Migration Reform and Fertility: Evidence from Rural China*

Wenchao Jin, Zhangfeng Jin

September 9, 2026

Abstract

How do institutional barriers to migration shape fertility in developing economies? We analyze the staggered removal of institutional barriers to rural-to-urban migration across 283 Chinese cities. We find that reducing these frictions led to a significant and persistent increase in fertility in *sending* rural communities. The average treatment effect is 0.011 newborns per household per year, representing approximately one-third of the sample mean. To interpret this result, we develop a unified household model endogenizing fertility and partial migration. The model identifies a positive income effect (higher expected lifetime earnings) that dominates the substitution effect (opportunity cost of time). Empirically, we show that the fertility response is concentrated in households with available grandparents and prior migration experience. This suggests that informal childcare provision is critical in neutralizing the time costs of migration, allowing rural households to realize the fertility gains from improved economic opportunities. These findings challenge the view that urbanization necessarily reduces fertility, highlighting instead how mobility restrictions acted to suppress fertility.

JEL classification: R23; J13; J22; J24
Keywords: migration barriers; fertility; China

*Jin: Business School, University of Sussex, (email: Wenchao.Jin@sussex.ac.uk); Jin (corresponding author): School of Economics, Zhejiang University of Technology and Global Labor Organization, 288 Liuhe Road, Hangzhou, China, 310000 (email: zhangfeng1019@gmail.com). We gratefully acknowledge the valuable comments and insights from Cevat Aksoy, Shuaizhang Feng, Clément Imbert, Mushfiq Mobarak, Rufei Guo, Klaus F. Zimmermann, Hillel Rapoport, and the other participants at the 8th IESR-GLO Joint Workshop on Fertility Decline and Family Policies, as well as the 18th AFD-World Bank Migration and Development Conference. This work was supported by the Fundamental Research Funds for the Provincial Universities of Zhejiang (Grant No. G26221310012). We declare no relevant or material financial interests related to the research described in this article.

1 Introduction

Internal migration is a dominant demographic force in developing economies, yet its relationship with fertility remains incompletely understood because the literature has focused almost exclusively on migrants in urban destinations. By restricting analysis to migrants, existing studies not only face an endogeneity challenge but also overlook a critical demographic reality: in many developing contexts, the vast majority of the rural-born population remains in rural areas. This paper addresses this gap by providing the first causal evidence on how reducing migration barriers affects fertility in *sending* rural regions. By shifting the focus from the destination to the source, we capture the impact of migration policy on the entire rural population—including non-migrants and left-behind families—rather than just the selected minority that migrate.

We examine this question in the context of China, where the Hukou (household registration) system has long acted as a significant friction to labor mobility. For decades, the Hukou system decoupled the ability to work in a city from the right to access local public services. Beginning in the early 2000s, a wave of reforms across 283 prefecture-level cities significantly reduced these barriers, replacing rigid quotas with more flexible entry criteria based on stable employment and housing. These reforms lowered the uncertainty and cost of future urban settlement, effectively raising the "option value" of migration for all rural households.

Using an event-study design (de Chaisemartin and D'Haultfoeuille, 2020) on a large panel of rural households, we find that removing these migration barriers led to a significant and persistent increase in fertility in rural areas. Contrary to the "modernization hypothesis"—which typically associates urbanization with fertility decline—our estimates show that the reform increased the likelihood of a rural household having a newborn by 1.1 percentage points per year, approximately one-third of the sample mean. This effect persists over the subsequent decade and is robust to a battery of checks for heterogeneous treatment effects and selective attrition.

To interpret this counter-intuitive finding, we develop a unified household model that endogenizes both fertility decisions and partial migration choices. The model highlights that migration reform affects fertility through competing channels: a positive income effect (higher expected lifetime earnings), a negative substitution effect (higher opportunity cost of time due to rising female wages), and a positive effect from lower price of raising children in urban destinations (gaining access to local schools). Our empirical tests are most consistent with a leading role for the income channel, while the price channel remains theoretically plausible but harder to isolate directly in our data. The reform significantly increased household consumption and male wages. Moreover, we find that the fertility response is concentrated in multi-generational households and those with healthy older adults. This suggests that the availability of informal childcare (grand-

parents) neutralizes the substitution effect, allowing households to realize the fertility gains from higher income.

This paper makes three main contributions to the development economics literature. First, we provide the first evidence on how migration liberalization affects fertility in *sending communities*, filling a critical gap in the literature. Existing studies focus on the population of migrants, those who have already moved to the receiving destinations or were induced to move by an instrument. Some found a positive fertility impact of pro-migration policies, as improved migration opportunities raise household income prospects, childbearing becomes more financially feasible (Amuedo-Dorantes *et al.*, 2023a; Lanari *et al.*, 2020; Dong *et al.*, 2023; Amuedo-Dorantes and Arenas-Arroyo, 2021). Others (Amuedo-Dorantes *et al.*, 2023b; Avitabile *et al.*, 2014; Jensen and Ahlburg, 2004) found opposite effects and point to either the substitution effect, whereby greater labor market opportunities increase the opportunity cost of child-rearing; or to the quality–quantity trade-off, whereby households choose to invest more in quality as improved migration opportunities enhance the expected returns to human capital.

Focusing on migrants misses the vast majority of the population that remains in rural areas. Figure 1 illustrates different groups of the originally-rural population. Our calculation (in Appendix 9.1) suggests that in 2010, of all rural-origin population in China, 42 percent were in non-migrant households, 48 percent were in households with migrating members (among which 20 percent were migrants and 28 percent were left-behind members), and 10 percent were "independent migrants", those who have no family members left behind in their rural home or are not considered family members any more by their rural relatives. This means migrants constitute less than 30% of the rural-origin population in China. By using a representative rural panel survey that captures non-migrants, left-behind families, and connected migrants, our analysis encompasses 90% of the relevant population, offering a more comprehensive understanding of how migration barriers affect fertility.

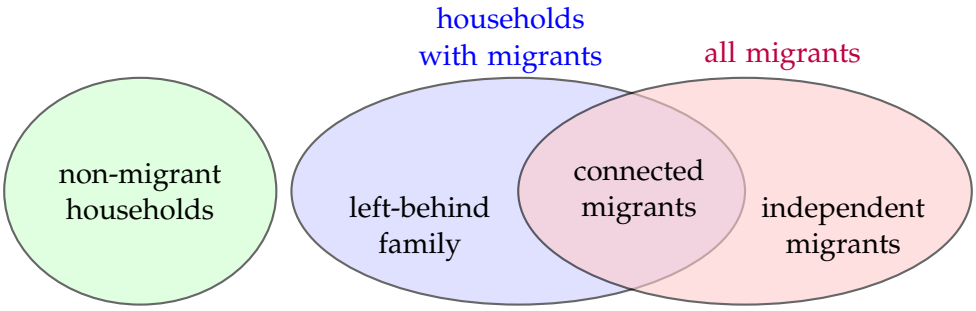


Figure 1: Groups of originally-rural population

Second, we contribute to the evolving literature on the relationship between migration and

fertility in developing economies. While urbanization is widely stylized as a driver of fertility decline due to higher housing and opportunity costs (Jensen and Ahlburg, 2004; White *et al.*, 2008; Guo *et al.*, 2012; Lee, 2025), the impact of *access* to migration on sending communities is theoretically ambiguous. Migration can depress rural fertility through spousal separation and the adoption of urban norms, but it can also stimulate fertility by relaxing household budget constraints through remittances and higher expected earnings (Naufal, 2015; Anelli and Balbo, 2021). We resolve this ambiguity by showing that removing Hukou barriers generates a positive income shock that releases suppressed fertility. This challenges the view that urban exposure necessarily accelerates fertility decline.

