertility decisions had completed by the time of policy enforcement.

fertility profiles across provinces, and c) birth cohort of the mother. Figure 4 illustrates the cohort and regional differences. Younger cohorts were more exposed, and women's mean exposure is higher in provinces with higher initial fertility rates and earlier policy enforcement conditional on the birth cohort.

Figure 4: Exposure measure by initial TFR and enforcement time




Age-specific exposure hypothesizes that women's fertility outcomes are more affected by family planning policies in a non-linear fashion among those younger at the enforcement time, as they spent more fertile years under the policy. The staggered roll-out of LLF thus implies that women of the same birth cohort who reside in provinces with earlier enforcement were younger when the policy was implemented, and consequently, they were exposed stronger to the policy.

The age-specific exposure can be captured by the sum of age-specific fertility rates (AFR) at all upcoming fertile ages, which accounts for both the length of exposure and the heterogeneous age profile of fertility. By using provincial AFR in the calculation, the variation of intensity across provinces is exploited as well. Women in provinces with higher initial fertility rates are more exposed to the LLF policy as LLF applied the same birth quota restrictions homogeneously to the entire country.

Among the 24 out of 28 provinces which are left in the final sample with complete information, one province established the Family Planning Leading Group the earliest in 1969, and 1 province formed their leading group the latest in 1975. Most of the provinces (16) had family planning groups set up in either 1971 or 1972. Table A1 in the appendix summarizes the leading team establishment years across provinces.

The cohorts of the sample are also reasonably affected by the LLF policy. Women would be impacted only if they were in their child-bearing years when the policy was in place. After the birth year selection, the minimum age is 18, and the maximum age is 49 at the time of policy enforcement. Table A2 in the appendix presents that 86% of the sample were under 35 years old when the policy was implemented in their respective provinces, which implies that the majority of the sample spent their most productive child-bearing years under the policy.

**Model specification** In a context where elderly parents rely on own labor income and support from children, fertility decline might lead to later retirement and longer work hours in old age. The following equation is estimated to examine the sign and the magnitude empirically:

$$LaborSupply_{icpt} = \alpha_0^{LS} + \beta_1 NrChild_{icpt} + \beta_2 SonRatio_{icpt} + \boldsymbol{\gamma x}_{icpt} + \delta_c + \delta_p + \delta_p \times c + \delta_t + \epsilon_{icpt}. \tag{2}$$

Labor supply mainly includes the extensive margin *Whether currently working* and the intensive margin *Hours worked per week*. The labor supply of individual i from family with a mother born in year c living in province p in survey year t is a function of: the total number of alive children, $NrChild_{icpt}$; the ratio of sons among living children, $SonRatio_{icpt}$; a vector of family and individual characteristics including age gap, whether widowed, highest educational attainment within the couple, wealth, and rural/urban residence, gender, age dummies, first hukou, current hukou, educational attainment, health status, annual pension benefits, $\boldsymbol{x}_{icpt}$; 5-year cohort group fixed effects, $\delta_c$; province fixed effects, $\delta_p$; province-specific cohort trend, $\delta_p \times c$; year fixed effects, $\delta_t$; and an error term, $\varepsilon_{icpt}$.

Two-stage least squares (2SLS) regression analysis is employed in this study, where two endogenous explanatory variables representing fertility outcomes – a) the total number of children and the sex composition of children or b) the number of sons and the number of daughters – are instrumented by the exposure to LLF policy ($Exp_{pc}$), the sex of the firstborn ($1stSon_{icpt}$), and their interaction ($Exp_{pc} \times 1stSon_{icpt}$). The following first stage regressions are estimated:

$$NrChild_{icpt} = \alpha_0^{NC} + \hat{\beta}_1 Exp_{pc} + \hat{\beta}_2 1stSon_{icpt} + \hat{\beta}_3 Exp_{pc} \times 1stSon_{icpt} + \hat{\boldsymbol{\gamma}} \boldsymbol{x}_{icpt} + \delta_c + \delta_p + \delta_p \times c + \delta_t + \eta_{icpt} \tag{3}$$

and

$$SonRatio_{icpt} = \alpha_0^{SR} + \hat{\beta}_1 Exp_{pc} + \hat{\beta}_2 1stSon_{icpt} + \hat{\beta}_3 Exp_{pc} \times 1stSon_{icpt} + \hat{\boldsymbol{\gamma}} \boldsymbol{x}_{icpt} + \delta_c + \delta_p + \delta_p \times c + \delta_t + \eta_{icpt}. \tag{4}$$

In an alternative specification, $NrDaughter_{icpt}$ and $NrSon_{icpt}$ are used to measure children's composition instead of the total number and the sex ratio. It provides more direct and intuitive results of the daughter/son effect for interpretation.

STATA command *ivreg2* with *gmm2s* and *partial* options are adopted in the main analyses (Baum, 2007).[15] Standard errors are clustered at the province-birth year level, given the construction of policy exposure. To examine the strength of the chosen instruments, Sanderson-Windmeijer F statistics for multiple endogenous variables are reported everywhere, instead of conventional first-stage F statistics (Sanderson & Windmeijer, 2016). The p-value of Hansen J statistic is provided for overidentification test as well, which provides inferences for the validity of instruments.

[15]Except Table 13, I partial out province dummies and province-specific cohort trends in every IV regression. It is to ensure that estimated covariance matrix of moment conditions are of full rank.

# 5 Main Results

## 5.1 First stage: Fertility Outcomes

Table 5: **First stage: Fertility Outcomes**

| | A | | B | |
|---|---|---|---|---|
| | Nr. Children | Son Ratio | Nr. Daughters | Nr. Sons |
| ***Rural*** | | | | |
| Exposure | -0.416*** | 0.021* | -0.300*** | -0.119** |
| | (0.083) | (0.013) | (0.070) | (0.059) |
| Exposure*1stSon | -0.049 | 0.013** | -0.004 | -0.046* |
| | (0.033) | (0.005) | (0.027) | (0.025) |
| 1stSon | -0.153 | 0.230*** | -1.027*** | 0.881*** |
| | (0.132) | (0.020) | (0.105) | (0.102) |
| Observations | 11,407 | 11,407 | 11,407 | 11,407 |
| Adjusted R-squared | 0.27 | 0.30 | 0.26 | 0.24 |
| ***Urban*** | | | | |
| Exposure | -0.393*** | -0.014 | -0.201*** | -0.194** |
| | (0.099) | (0.017) | (0.078) | (0.079) |
| Exposure*1stSon | -0.007 | 0.021*** | 0.009 | -0.015 |
| | (0.040) | (0.007) | (0.032) | (0.028) |
| 1stSon | -0.317** | 0.285*** | -1.159*** | 0.835*** |
| | (0.152) | (0.026) | (0.117) | (0.116) |
| Observations | 7,046 | 7,046 | 7,046 | 7,046 |
| Adjusted R-squared | 0.31 | 0.39 | 0.34 | 0.28 |

*Notes:* * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. Standard errors are clustered at the province-cohort level. Other controls, age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere. Mean exposure is 3.61 and 3.51 in the rural and the urban sub-sample, respectively. Rural and urban sample means of 1stSon are 0.53 and 0.51 respectively.

First stage results in Table 5 panel A show that exposure to LLF policy significantly decreases the total number of living children. Given that the rural (urban) sample mean of policy exposure is 3.61 (3.51), the policy on average reduces the family size by 0.416*3.61=1.50 (0.393*3.51=1.38) children among rural (urban) families with first-born daughters. The impact size is comparable to findings in Y. Chen and Huang (2020) and Y. Chen and Fang (2019), and slightly greater than the estimates of Babiarz et al. (2019). The positive rural-urban difference in the policy impact is also found by Y. Chen and Fang (2019).

The population policy also raises the sex ratio within households living in rural regions. Regardless of the sex of the firstborn, the policy impact is significantly positive when studying the son ratio among rural households. It suggests that there was an increase in stopping rule use under LLF, and such sex selection results in a biased within-family sex ratio (Babiarz et al., 2019). Son ratio in urban households, on the other hand, is not affected. The highly significant interaction term in both rural and urban indicates heterogeneous policy impact on sex ratio across families.