Third, we contribute a novel theoretical framework to the literature on household decision-making. We present the first economic model to endogenize both fertility decisions and *partial* migration choices. While previous studies have examined the joint determination of fertility and location (Mookherjee *et al.*, 2012; Tumen, 2012), they typically abstract from the option of "parent-only" migration and intergenerational co-residence. By explicitly modeling these features, we theoretically disentangle the competing income and substitution effects of migration reform. This framework reveals that extended family structures play a critical economic role: without the childcare support provided by grandparents, the opportunity costs of childrearing would likely dampen the positive fertility effects of migration access. This model is readily applicable to other developing contexts where partial migration and multi-generational households are common.

The rest of this article is organized as follows. Section 2 explains the Hukou system and the reform under investigation. Section 3 describes data sources and summary statistics. Section 4 details the event study specification and identification strategy. Section 5 presents our main findings and robustness checks. Section 6 introduces the theoretical framework endogenizing fertility and migration. Section 7 provides empirical evidence on the channels. The final section concludes.

2 Institutional background

2.1 The Hukou system

The Chinese household registration system (Hukou system), formally established in 1958, initially restricted internal migration by tying access to employment and social services to one's registered location. Individuals were classified at birth as either agricultural (AH) or non-agricultural Hukou (NAH) holders. AH holders were usually registered in rural areas, while NAH holders were usually registered in urban areas. In the first two decades of the adoption of the Hukou system, rural-urban migration was strictly controlled. For instance, without obtaining

a migration permit, rural individuals were not allowed to move to urban destinations for work. Given that an individual's Hukou was registered at the subdistrict level (including streets, towns and townships) within each city, the Hukou system not only strictly regulated migration between cities but also controlled migration from rural to urban areas within the same city. In this paper, the term 'city' refers to the official administrative definition of cities, which contain both urban centres and surrounding rural areas.

The first phase of relaxation (1980*s*-1990*s*) responded to dual pressures: urban labor shortages after market liberalization and rural labor surpluses following the Household Responsibility System's adoption in the agricultural sector. While temporary migration became permitted, migrants faced exclusion from local public services (including public education and social insurance) in urban destinations unless they obtained a local Hukou–a process governed by strict annual quotas and eligibility criteria favouring younger and educated applicants.

A second phase began in the late 1990*s*, when the government eased Hukou access in suburban areas of cities. But access to a local Hukou in most urban centres of cities remained tightly controlled. Then, from 2002, this is extended gradually to the remaining urban areas of prefecture-level cities including densely-populated areas. The common theme was to abolish quotas, remove age/education requirements, and standardise baseline criteria (stable housing and employment). In July 2014, the national government announced a national roll-out of the reform, mandating simplified Hukou access in all cities except a handful of mega-cities like Beijing and Shanghai. The staggered adoption of this phase of Hukou reform over 2002-2014 will be the subject of investigation in this paper.

To investigate this phase of reform, we collected data on migration policies implemented by local (prefecture-level) governments since the early 2000*s*.[1] We compiled data on the adoption dates of these reforms across 283 prefecture-level cities[2] from 2002 to 2015, sourcing information from official records, reputable news outlets, and specialised legal databases focusing on Hukou reform. This comprehensive approach allowed us to pinpoint the exact timeline for each city.

A key element of the Hukou reform was the transition from a rigid annual quota system to a more flexible, quota-free system. This new framework introduced basic criteria for local Hukou applications, removing the previous quotas on the number of migrants eligible for Hukou and eliminating age and educational (skill) requirements. These changes significantly improved the prospects for migrants from rural areas. After the reform adoption, the general prerequisites for obtaining Hukou included stable housing and employment or a reliable income source. The specific definitions of stable housing and employment varied by city. For instance, in cities

[1] In our context, provincial-level governments typically set general policy guidelines, while prefecture-level governments are responsible for working out the details and implementing the Hukou reforms at the local level.

[2] China is officially divided into 283 prefecture-level cities during our sample period.

like Nanjing, public rental housing might qualify, whereas in cities like Lianyungang, property ownership could be required. Most cities recognised employment in government entities, state-owned enterprises, joint ventures, or private firms with formal employment contracts as stable employment. Some cities also required a minimum length of employment contracts or prior contributions to local social security schemes. Figure 2 illustrates the timing of migration reform adoption across cities. 30% of cities adopted the policy by 2004. The share increased gradually to 61% by 2014, and in 2015 the policy was rolled out nationally.

Cities adopted the Hukou reform at different times, and this timing may be correlated with their economic and demographic conditions. The non-random timing is a potential threat to identification. To address this concern, we collect a set of economic and social characteristics for prefecture-level cities in 2001. These variables are selected based on mechanisms emphasized in the migration and local policy adoption literature, including labor market conditions, demographic structure, fiscal capacity, industrial composition, and migration patterns (Casarico *et al.*, 2018; Jin, 2024; Tian, 2024; Chen *et al.*, 2025). In particular, we include measures such as urban unemployment rate, rural and urban age structure (e.g., elderly population share and the share of younger cohorts), fiscal expenditure structure, fertility, state ownership, and migration rates. Appendix 9.6 describes the data sources, the definition of city-level controls, and reports their summary statistics.

Table 27 presents pairwise correlation between these initial conditions and the timing of reform adoption. The results suggest that reform timing was not entirely random and was systematically associated with several pre-existing characteristics, particularly those related to labor market conditions, fiscal capacity, and demographic structure. We further examine this relationship in a multivariate regression framework by regressing reform timing on initial city characteristics and identify a set of baseline factors that predict the timing of reform adoption (Table 28). The non-random timing of reform is addressed more formally in section 4, where we discuss the identification strategy and present additional evidence based on alternative specifications controlling for initial city conditions.

The reform over the 2000*s* has been shown to significantly increase rural-urban migration. Jin and Zhang (2023) estimated a 24.8% rise in inflows, with the impact being more pronounced for young and low- and medium-skilled workers. Moreover, the impact persisted and actually increased over 6 years (the time window examined in Jin and Zhang (2023)). Meanwhile, the reform has significantly increased the level of agricultural mechanization, facilitated the subcontracting of land by low productivity farmers, and improved Total Factor Productivity in agriculture (Ji and Nie, 2024).

Rural-to-urban migration was accelerating over our sample period. According to Gao *et al.*

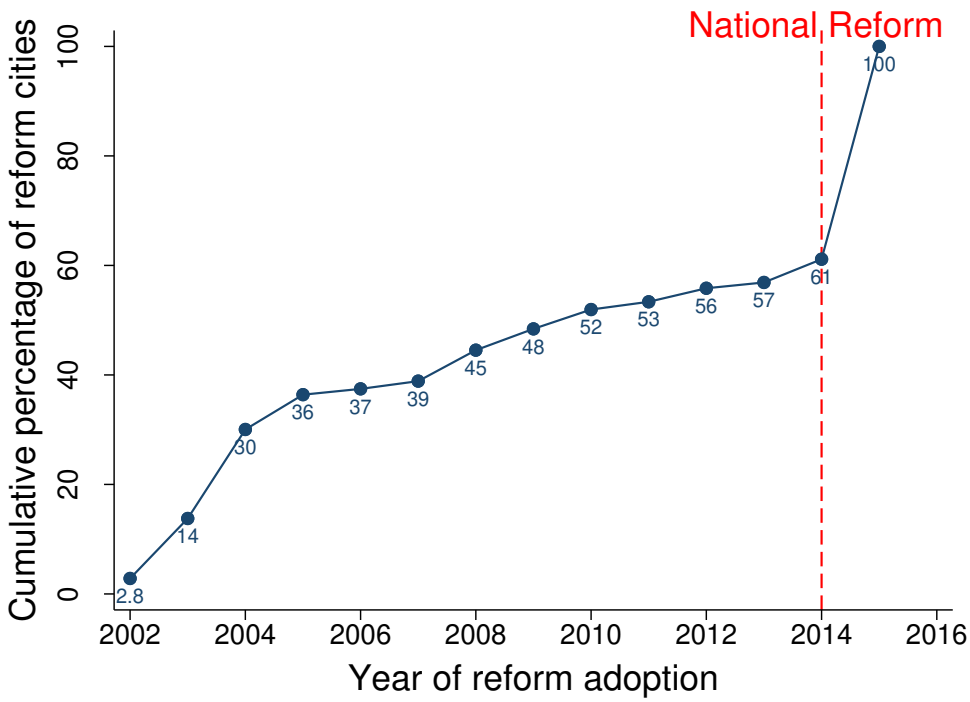


Figure 2: The Adoption of Migration Reform Over Time

Notes: This graph depicts the adoption of the migration reform over time. We assume that, following the announcement of the nationwide reform, the policy was de jure extended to all remaining eligible cities.

(2023), rural-urban migrants constituted 11% of the Chinese (rural) population in the 2005 census. By 2015, this share had increased to approximately 17%. In addition to Hukou reform, economic factors such as export-led growth and enhanced land tenure security were also significant contributors (Facchini *et al.*, 2019; Adamopoulos *et al.*, 2024).

2.2 Fertility and related policies

China's total fertility rate declined from 5.8 in 1970 to 2.75 in 1979 and stabilised at around $1.2 - 1.3$ between 2010 and 2020 (Yang *et al.*, 2022). As the population aged, concerns over low fertility grew, prompting the government to introduce a partial two-child policy in 2014, a universal two-child policy in 2016 and a three-child policy in 2021. However, these policies failed to significantly increase birth rates.

In rural China, fertility fell dramatically from above six births per woman in the 1960s to below 2 by the mid-90s and remained in the $1.5 - 2.0$ region through to 2010 (Zhang, 2017). Much of this decline occurred before the One-Child Policy (OCP) that was formally introduced in 1979, largely due to the earlier birth planning campaign known as "Later, Longer, and Fewer."

(Chen and Fang, 2021).

The strictness of OCP varied between rural and urban areas, across provinces, and over time (Zhang, 2017). In the mid-1980s, the government introduced a long list of exemptions for rural families, allowing second births under specific conditions. The most notable condition is if the first child was female. Moreover, enforcement was generally weaker in rural areas, as rural residents were less dependent on the state for jobs or earnings, and financial penalties were often harder to enforce on households with very little assets.

The majority of births reported in our sample were likely within the range allowed by policy. According to Guo *et al.* (2025), the policy in rural China restricted births at third and higher parities, but not at the second parity. In our sample, among all the newborns, 58% are the only child in their generation, 34% are the 2nd. This means at most 8% are the 3rd or higher-order child, and at most 42% are 2nd or higher-birth-order child. These are upper-bounds because we only observe individuals' relationship to the head of the household, not to each other. For instance, in a 3-generation household, the middle generation could have multiple couples and we cannot tell whether the grandchildren have the same parents.

Children of rural migrants were often excluded from urban public education unless their parents had obtained local Hukou (Chen and Feng, 2013). This means these children either have to go to school in the village of their rural Hukou, or attend lower-quality fee-paying schools designated for migrants' children (Chen and Feng, 2017).[3] High housing costs and limited child-care options in urban destinations further increased the cost of raising children in urban areas. According to Fang and Shi (2018), among children of rural-to-urban migrants in 2005, three quarters of those aged 0-14 were left behind; and the proportion of migrants' children left behind has declined during the early 2010s. The decline might be partly a result of Hukou reforms since the early 2000s. As the reform increased the chance of migrants obtaining the local urban Hukou, it reduced the costs of educating children in urban areas and improved the expectation of children's future incomes. This improvement in educational opportunities could have a positive impact on fertility. Section 6 will illustrate this intuition in a formal model, alongside other factors relevant for fertility choice.

3 Data and Summary Statistics

Our primary dataset is the National Fixed Point Survey (NFPS), a panel survey collected by the Research Center of Rural Economy (RCRE) of China's Ministry of Agriculture since 1986. The

[3] According to Chen and Feng (2017), the student-teacher ratio in public schools in urban areas is roughly half that of migrant schools. Moreover, teachers in public schools tend to be more experienced, have longer tenures at their current schools, possess higher levels of education, and receive higher salaries.

NFPS survey tracks a nationally representative sample of approximately 20,000 rural households from around 300 villages across all 31 mainland provinces of China. The survey is divided into three sub-surveys: the village survey, the household survey and the household member (individual) survey. The NFPS villages were selected based on region, income, cropping pattern, population, and non-farm activity to ensure representativeness. Within each village, around 50 to 100 households are randomly chosen to participate in the annual survey. These households maintain detailed diaries of income, expenditure, and labor activities, which are collected monthly by resident enumerators. The rigorous data collection and verification procedures minimize reporting and measurement errors (Benjamin *et al.*, 2005; Chen and Zhang, 2009).