Having a first-born son exogenously decreases the total number of living children and increases the son ratio. Since ultrasound technology was not widely available during the study period, the sex of the first-born child is reasonably exogenous and balanced. Having a first-born son should not impact the total number of children if there is no gender preference present. However, it significantly reduces the family size by 0.33 (-0.049*3.61-0.153) and 0.34 (-0.007*3.51-0.317) among rural and urban respondents. The significant impact of having a first-born son provides additional evidence for the existence of son preference.

Results in panel B imply that the LLF policy reduces both the number of daughters and the number of sons and that families do not control family sizes by curbing only the number of daughters. The decline in the number of daughters is homogeneous among families with different sexes of the first-born. However, the number of sons would be reduced more significantly if the eldest child were male in rural families.

## 5.2 Labor Supply Decisions

Results of the 2SLS regression analysis show that having more children could significantly reduce rural parents' labor supply in old age. As shown in the first column of Table 6, rural parents who are 60+ are 17.7 pp less likely to be working if they have one more living child. The magnitude of the impact is relatively large, compared to the average working incidence of 58.4% among the rural sample.

In addition, the significant coefficient, i.e., -0.238, of son ratio for the rural sample in panel A implies that the higher ratio of sons a rural respondent has, the less likely the elder respondent is working. Similar results can also be found in panel B. Having more daughters and having more sons both reduce parents' probability of working, whereas the impact of sons is greater than that of daughters (-0.198 versus -0.130).

The pattern above is more substantial and only statistically significant among rural households. The rural-urban heterogeneity suggests that parents living in rural regions have to rely more on children for support as they generally have less accessible resources. The more evident gender differences regarding sons among rural parents are potentially the result of a more traditional gender role of sons as support providers.

The intensive margin is examined in addition to the retirement decision, given the possibility of gradual retirement. Conditional on working, the total working hours on all jobs could also account for potential side jobs.According to both panels in Table 6, having more children, either daughter or son, significantly decreases the total working hours per week in the urban sample.

The instruments potentially suffer from the weak instrument problem given the urban sample's relatively small first-stage F statistics. Nevertheless, the urban results discussed above provide suggestive evidence at the intensive margin.

Results suggest that fertility outcomes impact rural parents' labor supply mainly through reducing the incidence of working, whereas it decreases urban parents' labor supply mainly by lowering the working intensity.

Table 6: **Labor Supply Decisions**

| | Rural | | Urban | |
|---|---|---|---|---|
| | Working | Hours/week | Working | Hours/week |
| *Panel A* | | | | |
| Nr. Children | -0.177*** | 0.077 | 0.026 | -0.298* |
| | (0.053) | (0.093) | (0.050) | (0.158) |
| Son Ratio | -0.238*** | 0.122 | 0.044 | -0.446 |
| | (0.084) | (0.149) | (0.059) | (0.282) |
| Hansen J test p-value | 0.86 | 0.77 | 0.20 | 0.81 |
| SW 1st stage F-stat: Nr. Children | 14.86 | 9.72 | 8.38 | 5.50 |
| SW 1st stage F-stat: Son Ratio | 21.00 | 12.42 | 12.75 | 7.57 |
| *Panel B* | | | | |
| Nr. Daughters | -0.130*** | 0.052 | 0.010 | -0.218* |
| | (0.040) | (0.074) | (0.042) | (0.122) |
| Nr. Sons | -0.198*** | 0.090 | 0.020 | -0.389* |
| | (0.059) | (0.108) | (0.060) | (0.202) |
| Hansen J test p-value | 0.49 | 0.70 | 0.16 | 0.98 |
| SW 1st stage F-stat: Nr. Dau | 16.86 | 11.08 | 8.28 | 6.17 |
| SW 1st stage F-stat: Nr. Son | 16.42 | 10.03 | 8.37 | 6.31 |
| Observations | 11,407 | 5,985 | 7,046 | 1,801 |

*Notes:* * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. Standard errors are clustered at the province-cohort level. All controls at family and individual levels described in Section 3.2, age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere. The P-value of Hansen's J statistic provides inferences for the overidentification test. *SW 1st stage F-stat* refers to the Sanderson-Windmeijer first-stage F statistics, which tests whether a particular endogenous regressor is weakly identified (Sanderson & Windmeijer, 2016). STATA command *ivreg2* and the partial option are adopted in above analyses. Working hours per week is the logrithmic of the actual value.

## 5.3 Heterogeneous impact of children

### 5.3.1 Age group

To investigate the elderly from which age group is the most affected by children, similar analyses are conducted separately for age groups 60-64, 65-69, 70-74, and 75+, respectively. Theoretically, elder parents would be more in need of child support due to aging, so a stronger impact of children is expected among the elder parents. However, aforementioned pattern is found to be stronger among relatively younger post-retirement-age respondents, as shown in Table 7.

An overall decreasing gradient in coefficients is present across age groups, where the child quantity impact is stronger among younger respondents and the most significant among elder parents who are between 65 and 74. Nevertheless, old parents who are 70+ still work significantly less if they have more children, both statistically and quantitatively. For instance, parents between 70-74 are 10.9 pp less likely to be working if they have one more child.

Sons also play a more important role among younger elderly, who are below 70. The large coefficients of son ratio (above 0.22) among groups 60-64 and 65-69 imply that, in the extreme case where families change from having no son to having all male children (the ratio of sons to all children increases by one), old respondents in these age groups are more than 22 pp more likely to retire. However, the results among the youngest group should be interpreted with care as the Sanderson-Windmeijer first-stage F statistics are relatively low, which might raise a weak instrument problem.

To more accurately account for parental needs and the aging process, I stratify elder parents by health status and pension benefits in the following analyses.

### 5.3.2 Ceaseless toil – health status

To directly examine if working in old age is as Davis (1991) described as *ceaseless toil* – parents in bad health still need to rely on their own labor income, respondents are further categorized based on their health status instead of age. The number of (I)ADL limitations is a good indicator of physical working ability, as the proportion of respondents who keep working decreases from 52.6% among individuals with no ADL to 27.5% among those with 3+ ADL limitations. Its correlation with age (corr. 0.23) suggests that age might be less accurate in measuring the elderly's ability to work.

Table 8 shows that the impact of children on reducing parental labor supply is increasing with parents' health limitations. Parents with 3+ (I)ADL limitations are 14.7 pp more likely to retire when having one more child, compared to parents with no ADL limitation whose labor supply is not significantly reduced by having more children.

Both the number of daughters and the number of sons negatively impact parental labor supply among parents with any ADL limitation, as shown in panel B. In the meantime, sons are more influential than daughters in preventing parents from working in bad health. If the within son ratio increases by one (from no son to all sons), parents' probability of working would decrease by 18.8 pp if they have more than 1 or 2 ADL limitations.

In addition, old parents are stratified based on their self-rated health (SRH). Self-perception of own health status might be a more crucial determinant of happiness, in the sense that parents would feel worse if they have to work when they consider themselves very unhealthy. SRH reported as very good, good, and fair are classified as good health, and SRH of poor and very poor are classified into poor health.

Results are reported in the last two columns of Table 8. Similar findings present: parents with poor health tend to substitute more work with children (-0.078 versus -0.137). The substitution effect between children and their own labor is significant in both good and poor SRH groups, whereas it is greater in magnitude in the latter. Such a pattern can be found among both daughters and sons. Having one more daughter decreases the probability of working by 11.1 pp among parents with poor SRH, and the number almost doubled from 6.2 pp among parents with good SRH.

Results imply that parents in worse health conditions rely more on children, and children play a more critical role in providing old-age security when parents are more in need. Since LLF significantly reduces family sizes as shown in Table 5, population policies are likely to threaten vulnerable elderly's well-being by compelling them to work into bad health. Having more children, especially sons, can potentially improve sick elderly's life satisfaction by reducing their needs to work.

#### 5.3.3 Pension

Another crucial determinant of retirement decision is pension benefits, which constitutes the post-retirement income. According to the hypothesized role of children as old-age assets, the retirement decision of those with high pension benefits would be less sensitive to the availability of children. During the past year, 25% of the sample received no pension or subsidy benefit, and around 50% received more than 1,000 yuan among those with a pension. I thus define three groups based on whether they received any pension benefit and whether they received more than the median amount of positive benefits.

As shown in Table 9, the impact is most evident among parents with no pension. The incidence of working is decreased by 24.1 pp if there is one more child in the family, among those who have no pension at all. Both daughters and sons reduce parental needs of own labor income, whereas sons' impact is greater. The coefficient of son ratio in panel A is statistically significant, and the coefficients of *Nr. daughters* and *Nr. sons* in panel B are both significant but unequal. It suggests the existence of gender differences in a child's impact on parental labor supply.

To account for regional and cohort differences in pension benefit levels and pension eligibility rules, the community-age group-specific median is defined as the local median and compared with the respondent's pension benefits. The last three columns in Table 9 demonstrate the heterogeneity across relative levels of pension benefits.

Having more children, either daughters or sons, significantly decreases parents' working incidence when parents' pension benefits are not higher than peers. Such child impact on labor supply is the strongest among old respondents whose pension is lower than the median pension benefit level of people with similar age living in the same neighborhood. The heterogeneity is most evident among daughters. Having one more daughter reduces parents' incidence of working by 7.5 pp when parents receive median level pension benefits, whereas the magnitude of the impact is 72% (from 7.5 pp to 12.9 pp) larger among parents with a below-median pension.

The substitution effect is most drastic among parents with no pension as they have to rely on either labor or child support. In addition, compared with similar-aged elderly in the same community, parents whose pension benefits are not higher than the median are significantly less likely to work if they have more children. The heterogeneous impact presented here provides powerful support to the life-cycle hypothesis that parents work longer to compensate for the consequence of having fewer children, especially sons.

To sum up, the impact of children is heterogeneous across parental health status and pension benefit levels, and it is more substantial among the more vulnerable groups. Figure 5 visualizes the point estimates and its 90 percent confidence intervals of the number of daughters and sons across parents in different groups. It shows a positive gradient in the impact of having more children, either daughters or sons, when the parent is more disadvantaged.

The heterogeneity analysis of health and pension also explains and decomposes the rural-urban differences in the impact of child quantity and sex ratio on the incidence of parental old-age labor supply. Since rural residents have worse health (corr. (rural (I)ADL) = 0.08) and receive lower pension benefits (corr. (rural pension) = -0.38), they are more vulnerable and more in need of child support. Nevertheless, urban parents with limited access to social security may also find it challenging to prevent themselves from working in old age when policies restrict the family size.

Table 7: **Labor Supply Decision by Age Group**

| | | Age | | | |
|---|---|---|---|---|---|
| | Total | 60-64 | 65-69 | 70-74 | 75+ |
| *Panel A* | | | | | |
| Nr. Children | -0.096** | -0.175 | -0.177*** | -0.109* | -0.048 |
| | (0.037) | (0.115) | (0.059) | (0.057) | (0.066) |
| Son Ratio | -0.107** | -0.280** | -0.220*** | -0.002 | -0.063 |
| | (0.050) | (0.131) | (0.080) | (0.084) | (0.087) |
| Hansen J test p-value | 0.60 | 0.36 | 0.86 | 0.47 | 0.32 |
| SW 1st stage F-stat: Nr. Children | 22.17 | 6.11 | 16.35 | 12.54 | 6.90 |
| SW 1st stage F-stat: Son Ratio | 31.96 | 6.02 | 20.51 | 20.41 | 12.92 |
| *Panel B* | | | | | |
| Nr. Daughters | -0.077** | -0.118 | -0.134*** | -0.109** | -0.044 |
| | (0.030) | (0.094) | (0.047) | (0.048) | (0.055) |
| Nr. Sons | -0.112*** | -0.234 | -0.214*** | -0.110 | -0.064 |
| | (0.042) | (0.147) | (0.069) | (0.069) | (0.075) |
| Hansen J test p-value | 0.87 | 0.27 | 0.88 | 0.44 | 0.37 |
| SW 1st stage F-stat: Nr. Dau | 24.13 | 5.51 | 17.12 | 12.40 | 6.78 |
| SW 1st stage F-stat: Nr. Son | 23.72 | 5.25 | 16.51 | 12.16 | 6.97 |
| Observations | 18,453 | 3,153 | 5,406 | 4,689 | 5,205 |

*Notes:* * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. Standard errors are clustered at the province-cohort level. The dependent variable is labor support decision at the extensive margin – whether currently working. All controls at family and individual levels described in Section 3.2, age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere. The P-value of Hansen's J statistic provides inferences for the overidentification test. *SW 1st stage F-stat* refers to the Sanderson-Windmeijer first-stage F statistics, which tests whether a particular endogenous regressor is weakly identified (Sanderson & Windmeijer, 2016). STATA command *ivreg2* and the partial option are adopted in above analyses.

Table 8: **Heterogeneous impact across health status**

| | Nr. (I)ADL | | | Self-rated Health | |
|---|---|---|---|---|---|
| | No (I)ADL | 1-2 (I)ADL | 3+ (I)ADL | Good | Poor |
| *Panel A* | | | | | |
| Nr. Children | -0.062 | -0.126*** | -0.147** | -0.078* | -0.137** |
| | (0.045) | (0.049) | (0.074) | (0.040) | (0.061) |
| Son Ratio | -0.049 | -0.188** | -0.149 | -0.087 | -0.139* |
| | (0.058) | (0.074) | (0.096) | (0.053) | (0.083) |
| Hansen J test p-value | 0.63 | 0.24 | 0.86 | 0.85 | 0.80 |
| SW 1st stage F-stat: Nr. Children | 14.25 | 13.97 | 7.55 | 16.71 | 13.87 |
| SW 1st stage F-stat: Son Ratio | 21.01 | 25.80 | 11.09 | 25.64 | 19.98 |
| *Panel B* | | | | | |
| Nr. Daughters | -0.050 | -0.100** | -0.116** | -0.062* | -0.111** |
| | (0.036) | (0.041) | (0.058) | (0.033) | (0.050) |
| Nr. Sons | -0.066 | -0.163*** | -0.159** | -0.092** | -0.154** |
| | (0.052) | (0.058) | (0.079) | (0.047) | (0.069) |
| Hansen J test p-value | 0.54 | 0.50 | 1.00 | 0.91 | 0.96 |
| SW 1st stage F-stat: Nr. Dau | 15.47 | 14.20 | 9.05 | 18.34 | 14.61 |
| SW 1st stage F-stat: Nr. Son | 15.55 | 14.74 | 8.36 | 17.83 | 14.50 |
| Observations | 10,273 | 4,660 | 3,520 | 12,054 | 5,480 |

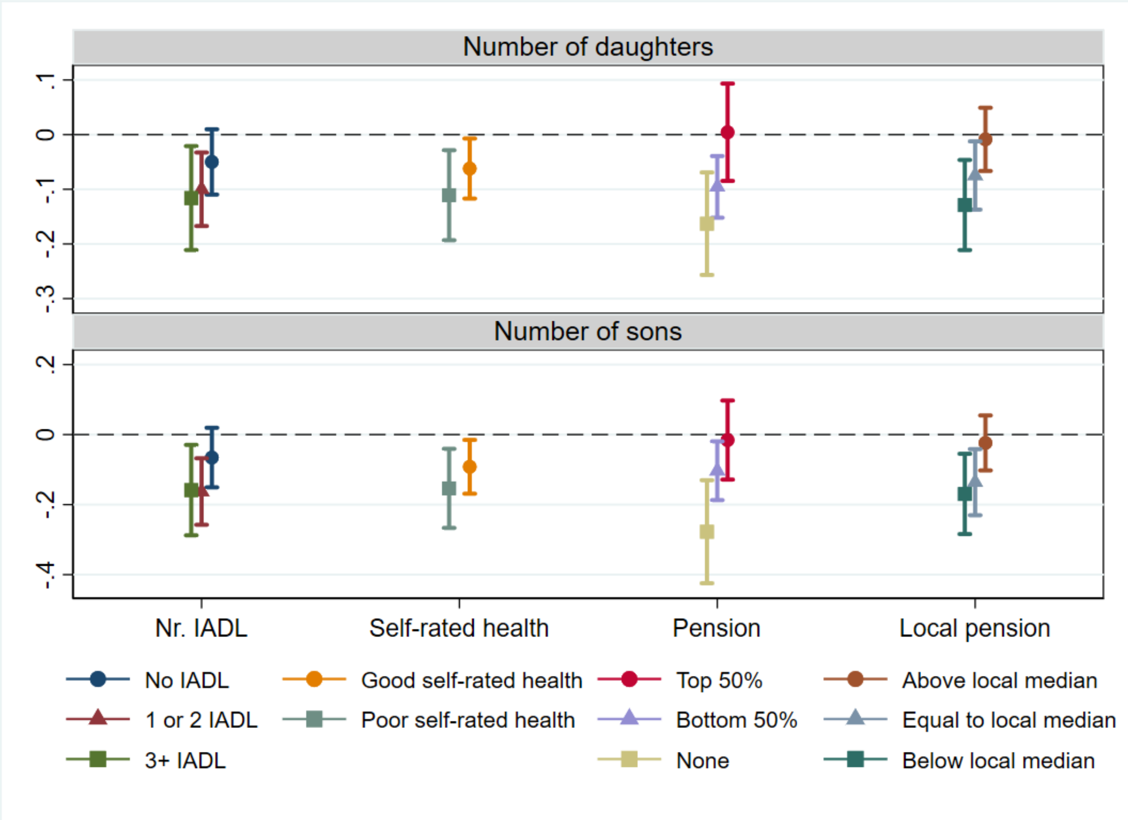
*Notes:* * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. Standard errors are clustered at the province-cohort level. The dependent variable is labor support decision at the extensive margin – whether currently working. All controls at family and individual levels described in Section 3.2, age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere. The P-value of Hansen's J statistic provides inferences for the overidentification test. *SW 1st stage F-stat* refers to the Sanderson-Windmeijer first-stage F statistics, which tests whether a particular endogenous regressor is weakly identified (Sanderson & Windmeijer, 2016). STATA command *ivreg2* and the partial option are adopted in above analyses.