The individual module within NFPS collects demographic data and labor supply. This includes variables such as gender, age, educational background, health status and days spent on different categories of work. These allow us to examine changes in household structure and labor supply following the implementation of migration policies. We match the individual survey with the household survey based on household-level ID to conduct the main analysis. We have access to the NFPS from 2003 to 2015. Before 2003, the NFPS collected household-level information only, and not individual-level information such as age.

Our data contain basic demographic information on *all* individuals within the household, whether the individual is physically present or not. An individual could be living outside the village for most of the year due to work, but as long as they send some income back home, they would still be reported as members of the household. By contrast, individuals who live elsewhere and maintain independent economic lives, such as an adult daughter married into a household in a different village or an urban area, would not be included. In our sample, about half of households had at least one person living at home for under 180 days.

Our main analysis is at the household-year level. There are 202,806 household-year observations over 2003-2015. Table 1 reports summary statistics of key variables. We define 'newborn' as a person aged 0 or 1 and has not been observed before in the same household. In rural China, the traditional way of counting age starts from 1 rather than 0. Hence we count individuals whose reported age is 1 as newborns as long as they did not appear as a 0-year-old in the preceding year. In our sample at the household-year level, 2.7% of observations reported a newborn.

On average, each household has 3.2 adults and 0.7 children (Table 1). This reflects the prevalence of multi-generational living arrangements, which we will explore shortly in Table 2. Table 1 shows the average number of days worked in local agriculture, local non-agriculture, and outside the local area.[4] Non-local work constitutes over one third of total labor supply from these rural

[4]The questionnaire used the term "township and town" as a single unit. Thus the term "local" refers to the respondent's registered township and town. Unlike the definitions used in developed countries such as the US or EU—which emphasize population density, settlement size, and functional linkages—the statistical classification of townships and

Table 1: Summary statistics at household-year level

variable	observations	mean	median	min	max
has newborn	202806	0.027	0.000	0.000	1.000
number of adults	202806	3.182	3.000	0.000	11.000
number of children	202806	0.670	0.000	0.000	9.000
whether head of household is male	201404	0.934	1.000	0.000	1.000
age of head	201223	53.616	54.000	18.000	97.000
years of education of the head	195874	6.809	7.000	0.000	22.000
any person not living at home	202806	0.495	0.000	0.000	1.000
mean outcome among adults in the household					
male	202780	0.517	0.500	0.000	1.000
age	202636	45.863	43.500	18.000	96.000
years of education	200229	7.050	7.000	0.000	22.000
total number of days worked	202417	171.756	175.000	0.000	365.000
days worked in local agriculture	202803	67.074	43.333	0.000	365.000
days worked in local non-agricul	202804	41.638	0.000	0.000	365.000
days worked non locally	202794	66.790	30.000	0.000	365.000
number of days lived at home	199364	294.569	341.250	0.000	365.000

Note: This table is based on our cleaned sample over 2003-2015, at the household-year level. We drop households where everyone is under 18 or the head is under 18. We drop households with multiple heads. Adults refer to anyone aged 18+. 'Children' refer to anyone below age 18. A small fraction of households in the sample do not have any 'head of household' and hence are missing the gender and age of the head. Missing values in 'number of days of worked' are due to reported values below zero or above 365. 'Any person not living at home' means the household contained a person whose number of days living at home under 180.

households.

The mean number of days of living at home last year was 294 among adults, suggesting that at the individual level, migrants were a substantial minority rather than the majority. Because these households contain 3+ adults on average, the proportion of households having at least one migrant is high. 49.5% of the sample contained at least one person living at home for under 180 days.

Table 2 splits the sample by household type. The most common type of household is parents

towns in China is rooted in administrative hierarchies and the extent of physical construction. A township (xiang) is a township-level administrative unit under a county, consisting mainly of dispersed rural villages and a small administrative seat that is typically too small and sparsely built to be classified as urban. In contrast, towns (zhen) include county seats, larger township centers, and independent settlements with at least 3,000 permanent residents, provided that their built-up areas are contiguous with the respective administrative seats. Importantly, only the built-up core of a town is classified as urban, while the surrounding rural villages under the same town's jurisdiction remain classified as rural. According to the 2010 Population Census, 87% of individuals living in township and town jurisdictions held agricultural Hukou, with the proportions being 96% for townships and 64% for towns, respectively. This highlights that, despite being labeled "towns," many of these areas still retain predominantly rural characteristics.

Table 2: Summary statistics by household type

household type	observations	mean of number of adults	number of children	has any newborn	age of head	has any migrant
single or couple only	41771	1.849	0.003	0.000	61.718	0.112
parents+ children$<$18	22479	1.983	1.546	0.017	39.335	0.424
parents+adult children	67809	3.416	0.235	0.001	51.068	0.662
three generations	65850	4.075	1.202	0.070	56.001	0.582
4 or 5 generations	3232	5.155	1.261	0.123	53.495	0.592
other	1665	4.155	1.062	0.083	55.053	0.712

Note: This table shows the number of observations and average characteristics by household type. The data is at the household-year level. Adults refer to anyone aged 18+. 'Children' refer to anyone below age 18. 'Any migrant' is a binary indicator for whether the household contained any member who lived at home for less than 180 days. 'Other' means all other households except the aforementioned ones: for example, households that contain siblings of the head of household, or nephews/nieces of the head. Household type is constructed on the basis of all individuals' relations to the head and their ages.

plus adult children; followed by three-generation households. Each of them constitutes about one third of all households. The third most common type of households are single-generation households. They are mostly elderly couples or single elders living alone. Nuclear families (parents plus children under 18) constitute only 11% of the sample.

Unsurprisingly, household structure is closely tied to fertility patterns. Among three-generation households, 7% have newborns. By contrast, in nuclear families (parents plus children under 18 only), only 1.7% have newborns. In fact, of all newborns in the sample, 84% were born in three-generation households and 7% were born in four-five-generation households. This reflects the prevalence of multi-generational co-residence and it means that prospective grandparents could be an important source of childcare provision. Grandparents in the sample are typically not very old, the median age of household head is 56 in three-generation households. Later in subsection 7.1 we will examine the impact of the reform by the presence of healthy prospective grandparents. The survey asks respondents to categorize health status into 5 categories, indexed 1 to 5 from the best to the worst. As shown in Table 19 in Appendix 9.2, among both genders over 45, the median health status is 2 for 'good', the 2nd category; and the mean is also close to 2.