Table 9: **Heterogeneous impact across pension benefits**

| | Pension | | | Local pension | | |
|---|---|---|---|---|---|---|
| | Top 50% | Bottom 50% | None | Above median | Equal median | Below median |
| *Panel A* | | | | | | |
| Nr. Children | 0.004 | -0.105** | -0.241*** | -0.010 | -0.118** | -0.163** |
| | (0.060) | (0.044) | (0.079) | (0.041) | (0.051) | (0.065) |
| Son Ratio | -0.043 | -0.036 | -0.377*** | -0.034 | -0.205** | -0.136* |
| | (0.052) | (0.080) | (0.127) | (0.053) | (0.084) | (0.081) |
| Hansen J test p-value | 0.28 | 0.64 | 0.56 | 0.20 | 0.70 | 0.73 |
| SW 1st stage F-stat: Nr. Children | 8.99 | 20.36 | 10.15 | 16.62 | 18.61 | 11.16 |
| SW 1st stage F-stat: Son Ratio | 14.47 | 31.79 | 12.15 | 28.74 | 25.60 | 14.54 |
| *Panel B* | | | | | | |
| Nr. Daughters | 0.004 | -0.096*** | -0.163*** | -0.009 | -0.075** | -0.129** |
| | (0.054) | (0.034) | (0.057) | (0.035) | (0.038) | (0.050) |
| Nr. Sons | -0.016 | -0.103** | -0.277*** | -0.024 | -0.136** | -0.169** |
| | (0.069) | (0.051) | (0.089) | (0.048) | (0.057) | (0.070) |
| Hansen J test p-value | 0.37 | 0.58 | 0.22 | 0.24 | 0.43 | 0.50 |
| SW 1st stage F-stat: Nr. Dau | 9.10 | 23.26 | 10.28 | 17.04 | 21.42 | 12.63 |
| SW 1st stage F-stat: Nr. Son | 9.27 | 22.38 | 9.91 | 17.17 | 20.64 | 11.99 |
| Observations | 6,579 | 7,069 | 4,805 | 7,301 | 4,484 | 6,668 |

*Notes:* * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. Standard errors are clustered at the province-cohort level.The dependent variable is labor support decision at the extensive margin – whether currently working. All controls at family and individual levels described in Section 3.2, age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere. The P-value of Hansen's J statistic provides inferences for the overidentification test. *SW 1st stage F-stat* refers to the Sanderson-Windmeijer first-stage F statistics, which tests whether a particular endogenous regressor is weakly identified (Sanderson & Windmeijer, 2016). STATA command *ivreg2* and the partial option are adopted in above analyses.

Figure 5: Heterogeneous impact of children




# 6 Discussion

## 6.1 Child support

The findings above suggest that own labor supply at old age is a substitute for children, especially sons. I next examine the influence of child quantity and within-family sex ratio on parents' received material support from children to test the potential mechanism of child support.

Theoretically, when the child quantity-quality trade-off exists, the total material support from children is not necessarily higher among families with larger sizes. There is evidence showing that along with the decrease in child quantity is the improvement of child quality (H. Li, Zhang, & Zhu, 2008; Liu, 2014; Oliveira, 2016; B. Li & Zhang, 2017). Children with fewer siblings might receive better human capital investment and consequently earn more and transfer more to parents in adulthood. The higher quality of children thus might attenuate the negative impact of having fewer children on parental old-age security. The same applies to the quality and quantity of sons, as sons are considered the responsible caregivers in the traditional Chinese context.

The first two columns of Table 10 show that having more children (daughter/son) significantly increases the probability and the amount of transfers they received from children. Parents with one more child are 6.0 pp more likely to receive any material transfer from children, and they receive 28.0% higher transfer conditional on receiving transfers, which amounts to 770.5 yuan at the median level.[16]

[16]The amount value is calculated based on the percentage change and the conditional median of material transfers, i.e., $(exp(0.247) - 1) \times 2750 = 770.5$.

Surprisingly, no gender difference presents in the total help from children. It suggests that daughters are equivalently important as their male siblings in providing material support, although sons have a more significant impact on parental labor supply. There might be a discrepancy between parental perception and the actual transfer, in the sense that parents may consider sons as more reliable although daughters provide similar amounts of material support.

Regular transfer results presented in the last columns show slightly different patterns, where only the incidence of regular transfer is increased significantly by the ratio of sons within families. Sons do not provide more material transfers in total, but they are more likely to transfer regularly. Although results suggest gender differences in terms of regular transfer, neither the number of daughters nor that of sons statistically significantly contributes to a higher probability of regular transfer, according to panel B.[17]

To provide suggestive evidence on the quality-quantity trade-off, I also test the impact of fertility outcomes on the conditional amount of material transfer per child. The third and sixth columns of Table 10 show that each child provides less total and regular support on average when there are more siblings. Such reduction might be due to the substitution between siblings, and it might also reflect siblings' lower socioeconomic statuses on average. In short, although support from each child decreases with family sizes, the sum of support received from all children is more when the family is larger.

[17] I also experimented with net transfer from children, which is the transfers received minus downward transfers to children. Results are quantitatively and qualitatively very similar, see Table A4.

Table 10: **Material support from children**

| | Total | | | Regular | | |
|---|---|---|---|---|---|---|
| | Incidence | Amt | Amt per child | Incidence | Amt | Amt per child |
| *Panel A* | | | | | | |
| Nr. Children | 0.060** | 0.247** | -0.535*** | 0.049 | 0.087 | -0.451*** |
| | (0.027) | (0.124) | (0.087) | (0.031) | (0.160) | (0.077) |
| Son Ratio | -0.015 | 0.011 | 0.172 | 0.092** | 0.039 | 0.156 |
| | (0.037) | (0.170) | (0.133) | (0.044) | (0.278) | (0.144) |
| Hansen J test p-value | 0.08 | 0.20 | 0.88 | 0.26 | 0.19 | 0.34 |
| SW 1st stage F-stat: Nr. Children | 22.17 | 18.20 | 18.20 | 22.17 | 13.85 | 13.85 |
| SW 1st stage F-stat: Son Ratio | 31.96 | 30.31 | 30.31 | 31.96 | 25.45 | 25.45 |
| *Panel B* | | | | | | |
| Nr. Daughters | 0.059*** | 0.259** | -0.567*** | 0.028 | 0.102 | -0.486*** |
| | (0.022) | (0.104) | (0.070) | (0.026) | (0.134) | (0.067) |
| Nr. Sons | 0.049 | 0.274** | -0.517*** | 0.055 | 0.137 | -0.451*** |
| | (0.031) | (0.139) | (0.098) | (0.036) | (0.189) | (0.091) |
| Hansen J test p-value | 0.10 | 0.22 | 0.69 | 0.14 | 0.22 | 0.25 |
| SW 1st stage F-stat: Nr. Dau | 24.13 | 19.47 | 19.47 | 24.13 | 14.42 | 14.42 |
| SW 1st stage F-stat: Nr. Son | 23.72 | 19.79 | 19.79 | 23.72 | 14.53 | 14.53 |
| Observations | 18,453 | 14,330 | 14,330 | 18,453 | 4,913 | 4,913 |

*Notes:* * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. Standard errors are clustered at the province-cohort level. All controls at family and individual levels described in Section 3.2, age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere. The P-value of Hansen's J statistic provides inferences for the overidentification test. *SW 1st stage F-stat* refers to the Sanderson-Windmeijer first-stage F statistics, which tests whether a particular endogenous regressor is weakly identified (Sanderson & Windmeijer, 2016). STATA command *ivreg2* and the partial option are adopted in above analyses. The amounts of material transfers are the natural logarithm of the actual values.

## 6.2 Lifetime labor supply

The analysis above reveals a strong impact of fertility outcomes on old-age labor force participation, whereas less is known about these elderly's lifetime labor supply. Although it is not the paper's primary focus, investigating the work intensity through the lifetime offers valuable insight into the implication on parents' overall well-being. The following exercises use work history information from the life history questionnaire to provide suggestive evidence on fertility's lifetime labor supply impact.

Retrieving recall information on all jobs during the past from the life history data, I measure the lifetime work intensity by the weighted average of monthly working hours.[18] To better represent the potentially unhealthier case of overwork, a dummy variable indicating whether working over 55 hours per week is additionally studied. The analysis further examines the labor supply impact before having children by studying the first job's working hours to highlight the channel of fertility outcomes. Some observations are unavoidably lost due to missing recall data.

Results related to monthly working hours during the lifetime are presented in the first two columns in Table 11. Having more children significantly contributes to fewer working hours and a lower probability of overwork on average. As can be observed from Panel B, both sons and daughters matter, albeit the latter has a relatively smaller influence in magnitude given the significantly negative son ratio. Quantitatively, parents with one more son reduce monthly working hours by 15.6 percent, which is roughly 30 hours per month at the median level, and they are 16.2 pp less likely to overwork during the lifetime.[19]

The third column in Table 11 provides suggestive evidence for the fertility channel of the labor supply impact. Around 90 percent of the sample started their first job no later than age 20, and it is consistent with their low educational attainment – only six percent attained higher than middle school.[20] People likely start their first job right after stopping school education and before making fertility decisions. Therefore, examining the labor supply at the first job helps us disentangle the fertility channel from other confounding mechanisms. Statistically insignificant results of the first job imply that unrealized fertility outcomes have no impact on labor supply and that fertility outcomes are the reason behind the strong lifetime labor supply effect.

Findings on lifetime labor supply indicate that restricted by population policies, parents with fewer children might experience great strains from longer working hours, even overwork, throughout the lifetime in addition to prolonged working life in old age. In contrast, parents with more children are less concerned about old-age support, so they work less and save less. These findings are well in line with previous studies that found a negative impact of fertility on household savings, e.g., Banerjee et al. (2014); Choukhmane, Coeurdacier, and Jin (2013); Ge et al. (2018).

[18] Working hour questions were only asked with regard to the first job regardless of the length and any other jobs that lasted for at least five or twenty years. The weighted average monthly working hours is constructed based on the job length in years, months worked per year, days worked per month, and hours worked per day for each job over five years with relevant information.

[19] The amount is calculated based on the percentage change and the median of average monthly working hours, i.e., $(exp(-0.170) - 1) \times 191 = -29.9$.

[20] Teenagers usually graduate from middle school at around age 15.

Table 11: **Lifetime Labor Supply Impact**

| | Lifetime weighted mean | | First job |
|---|---|---|---|
| | Hours/month | Over 55 hours/week | Hours/month |
| *Panel A* | | | |
| Nr. Children | -0.141** | -0.147** | 0.006 |
| | (0.063) | (0.058) | (0.066) |
| Son Ratio | -0.249*** | -0.170** | -0.087 |
| | (0.083) | (0.082) | (0.089) |
| Hansen J test p-value | 0.58 | 0.70 | 0.77 |
| SW 1st stage F-stat: Nr. Children | 17.69 | 17.69 | 17.69 |
| SW 1st stage F-stat: Son Ratio | 23.81 | 23.81 | 23.81 |
| *Panel B* | | | |
| Nr. Daughters | -0.092* | -0.110** | 0.019 |
| | (0.052) | (0.045) | (0.054) |
| Nr. Sons | -0.170** | -0.162** | -0.011 |
| | (0.073) | (0.066) | (0.078) |
| Hansen J test p-value | 0.24 | 0.42 | 0.94 |
| SW 1st stage F-stat: Nr. Dau | 18.67 | 18.67 | 18.67 |
| SW 1st stage F-stat: Nr. Son | 18.50 | 18.50 | 18.50 |
| Observations | 14,469 | 14,469 | 14,469 |

*Notes:* * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. Standard errors are clustered at the province-cohort level. Age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere. Late life controls, including current health status, value of household assets and pension benefits, are substituted with health status before age 16, family economic status before age 17, mother's education, and age when clean water is available, to avoid bad control problem. The P-value of Hansen's J statistic provides inferences for the overidentification test. *SW 1st stage F-stat* refers to the Sanderson-Windmeijer first-stage F statistics, which tests whether a particular endogenous regressor is weakly identified (Sanderson & Windmeijer, 2016). STATA command *ivreg2* and the partial option are adopted in above analyses. Working hours per month is the logrithmic of the actual value.

## 6.3 Robustness

This section provides a set of sensitivity analyses to examine the robustness of the results. There are two fundamental requirements of an instrumental variable approach: relevance and exogeneity. The first-stage results show that the chosen IVs – policy exposure, firstborn son, and their interaction – are proven to be highly relevant to fertility outcomes. The Sanderson-Windmeijer first-stage F statistics are also usually greater than 10, supporting that neither of the fertility outcomes is weakly identified. As ultrasound B technology was not prevalent during the study period, prenatal sex selection is not a concern, and the sex of the firstborn child is arguably exogenous (H. Li & Zheng, 2009; Y. Chen et al., 2013). In addition, the exposure to family planning policies is an aggregate macro variable, whereas the outcome is individual labor supply decisions. The relevance and exogeneity requirements are likely to be satisfied. Following tests help rule out the possibility that the impact of instruments on labor supply operates via channels other than fertility outcomes.

First, concerns may arise from the confounding impact of other contemporaneous historical events, including the Cultural Revolution (1966-1976) and the Send-down Movement (1967-1978). Cultural Revolution was a political event that caused catastrophic violence and disturbance throughout the country, and Send-down Movement was a political migration movement that sent tremendous urban youths to the countryside (Gu, 2009). Experiencing either violent and unrest social environment or forced migration to unfamiliar places when young is likely to directly influence individuals' attitudes and behavior regarding fertility, work, and other later-life outcomes.

To address the concern, I control each event's intensity at the provincial level that interacted with the mother's birth cohort. The intensity of the Cultural Revolution is measured by aggregate fatalities during the event as a share of the province's population in 1965 (Walder, 2017). The total number of people who were sent down during the movement as a share of the provincial population born between 1945 and 1960 is used to measure the intensity of the Send-down Movement (Gu, 2009). Mothers within different birth cohorts may experience differential exposure to the same historical events, so the interactions sufficiently capture such cohort-specific impact. It also tackles the multicollinearity issue that province dummies have already captured the provincial intensity of these events. Results in column (2) in Table 12 suggest that the negative impact of child quantity and son ratio are robust to the inclusion of historical events. Both coefficients and standard errors barely change after introducing the cohort-specific event intensities, compared to original estimates displayed in column (1).

Secondly, it is possible that people with better health care access and a better economic environment when they were young would make systematically different fertility and labor supply decisions from others. Such different initial socioeconomic conditions also correlate with their cohort and province, based on which the policy exposure instrument is constructed. To eliminate socioeconomic development's impact on fertility and labor supply (the concern that the policy instrument affects labor supply via socioeconomic development other than fertility), I include province-specific cohort trends in every regression analysis. To account for more detailed cohort-province-specific socioeconomic differences, provincial health care, economic development, demographic and educational conditions when the mother turned age 20 are additionally included as the second set of sensitivity analyses.