In recent years, the share of households having newborns has declined in our sample. Figure 3 shows that it used to be around 3% in the 2000s, and has started declining around 2010, to 2.1% by 2015. The decline is driven by first-borns (shown in red), rather than higher-order births (shown in green). The decline of first-borns could be due to postponement of marriage or having a first child, or simply a decline in the share of young people. Figure 13 in Appendix shows that

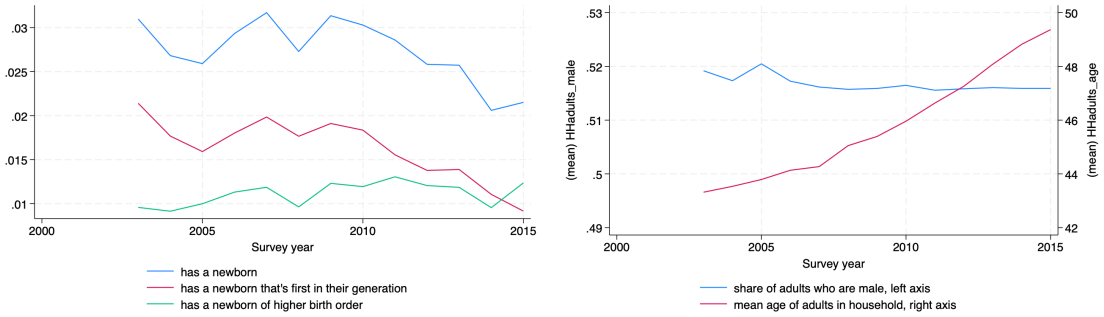


Figure 3: Rural demographic composition over time
Notes: the sample is all households in cleaned data as described in Table 1.

the share of households with a woman under 40 also started to fall around 2010. This can be attributed to the aging of the rural population rather than gender-biased migration (right sub-figure in Figure 3). There was huge variation in cohort size in rural China, because the Total Fertility Rate fell from over 6 births per woman in 1970 to around 3 in 1980 (Zhang, 2017).

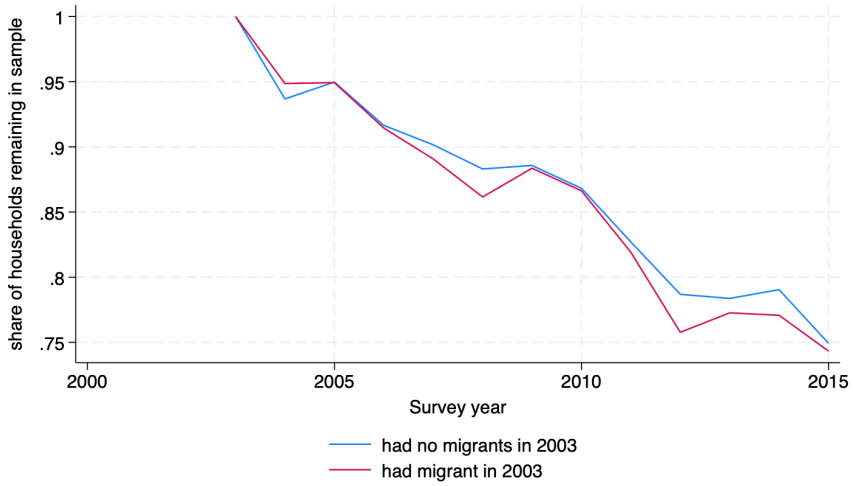


Figure 4: Sample attrition by initial status of migrant work

Notes: This graph illustrates what percentage of rural households present in 2003 remained in the sample over time. It differentiates households by whether it contained a person working non-locally for 180+ days in 2003. There are 7893 households without migrant workers in the 2003 sample, and 5894 with.

The NFPS tracks households over time. For most of our analysis, we do not restrict the sample to be balanced. Table 20 in Appendix 9.2 shows the attrition rate year on year. Overall the annual attrition rate is 4.5%. Some households left the survey and then came back in later years. Among all households observed in 2003, 74.7% were also observed in 2015. Moreover, the probability of

remaining in the sample did not differ much by initial migration status, as shown in Figure 4.[5] The cumulative attrition rate of 25.3% over 12 years, or 2.4% per year, is comparable to the annual attrition rate of 2.5-3.0% in the PSID for the US (Fitzgerald *et al.*, 1998) and lower than Chinese Family Panel Studies, another commonly used household panel survey data in China (Gai *et al.*, 2024). Overall, the attrition rate is relatively low, and its potential threat to our estimation will be discussed in detail in Section 5.3.

4 Reduced-form Specification and Identification

This section outlines the econometric method and discusses potential challenges to identification. The main method is an event-study model with group and time fixed effect. The treatment is at the group-time level, in our context, the prefecture-level city and year level. Subscript s denotes city s, and t for year t. As mentioned earlier, we examine the impact of prefecture-level governments' relaxation of migration barriers on the fertility behaviour of rural households residing in villages under their jurisdiction.

The event study model takes the following form:

$$Y_{hst} = \alpha_s + \beta_t + \sum_{k=-5}^{10} \gamma_k 1[t - policy_s = k] + \varepsilon_{ht} \tag{1}$$

where Y_{hst} is a binary variable indicating whether household h in city s in year t had a newborn, $policy_s$ is the year when the reform took place for city s, α_s is city fixed effect, β_t is year fixed effect, and ε_{ist} is the error term. The parameters of interest γ_k capture the effects of relaxing migration barriers on rural households k years after the reform. γ_k for $k < 0$ allows us to check pre-trends. Standard errors are clustered at the city level.

Several challenges exist in ensuring consistency of the method. First, there are likely heterogeneous treatment effects. Recent studies highlight that under the assumption of a constant treatment effect across groups and over time, the two-way fixed effects regression employed here identifies the effect under the standard 'parallel trends' assumption. However, if the assumption does not hold, the two-way fixed effects regression identifies weighted sums of the average treatment effects (ATEs) across groups and periods, with these weights potentially being negative, as shown in de Chaisemartin and D'Haultfoeuille (2020), Sun and Abraham (2020), Athey and Imbens (2022), Callaway and Sant'Anna (2021), Imai and Kim (2018), and Goodman-Bacon (2018).

[5] Among households who had no member working non-locally for 180+ days in 2003, 74.9% remained in the sample in 2015; for households with at least one person working non-locally for 180+ days in 2003, 74.3% remained.

In our context, the treatment effect is likely to be heterogeneous across cities for two reasons. First, each city sets its own detailed rules for local Hukou applications. While the rule typically requires having stable accommodation and a stable job, what constitutes stable accommodation and stable jobs varied across cities. Second, cities differ in the composition of its population in rural areas. The impact of the Hukou reform on households is likely to vary depending on their education and income.[6]

Based on de Chaisemartin and D'Haultfoeuille (2020), we compute a few statistics that highlight the need for taking heterogeneous treatment effects seriously. The Two-Way-Fixed-Effect (TWFE) method gives an estimate of 0.0016, which is a weighted average of 1271 ATTs. Among the 1271 ATTs, 324 have negative weights. The minimum value of the standard deviation of the ATEs compatible with the coefficient from the TWFE regression and ATE across all treated (g, t) cells being equal to zero is 0.0008. And the minimum value of the standard deviation of the ATEs compatible with the coefficient from the TWFE regression and ATE across all treated (g, t) cells having different signs is 0.0015. These two summary statistics are low in this empirical context: the outcome variable has a mean of 0.027 and standard deviation of 0.16. Thus, it is clear that TWFE can give rather misleading results in this context.