These additional controls consist of the numbers of hospital beds and doctors per

10,000 persons, GDP, GDP per capita, the share of primary industry in GDP, total population, sex ratio, death rate, the number of high, secondary, and primary school teachers per 10,000 persons, at the province level. Results are reported in column (3) in Table 12. The number of observations decreases due to missing values of some cohort-province variables, while main results remain robust – having more children, especially sons, reduces the parental incidence of working at post-retirement age. Column (4) reports similar results when considering both historical events and socioeconomic status.

Next, I check the robustness of the results to different clustering methods. In the main analysis, standard errors are clustered at the province-cohort level as an individual's exposure to family planning policies is defined based on province and mother's birth year. To allow for within-family correlation specifically, I cluster at the household level. The resulting standard errors are reported in column (5). It is not surprising that standard errors under the two different correction methods are highly similar, given the already relatively large number of clusters (548) in the main analysis. Overall, results are always statistically significant, regardless of the correction method.

Finally, regressions with individual birth years are analyzed to justify the choice of 5-year-interval birth cohort groups. There is a strong tendency for the policy exposure measure and the mother's birth year to be highly correlated because the former is constructed as the aggregate provincial age-specific fertility rate based on the latter. Besides, the paper's primary goal is to explain the elderly labor supply instead of evaluating the causal impact of population policies. Therefore, in order to capture cohort effects and at the same time circumventing the multicollinearity issue, the functional form of cohorts should be carefully chosen.

To test if the five cohort interval dummies (1926-1931, 1932-1936, 1937-1941, 1942-1946, and 1947-1951) have sufficiently captured cohort effects, I include 21 extra birth year dummies (26 birth years - 5 cohort groups) in the baseline regression. Results in column (2) in Table 13 reveal that indeed adding the extra year of birth dummies takes away the power of IV. However, a joint significance test on the extra year dummies is not rejected (p=1.00), suggesting that these extra variables do not provide additional information and that the 5 group dummies are sufficient. In the meantime, the original cohort group FE and province-specific cohort trend are still jointly significant (p=0.01).[21]

I further substitute cohort group dummies with the year of birth dummies everywhere in column (3) and test the joint significance of birth year FE and province-specific birth year trend. The null hypothesis of joint insignificance is not rejected at any level (p=0.60). In contrast, a similar test of cohort group FE and province-specific cohort trend in the baseline model is strongly rejected at the 1% level (p=0.00). The different test results indicate that cohort groups better explain the variance of old-age labor supply, the variable of interest, than years of birth. In addition, I test the assumption that the five or six birth years within each cohort group have equal coefficients. Very high p-values (ranging from 0.81 to 0.96) support the equality assumption. There is also no overall improvement in model fit (based on adjusted R-squared) when switching from cohort group to birth year.

In short, to avoid wrongly diminishing the policy instrument's power or mistaking cohort difference as the fertility impact, the grouped cohorts are a better choice. For the sake of simplicity of the model specification, I adopt 5-year intervals in the paper.

In conclusion, the main results are robust to various sensitivity tests.

[21] The additional year of birth dummies are not jointly significant (p=0.92 and 0.41) in the first stage regressions either, see column (6) and column (7).

Table 12: **Sensitivity analysis**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| *Panel A* | | | | | |
| Nr. Children | -0.096** | -0.097*** | -0.102** | -0.107** | -0.097*** |
| | (0.037) | (0.037) | (0.047) | (0.049) | (0.037) |
| Son Ratio | -0.107** | -0.109** | -0.099* | -0.106* | -0.106** |
| | (0.050) | (0.050) | (0.056) | (0.058) | (0.051) |
| Hansen J test p-value | 0.60 | 0.58 | 0.74 | 0.74 | 0.57 |
| SW 1st stage F-stat: Nr. Children | 22.17 | 21.88 | 13.48 | 13.32 | 23.61 |
| SW 1st stage F-stat: Son Ratio | 31.96 | 31.97 | 21.50 | 21.57 | 38.74 |
| *Panel B* | | | | | |
| Nr. Daughters | -0.077** | -0.078** | -0.077** | -0.081** | -0.077*** |
| | (0.030) | (0.030) | (0.038) | (0.038) | (0.029) |
| Nr. Sons | -0.112*** | -0.114*** | -0.107** | -0.113** | -0.112*** |
| | (0.042) | (0.043) | (0.051) | (0.052) | (0.042) |
| Hansen J test p-value | 0.87 | 0.86 | 0.52 | 0.51 | 0.86 |
| SW 1st stage F-stat: Nr. Dau | 24.13 | 23.77 | 14.76 | 14.58 | 25.67 |
| SW 1st stage F-stat: Nr. Son | 23.72 | 23.55 | 14.75 | 14.72 | 25.45 |
| Historical events | | ✓ | | ✓ | |
| Econ, health, demo & educ status | | | ✓ | ✓ | |
| Cluster at household | | | | | ✓ |
| Observations | 18,453 | 18,453 | 14,532 | 14,532 | 18,453 |

*Notes:* The dependent variable is labor support decision at the extensive margin – whether currently working. * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. If not specified, standard errors are clustered at the province-cohort level. Other controls, age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere. Historical events include Cultural Revolution and Send-down Movement, where we allow the impact of their provincial intensity to vary across cohorts. Provincial economic, health, demographic and education status at age 20 are captured by a group of variables.

Table 13: **Sensitivity analysis: birth cohorts**

| | IV | | | First stage | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Baseline | Extra yob dummies | Yob instead of cohort | Baseline | | Extra yob dummies | | Yob instead of cohort | |
| | | | | Nr. Children | Son Ratio | Nr. Children | Son Ratio | Nr. Children | Son Ratio |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Number of living children | -0.096** | -0.220 | -0.253 | | | | | | |
| | (0.037) | (0.251) | (0.270) | | | | | | |
| Ratio of sons | -0.107** | -0.236 | -0.270 | | | | | | |
| | (0.050) | (0.264) | (0.283) | | | | | | |
| Exposure | | | | -0.412*** | 0.006 | -0.218 | 0.042 | -0.272 | 0.059* |
| | | | | (0.066) | (0.010) | (0.252) | (0.033) | (0.272) | (0.034) |
| Exposure*1stSon | | | | -0.043 | 0.016*** | -0.044 | 0.016*** | -0.042 | 0.016*** |
| | | | | (0.027) | (0.005) | (0.028) | (0.005) | (0.028) | (0.005) |
| 1stSon | | | | -0.176* | 0.252*** | -0.170 | 0.251*** | -0.176* | 0.251*** |
| | | | | (0.105) | (0.017) | (0.106) | (0.017) | (0.106) | (0.017) |
| Adjusted R-squared | | | | 0.29 | 0.33 | 0.29 | 0.33 | 0.29 | 0.33 |
| Observations | 18,453 | 18,453 | 18,453 | 18,453 | 18,453 | 18,453 | 18,453 | 18,453 | 18,453 |
| Hansen J test p-value | 0.60 | 1.00 | 0.92 | | | | | | |
| SW 1st stage F-stat: Nr. Children | 22.17 | 0.74 | 0.74 | | | | | | |
| SW 1st stage F-stat: Son Ratio | 31.96 | 0.78 | 0.78 | | | | | | |
| **Joint significance test (p-value)** | | | | | | | | | |
| Extra yob dummies | | 1.00 | | | | 0.92 | 0.41 | | |
| i.yob & c.yob#i.province | | | 0.60 | | | | | 0.00 | 0.03 |
| i.cohort & c.cohort#i.province | 0.00 | 0.01 | | 0.00 | 0.03 | 0.00 | 0.02 | | |
| Equality of yobs in each cohort | | | 0.81-0.96 | | | | | 0.39-0.91 | 0.20-0.87 |

*z* statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1
*Note:* The dependent variable is labor support decision at the extensive margin – whether currently working. Robust standard errors in parentheses. Standard errors are clustered at the province-cohort level. Other controls, age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere.

# 7 Conclusion

This study investigates the impact of child quantity and child sex composition on parents' old-age labor supply decisions and examines if children can prevent vulnerable old parents from working into old age. Microdata from the China Health and Retirement Longitudinal Study (CHARLS) is used, and an instrumental variable method is adopted to account for endogenous fertility decisions. IVs in this paper utilize the differential exposure to family planning policies in the early 1970s and the exogenous sex of the firstborn child. Heterogeneity across health statuses and pension benefits of the parent is further explored to test if children are more important when parents are more in need.