A second identification issue arises from the non-random nature of reform timing and its implementation. For example, cities that had a lower urban unemployment rate may choose to adopt the reform earlier and implement less restrictive migration policies. If low unemployment in urban destinations directly causes an upward trend in fertility in rural areas, then our result would be biased. However, if urban unemployment rate is correlated with the level of fertility only but not its growth, this would be absorbed by the city fixed effect α_s in (1).

One way to examine this potential endogeneity concern is to examine pre-trends. If an omitted factor affected both the timing of reform adoption and the subsequent increase in fertility, we would expect to observe differential trends before reform adoption. The second approach is to control for interactions between initial city-level conditions and year fixed effects. These controls absorb potential correlation between baseline city characteristics and prefecture-specific fertility trends.

Furthermore, we examine whether differential trends may also exist across households with different initial characteristics. Specifically, we control for predetermined household characteristics measured before reform exposure interacted with year fixed effects. These household-level characteristics include demographic composition, parental characteristics, migration-related attributes, and the availability of potential caregivers. This specification allows households with

[6]The Hukou reform increased migration into the cities for the young, low- and mid- skilled. The highly educated and higher-income people from rural areas had smaller difficulties to migrate to start with.

different initial conditions to follow flexible differential trends over time.

Third, the fertility behaviour we observe among remaining rural households may be complicated by non-random attrition and migration. Migration can happen to the entire household as well as to a subset of household members.

When the whole household moves and drops out from the rural sample, this attrition could be problematic for us if it's correlated with some unobservable factor like fertility preference or if it's caused by the reform. For example, if child-loving households are less likely to migrate after the reform, then in the unbalanced sample the average fertility would increase after reform just due to this compositional shift towards child-loving ones. This selection bias will be addressed in depth in subsection 5.3.

Meanwhile, when a subset of household members migrate, there is no attrition at the household level, but potentially measurement errors. For example, when a young couple migrates while their elderly parents stay in our rural sample, it is not certain whether the couple's newborn will be observed in the survey. In principle, as long as the young couple remain economically linked to the left-behind elderly, the elderly should report the young couple as still family members but living at home for very few days.[7] Due to the lack of affordable childcare in urban destinations, it's highly likely that the baby would have to stay behind with grandparents in the rural village and therefore most of newborns should be observed. If the reform induces the younger generation to migrate and all babies are left behind, our estimate will not be biased. If the reform caused some babies to migrate with their parents and they were not reported by the left-behind elderly, then our outcome measure would understate fertility after the reform, and therefore our estimate of the policy impact would be biased downwards.

Finally, it is possible that some other unobserved city-specific shocks are correlated with both the Hukou reform and fertility trends. As a falsification test, we will examine subgroups who are less affected by the Hukou reform directly, and check if the estimated impact on them is small in subsection 5.2.

5 Reduced-form Results

5.1 Impact on fertility

Figure 5 shows our main result: the dynamic effect of the Hukou reform on rural household's fertility. The horizontal axis represents the number of years relative to the de facto removal of migration barriers, with 1 being the first year of reform. The blue points represent the estimated

[7] In the sample, the number of days lived at home has a mean of 222 and standard deviation of 147 for men under age 40.

treatment effects in the treatment periods and placebo effects in the pretreatment periods, while the red intervals indicate the 95% confidence intervals of these estimates and placebos. The estimated effect on fertility is up to 0.02 per year and it is broadly increasing over time.

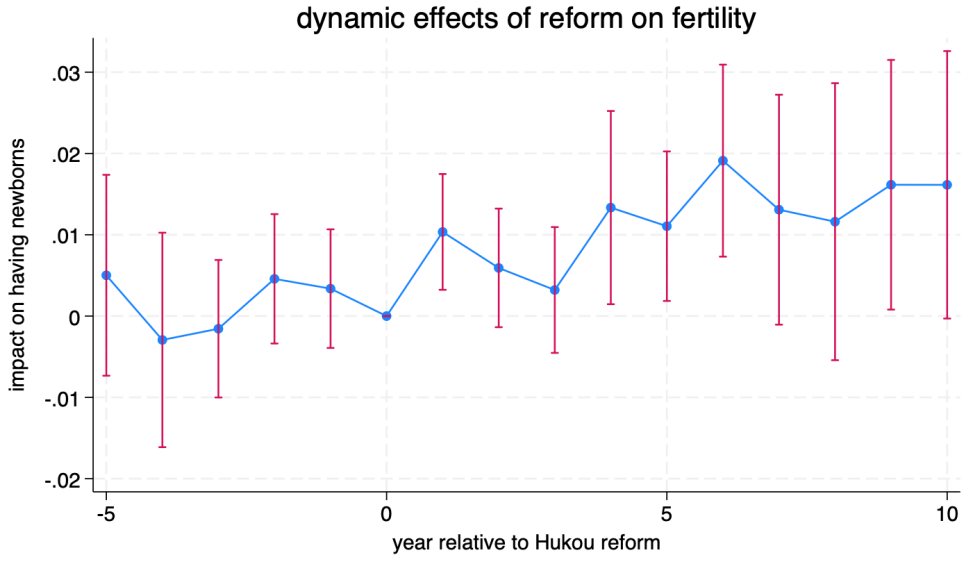


Figure 5: Effect of removing migration barriers on the probability of having a newborn

Notes: The points represent the estimated treatment effects in the treatment periods and placebo effects in the pretreatment periods, and the error bars represent 95% confidence intervals of the estimated effects and placebos. To draw these figures, we followed the estimation strategy proposed by de Chaisemartin and D'Haultfoeuille (2020).

The Average Treatment Effect on the Treated is 0.0110, with a standard error of 0.0040, as listed in Table 3. As the sample mean is 0.0274, the ATT is more than one third of it.

As our main sample contains a lot of households for whom fertility is extremely likely to be zero, our main estimate of 0.0110 per household per year has been diluted. Typically, having a baby requires a woman of child-bearing age. While it's theoretically possible for a household to adopt, or for the new mother to be absent from the survey (e.g. due to divorce), or an older woman to have babies, it's rare in practice. 96% of newborns in the sample were born to households where an adult woman under 40 was present.

As shown in Table 3, about 55% of the sample has a woman aged 18-40. The ATT conditional on having a woman aged 18-40 is 0.0216, almost twice as large as the ATT for the whole sample. Among the group without women aged 18-40, both the mean of newborn and the ATT are close to zero (Table 23). If we draw the line at age 45 instead of 40, the ATT among households with young women is 0.0179 and still highly significant. This is nearly half of the mean of outcome in

this subsample.

Table 3: Average treatment effect by presence of woman of fertile age

Sub-sample	sample size	mean of outcome	estimated ATT	standard error
all households	202806	0.027	0.011***	0.004
households with woman aged 18-40	112949	0.048	0.022***	0.006
households with woman aged 18-45	129369	0.042	0.018***	0.006
households without woman aged 18-40	89857	0.002	-0.002	0.002
households without woman aged 18-45	73437	0.002	-0.001	0.002

Note: each row uses the same event study specification as in Figure 5 and Table 4. Each row selects a subsample as labelled.
*** p<0.01, ** p<0.05, * p<0.10.

Note the presence of a woman of fertile age might itself be affected by the reform in ambiguous ways. On the one hand, younger women may be encouraged by the reform to move away. On the other, higher incomes may make it easier for local young men to attract potential wives. Figure 15 in the Appendix shows the presence of women had a weakly declining trend in the treated regions both before and after the reform. Thus, we cannot eliminate the concern that the presence of women of child-bearing age is endogenous. Therefore, the following analysis will not restrict sample by the contemporaneous presence of women under any age threshold.

Our finding of 0.0179 extra newborn among households with women aged 18-45 is larger than the "0.9 percent increase" in fertility found in Dong *et al.* (2023). Note that Dong *et al.* (2023) used cross-sectional surveys of *migrant* women aged 15-45 who had been at their urban destinations for at least one year, whereas our sample is a panel of households observed in rural areas.[8] Looking at the pre-existing long-term migrants alone misses the impact on the large numbers of households remaining in rural areas, with or without temporary migrating members. The latter rural population was much larger than the long-term migration population.[9] Viewing our results together with the estimate from Dong *et al.* (2023), we can now confidently say that the Hukou reform significantly increased fertility for the whole original rural population.

A critical assumption underlying the event study approach is the existence of common parallel trends after treatment. While this assumption cannot be tested, we can examine parallel

[8]There are also differences in the timing of policy, sample period, and nature of data. Dong *et al.* (2023)'s main sample is over 2011-17 and the policy they investigate was the relaxation of Hukou system since 2014. We collected city-level data on Hukou policy from the early 2000s to 2015, and our sample also covers a longer period. Moreover, the longitudinal nature of our data, covering 2003–2015, has allowed us to test for selective migration (in subsection 5.3) and therefore be more confident of our estimates. Our sample does not include reforms after 2015, as these fall outside our data coverage.

[9]According to China's Population Census in 2010, long-term migrants, defined as those who lived away from their registered places for more than one year, accounted for about 15.7% of total rural-registered population.

trends before treatment as an indicative test. If pre-trends are significantly different, we would be reluctant to accept the assumption of parallel trends after treatment.

Figure 5 shows that the pre-treatment effects are close to zero and not statistically significant at the conventional level.[10] Recall that a city's timing of reform is not random and correlated with initial conditions. The absence of pre-trends supports the view that city's initial conditions are not strongly correlated with the trend of rural fertility.

The rest of this section conducts a range of checks to establish the credibility of our main finding. Section 6 will develop a model to explain the mechanisms of this positive fertility response. Section 7 will examine empirical evidence regarding those mechanisms.

5.2 Robustness checks

First, we apply different estimators from the recent literature on staggered diff-in-diff and heterogeneous treatment effects. Figure 6 shows that both methods from Callaway and Sant'Anna (2021) and Borusyak *et al.* (2024) give similar results to de Chaisemartin and D'Haultfoeuille (2020), in that the reform increases the probability of having newborns in the next 10 years by about 1%. By contrast, the standard TWFE model gives estimates that are much closer to zero and insignificant. This contrast is striking. Recall that TWFE will be misleading when there are heterogeneous treatment effects and negative weights, and that about one quarter of groups have negative weights in our setting (shown in section 4). These results highlight the importance of correcting for negative weights in event study analysis with heterogeneous treatment effects.

Second, Table 4 looks at the number of newborns, rather than the binary outcome used in the main analysis. This makes very little difference – the ATT is still 0.011. This is because in the vast majority of households with newborns, there was only one newborn in that year. The last two rows in Table 4 distinguish newborns by whether they are the first in their generation. This is a proxy to distinguish first-borns and higher birth orders, because the survey only identifies relationship between each individual and the head. The reform increased the birth of first-borns as well as babies of high birth orders. Because childlessness was and still is rarely an active choice of couples in China, the positive impact on first-borns suggests the reform may have increased the marriage rate or accelerated the arrival of the first child.

Third, to address the concern that migration reforms were not randomly assigned across cities, we allow for differential trends across cities by interacting pre-reform city characteristics with year fixed effects. Specifically, we construct a comprehensive set of baseline characteristics measured in 2001, capturing key dimensions emphasized in the migration and local policy

[10] Test of joint nullity of pre-trends has $p - value = 0.249$.

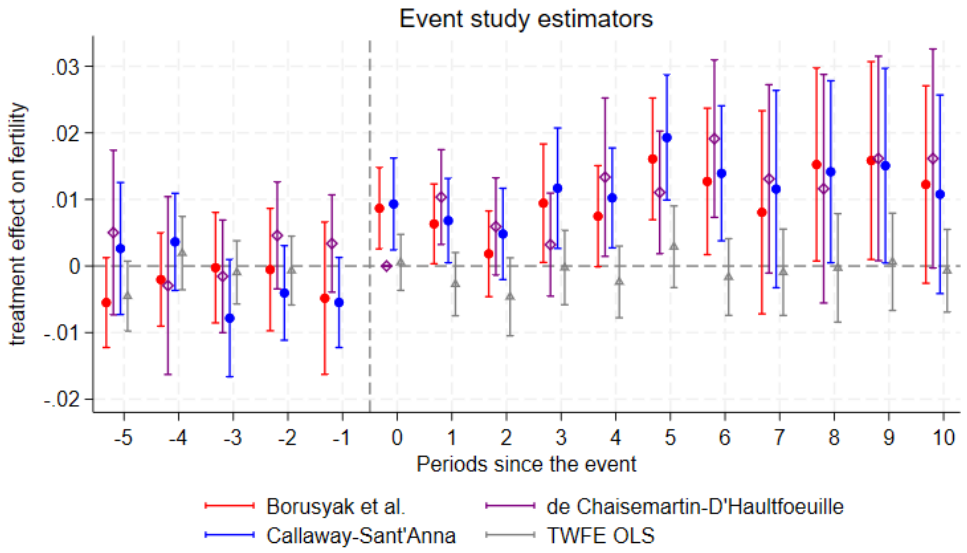


Figure 6: Effect of reform on the probability of having a newborn across different event study estimators

adoption literature (Casarico *et al.*, 2018; Jin, 2024; Tian, 2024; Chen *et al.*, 2025). These characteristics include the urban unemployment rate, the share of fiscal expenditure on education, fertility rates, the employment share of the state sector, migration rates, and both rural and urban age structure (e.g., the share of elderly population aged 65+ and the share of younger cohorts). In our preferred specification, we include the full set of these pre-reform characteristics interacted with year fixed effects, thereby allowing cities with different initial conditions to follow flexible differential trends over time.