Overall, the results provide strong evidence for children's substitution effect on old parents' own labor supply – having more children significantly reduces parents' probability of working at post-retirement age. The impact is more substantial among more vulnerable groups, for instance, parents with worse health, and parents with no pension. Although there might be a quality-quantity trade-off in children, a larger family size still results in more material transfers received by parents, explaining the child impact.

Findings in this paper suggest that population policies potentially threaten the elderly well-being by reducing family sizes. When social security systems are not improving correspondingly, old parents would lack support from children and have to work till physically incapable.

Understanding the role of children in preventing old parents from strenuous work provides crucial implications to many aging societies with immature public security systems. Special public supporting schemes should be directed to the unprepared generation impacted the most by the family size shrinkage to complement the traditional yet most important family support.

# References


Agüero, J. M., & Marks, M. S. (2011). Motherhood and female labor supply in the developing world evidence from infertility shocks. *Journal of Human Resources*, *46*(4), 800–826.

Angrist, J. D., & Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, *88*(3), 450–477.

Arnold, F., & Zhaoxiang, L. (1992). Sex preference, fertility, and family planning in china. In *The population of modern china* (pp. 491–523). Springer.

Babiarz, K. S., Ma, P., Miller, G., & Song, S. (2019). *The Limits (And Human Costs) of Population Policy: Fertility Decline and Sex Selection in China under Mao* (Working Papers No. 505). Center for Global Development.

Banerjee, A., Meng, X., Porzio, T., & Qian, N. (2014). *Aggregate fertility and household savings: A general equilibrium analysis using micro data* (Tech. Rep.). National Bureau of Economic Research.

Banerjee, A., Meng, X., & Qian, N. (2010). The life cycle model and household savings: Micro evidence from urban china. *National Bureau of Demographic Dividends Revisited*, *21*.

Baum, C. F. (2007). ivreg2: Stata module for extended instrumental variables/2sls, gmm and ac/hac, liml and k-class regression. *http://ideas. repec. org/c/boc/bocode/s425401. html*.

Boldrin, M., & Jones, L. E. (2002). Mortality, fertility, and saving in a malthusian economy. *Review of Economic Dynamics*, *5*(4), 775–814.

Cai, Y. (2010). China's below-replacement fertility: government policy or socioeconomic development? *Population and Development Review*, *36*(3), 419–440.

Caldwell, J. C. (1978). A theory of fertility: from high plateau to destabilization. *Population and development review*, 553–577.

Chen, X., Smith, J., Strauss, J., Wang, Y., & Zhao, Y. (2017). China Health and Retirement Longitudinal Study (CHARLS). *Encyclopedia of Geropsychology*, 463–469.

Chen, Y., & Fang, H. (2019, May). *The Long-Term Consequences of Family Planning in Old Age: Evidence from China's "Later, Longer, Fewer" Campaign* (Working Papers No. 34). University of Pennsylvania Population Center.

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

Chen, Y., Li, H., & Meng, L. (2013). Prenatal sex selection and missing girls in china: Evidence from the diffusion of diagnostic ultrasound. *Journal of Human Resources*, *48*(1), 36–70.

Choukhmane, T., Coeurdacier, N., & Jin, K. (2013). The one-child policy and household savings.

Clark, S. (2000). Son preference and sex composition of children: Evidence from india. *Demography*, *37*(1), 95–108.

Coale, A. J., & Chen, S. L. (1987). Basic data on fertility in the provinces of china, 1940-82.

Davis, D. (1991). *Long lives: Chinese elderly and the communist revolution.* Stanford University Press.

Ebenstein, A., & Leung, S. (2010). Son preference and access to social insurance: evidence from china's rural pension program. *Population and Development Review*, *36*(1), 47–70.

Fang, H., & Feng, J. (2018). *The chinese pension system* (Tech. Rep.). National Bureau of Economic Research.

Ge, S., Yang, D. T., & Zhang, J. (2018). Population policies, demographic structural changes, and the chinese household saving puzzle. *European Economic Review*, *101*, 181–209.

Greenhalgh, S. (2008). *Just one child: Science and policy in deng's china*. University of California Press.

Gu, H. (2009). *Chinese educated city youth: The whole story (in chinese)*. People's Daily Publishing House.

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

He, X., & Zhu, R. (2016). Fertility and female labour force participation: Causal evidence from urban c hina. *The Manchester School*, *84*(5), 664–674.

Ho, D. Y.-F. (1994). Filial piety, authoritarian moralism and cognitive conservatism in chinese societies. *Genetic, Social & General Psychology Monographs*, *120*(3), 349.

Huang, W., Lei, X., & Sun, A. (2016). Fertility restrictions and life-cycle outcomes: Evidence from the one-child policy in china. *Review of Economics and Statistics*, 1–41.

Kwan, K.-L. K. (2000). Counseling chinese peoples: Perspectives of filial piety. *Asian Journal of Counseling*, *7*(1), 23–41.

Li, B., & 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., & 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., & Zhang, J. (2007). Do high birth rates hamper economic growth? *The Review of Economics and Statistics*, *89*(1), 110–117.

Li, H., Zhang, J., & Zhu, Y. (2008). The quantity-quality trade-off of children in a developing country: Identification using chinese twins. *Demography*, *45*(1), 223–243.

Li, H., & Zheng, H. (2009). Ultrasonography and sex ratios in china. *Asian Economic Policy Review*, *4*(1), 121–137.

Li, L., & Wu, X. (2017). The consequences of having a son on family wealth in urban c hina. *Review of Income and Wealth*, *63*(2), 378–393.

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.

Mauldin, W. P. (1982). The determinants of fertility decline in developing countries: An overview of the available empirical evidence. *International Family Planning Perspectives*, 116–121.

Miller, G. (2010). Contraception as development? new evidence from family planning in colombia. *The Economic Journal*, *120*(545), 709–736.

Modigliani, F., & Cao, S. L. (2004). The chinese saving puzzle and the life-cycle hypothesis. *Journal of economic literature*, *42*(1), 145–170.

Oliveira, J. (2016). The value of children: Inter-generational support, fertility, and human capital. *Journal of Development Economics*, *120*, 1–16.

Rosenzweig, M., & Zhang, J. (2014). *Co-residence, life-cycle savings and inter-generational support in urban china* (Tech. Rep.). National Bureau of Economic Research.

Sanderson, E., & Windmeijer, F. (2016). A weak instrument f-test in linear iv models with multiple endogenous variables. *Journal of econometrics*, *190*(2), 212–221.

Scharping, T. (2013). *Birth control in china 1949-2000: Population policy and demographic development*. Routledge.

Song, Z., Coupé, T., & Reed, W. R. (2020). Estimating the effect of the one-child policy on chinese household savings-evidence from an oaxaca decomposition. *China Economic Review*, 101570.

Sung, K.-T. (1998). An exploration of actions of filial piety. *Journal of Aging Studies*, *12*(4), 369–386.

Tertilt, M. (2005). Polygyny, fertility, and savings. *Journal of Political Economy*, *113*(6), 1341–1371.

Walder, A. G. (2017). *China political events dataset, 1966-1971*. Retrieved from `https://stanford.app.box.com/s/1p228gewy2pjd3817ksq9kd4d6cz3jy8`

Wang, F. (2016). *Using New Measures to Reassess the Fertility Effects of China's Family Planning Policies* (Working Papers). Renmin University of China, School of Labor and Human Resources.

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

White, T. (2009). *China's longest campaign: Birth planning in the people's republic, 1949–2005*. Ithaca: Cornell University Press. Retrieved from `https://www.degruyter.com/cornellup/view/title/552318` doi: https://doi.org/10.7591/9781501726583

Whyte, M. K. (Ed.). (2003). *China's revolutions and intergenerational relations*. University of Michigan Press. Retrieved from `http://www.jstor.org/stable/10.3998/mpub.19840`

Whyte, M. K., Feng, W., & Cai, Y. (2015). Challenging myths about china's one-child policy. *The China Journal*(74), 144–159.

Wu, X., & Treiman, D. J. (2004). The household registration system and social stratification in china: 1955–1996. *Demography*, *41*(2), 363–384.

Yamaguchi, K. (1989). A formal theory for male-preferring stopping rules of childbearing: Sex differences in birth order and in the number of siblings. *Demography*, *26*(3), 451–465.

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.

Zhao, Y., Strauss, J., Yang, G., Giles, J., Hu, P., Hu, Y., … Wang, Y. (2013). *China Health and Retirement Longitudinal Study 2011–2012 national baseline users' guide*. Beijing, China: National School of Development, Peking University.

Table A1: **Family Planning Leading Group Establishment Year**

| Leading group | Provinces | Percentage |
|---|---|---|
| 1969 | Guangdong | 4.17% |
| 1970 | Shandong | 4.17% |
| 1971 | Gansu, Hubei, Hunan, Shaanxi, Shanxi, Sichuan, Zhejiang | 29.17% |
| 1972 | Anhui, Fujian, Hebei, Heilongjiang, Inner Mongolia, Jiangxi, Jilin, Liaoning, Yunnan | 37.50% |
| 1973 | Henan, Jiangsu | 8.33% |
| 1974 | Guangxi, Qinghai | 8.33% |
| 1975 | Guizhou, Xinjiang | 8.33% |
| Total | | 100% |

Data sources: Y. Chen and Huang (2020).

Figure A1: TFR trend

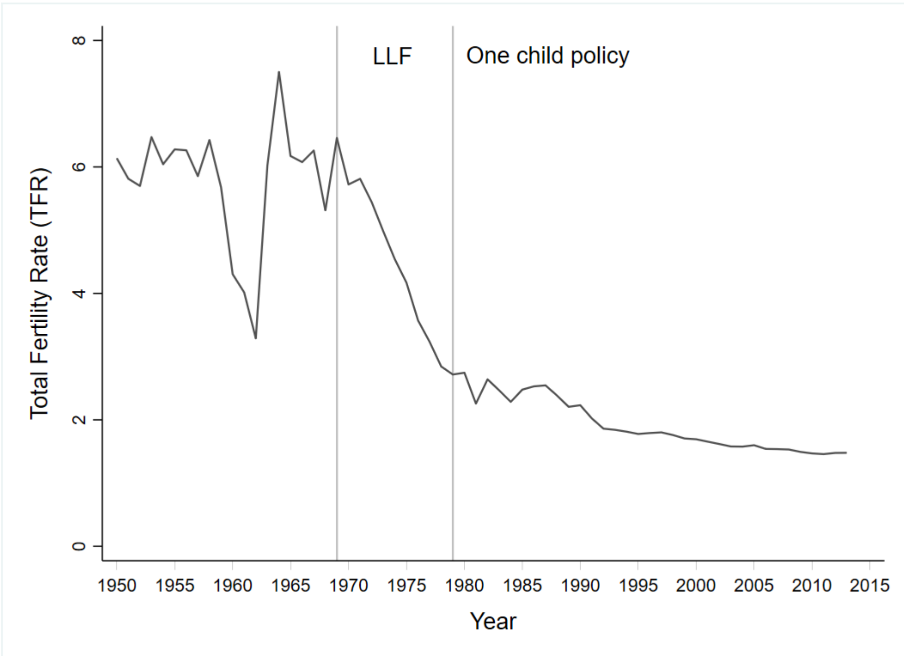



Table A2: **Distribution of Age at Enforcement**

| Age at Enforcement | Percentage |
|---|---|
| -24 | 39.38% |
| 25-29 | 28.76% |
| 30-34 | 17.61% |
| 35-39 | 9.50% |
| 40-44 | 4.19% |
| 45+ | 0.56% |
| Total | 100% |

Table A3: Descriptive statistics

| Variables | Mean | S.D. | OBS |
|---|---|---|---|
| **Basic characteristics** | | | |
| Age | 70.96 | 6.40 | 18,453 |
| Age gap | 3.35 | 3.13 | 18,453 |
| Mother's birth year | 1944.11 | 5.84 | 18,453 |
| Male | 0.44 | 0.50 | 18,453 |
| Residence rural | 0.62 | 0.49 | 18,453 |
| First hukou rural | 0.91 | 0.29 | 18,453 |
| Current hukou rural | 0.76 | 0.43 | 18,453 |
| Educational attainment | | | |
| -illiterate | 0.37 | 0.48 | 18,453 |
| -primary school | 0.45 | 0.50 | 18,453 |
| -secondary school | 0.11 | 0.31 | 18,453 |
| -high school and above | 0.07 | 0.25 | 18,453 |
| Couple's highest educ | | | |
| -illiterate | 0.20 | 0.40 | 18,453 |
| -primary school | 0.52 | 0.50 | 18,453 |
| -secondary school | 0.17 | 0.38 | 18,453 |
| -high school and above | 0.11 | 0.31 | 18,453 |
| Widowed | 0.23 | 0.42 | 18,453 |
| Nr. chronic diseases | 1.70 | 1.59 | 18,453 |
| Nr. (I)ADL | 1.40 | 2.36 | 18,453 |
| Wealth | 250,906 | 2994012 | 18,453 |
| Pension | 6448 | 13772 | 18,453 |
| **Labor supply** | | | |
| *Currently working* | | | |
| 60-64 | 0.65 | 0.48 | 3,153 |
| 65-69 | 0.58 | 0.49 | 5,406 |
| 70-74 | 0.46 | 0.50 | 4,689 |
| 75+ | 0.25 | 0.43 | 5,205 |
| Total | 0.47 | 0.50 | 18,453 |
| *Condi. on currently working* | | | |
| Hours/week | 36.57 | 23.77 | 7,786 |
| More than part-time | 0.70 | 0.46 | 7,786 |
| **Fertility outcomes** | | | |
| Number of children | 3.55 | 1.49 | 18,453 |
| Ratio of sons | 0.54 | 0.28 | 18,453 |

*Notes:* Wealth and working hours are logarithmic transformed in regression analyses.

Table A4: **Net material support from children**

| | Total | | | Regular | | |
|---|---|---|---|---|---|---|
| | Incidence | Amt | Amt per child | Incidence | Amt | Amt per child |
| *Panel A* | | | | | | |
| Nr. Children | 0.062** | 0.262** | -0.495*** | 0.047 | 0.024 | -0.477*** |
| | (0.030) | (0.121) | (0.085) | (0.030) | (0.170) | (0.079) |
| Son Ratio | -0.054 | 0.115 | 0.242* | 0.091** | 0.035 | 0.142 |
| | (0.044) | (0.180) | (0.137) | (0.041) | (0.304) | (0.148) |
| Adjusted R-squared | 0.16 | 0.06 | 0.53 | 0.07 | 0.01 | 0.51 |
| Hansen J test p-value | 0.19 | 0.08 | 0.73 | 0.58 | 0.43 | 0.15 |
| F-stat 1st stage: Nr. Children | 22.17 | 17.38 | 17.38 | 22.17 | 13.42 | 13.42 |
| F-stat 1st stage: Son Ratio | 31.96 | 29.36 | 29.36 | 31.96 | 23.30 | 23.30 |
| *Panel B* | | | | | | |
| Nr. Daughters | 0.069*** | 0.259** | -0.543*** | 0.029 | 0.032 | -0.514*** |
| | (0.025) | (0.101) | (0.068) | (0.025) | (0.139) | (0.068) |
| Nr. Sons | 0.047 | 0.307** | -0.476*** | 0.057 | 0.057 | -0.488*** |
| | (0.035) | (0.135) | (0.096) | (0.035) | (0.199) | (0.093) |
| Adjusted R-squared | 0.16 | 0.05 | 0.53 | 0.07 | 0.01 | 0.51 |
| Hansen J test p-value | 0.27 | 0.12 | 0.50 | 0.36 | 0.46 | 0.11 |
| F-stat 1st stage: Nr. Dau | 24.13 | 19.34 | 19.34 | 24.13 | 14.05 | 14.05 |
| F-stat 1st stage: Nr. Son | 23.72 | 19.53 | 19.53 | 23.72 | 14.24 | 14.24 |
| Observations | 18,453 | 13,000 | 13,000 | 18,453 | 4,631 | 4,631 |

*Notes:* * significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses. Standard errors are clustered at the province-cohort level. Other controls, age FE, cohort group FE, province FE, province-specific cohort trend and year FE are included everywhere. The amounts of material transfers are natural logarithmic transformed to correct for the right skewed distribution.