To further examine whether our results are sensitive to the selection of baseline city controls, we consider alternative approaches of selecting these characteristics. First, following Table 27, we estimate specifications controlling for baseline characteristics that exhibit relatively strong pairwise correlations with reform timing.[11] Second, we examine the relationship between reform timing and initial city conditions using a multivariate regression and include the baseline characteristics that are statistically significant predictors of reform timing as additional controls, as shown in Table 28.[12]

The results reported in Table 5 show that the estimated effects remain stable after controlling

[11]The subset used in robustness checks consists of variables that exhibit relatively stronger correlations with reform timing (e.g., correlation coefficients above 0.1). Results are robust to alternative selections of baseline controls.

[12]The second set of controls consists of seven baseline city characteristics identified as statistically significant predictors of reform timing in the multivariate regression. These variables include urban unemployment rate, the share of fiscal expenditure on education, fertility rate, the ratio of fiscal expenditures to fiscal revenue, rural employment share in the secondary sector, urban elderly population share, and urban population share aged below 30.

Table 4: Average treatment effect on fertility

Outcome	sample size	mean of outcome	estimated ATT	standard error
has newborn	202806	0.027	0.011***	0.004
number of newborns	202806	0.028	0.011***	0.004
has a newborn that's 1st in that generat	202806	0.016	0.007**	0.003
has a newborn that's 2nd or higher birth	202806	0.011	0.004**	0.002

Note: All rows in this table use the same sample. 'ATT' is the Average Treatment Effect averaged across all Treated (g,t) cells. ATT and standard errors are produced by the did_multiplegt package in STATA. Sample size is number of observations at household-year level. 'Mean of outcome' is the mean of the outcome variable in the sample.
*** p<0.01, ** p<0.05, * p<0.10.

for alternative sets of initial city characteristics. Including the full set of baseline city controls yields an ATT of 0.0150, which is very close to the benchmark estimate of 0.0110. Approximately 7% of observations are lost due to missing values in baseline city characteristics. The similarity of estimates across different specifications suggests that our findings are not driven by specific choices of city-level controls.

We next examine whether other contemporaneous policy changes could confound the estimated effect of migration reform on fertility. Specifically, we control for three major policy shocks that may affect fertility decisions: city-level reductions in trade policy uncertainty following China's accession to WTO (Facchini *et al.*, 2019), city-level minimum wages in urban areas (Fang and Lin, 2015), and city-level abolition of agricultural taxes (Chen and Wang, 2014). The results are reported in Table 6. Although information on these additional policy shocks is unavailable after 2009, leading to a substantial reduction in sample size, the estimated effects remain consistent. With all controls included, the ATT is 0.0233, compared with 0.0120 in the same sample without additional controls. [13] Given the substantial reduction in sample size and the limited additional explanatory power of these controls, our preferred specification adopts a more parsimonious approach without these contemporaneous policy controls. [14]

A second concern is that the estimated effect may be driven by differential trends across households with different initial characteristics. Since households may differ in demographic composition and migration-related characteristics before migration reform exposure, we control

[13] We find similar patterns when including missing-value indicators for city-level controls with incomplete information, which preserves a consistent sample size across specifications. In this case, the ATT with all controls included remains close to the benchmark estimate.

[14] Using an event-study specification, we further conduct a joint test of the city-level control variables. While the null hypothesis of joint insignificance is marginally rejected at conventional levels (F(21,185)=1.65, p=0.0426), the inclusion of these controls leaves the estimated treatment effects largely unchanged.

for a set of predetermined household characteristics measured in 2003 interacted with year fixed effects. Specifically, we sequentially add characteristics of potential mothers, potential fathers, and the availability of potential caregivers, including parental age, education, migration status, the presence of young adult women, and the presence of grandparents. These controls allow households with different initial characteristics to follow flexible differential trends over time.

As reported in Table 7, the estimated effects remain stable after incorporating these household-level controls. The ATT increases slightly from 0.011 in the benchmark specification to 0.015 in the most comprehensive specification, while remaining statistically significant. These results suggest that the estimated impact of migration reform on fertility is unlikely to be driven by differential trends associated with initial household characteristics.

Table 5: Effect of removing migration barriers on the probability of having a newborn, with and without cities' initial controls

Outcome	sample size	mean of outcome	estimated ATT	standard error
No_controls	187880	0.027	0.010**	0.004
All_controls18	187880	0.027	0.015***	0.005
Selected_controls8	187880	0.027	0.013***	0.004
Regression_controls7	187880	0.027	0.016***	0.005

Note: All rows in this table use the same sample. 'ATT' is the Average Treatment Effect averaged across all Treated (g, t) cells. ATT and standard errors are produced by the did_multiplegt package in STATA. Sample size is number of observations at household-year level. 'Mean of outcome' is the mean of the outcome variable in the sample. The full controls include interactions between year fixed effects and a comprehensive set of 18 pre-reform city characteristics. The 8 selected pre-reform city characteristics include urban unemployment rate, the share of fiscal expenditure on education, prefecture fertility rate, and the proportion of state-owned firms in employment, the rural out-migration rate, the rural out-migration rate among individuals aged 20-34, the rural elderly population share (65+), and rural population share aged below 30. The 7 pre-reform city characteristics that are statistically significant in a multivariate regression examining the relationship between reform timing and initial city conditions include urban unemployment rate, the share of fiscal expenditure on education, fertility rate, the ratio of fiscal expenditures to fiscal revenue, rural employment share in the secondary sector, urban elderly population share (age 65+), and urban population share aged below 30.
*** p<0.01, ** p<0.05, * p<0.10.

Fourth, to examine whether reform timing is correlated with pre-existing fertility trends, we conduct a placebo test using data from the 1990 and 2000 Population Censuses. Specifically, we regress changes in rural total fertility rates between 1990 and 2000 on subsequent reform timing. Since all Hukou reforms in our sample took place after 2002 (between 2002 and 2015), this specification relies entirely on pre-reform variation. We find no significant relationship, suggesting that cities that adopted Hukou reform earlier were not systematically different in terms of pre-sample fertility trends (Table 8). This alleviates concerns that our estimates are

Table 6: Effect of removing migration barriers on the probability of having a newborn, with and without cities' initial controls and contemporaneous policy shocks

Outcome	sample size	mean of outcome	estimated ATT	standard error
No controls	99533	0.029	0.011**	0.005
Full controls	99533	0.029	0.015**	0.007
Full controls + 3 policy shocks	99533	0.029	0.023***	0.006

Note: All rows in this table use the same sample. 'ATT' is the Average Treatment Effect averaged across all Treated (g, t) cells. ATT and standard errors are produced by the did_multiplegt package in STATA. Sample size is number of observations at household-year level. 'Mean of outcome' is the mean of the outcome variable in the sample. The controls are interactions of year and 18 pre-reform city characteristics, and three contemporaneous policy shocks, including reductions in trade policy uncertainty after China's accession to WTO, minimum wages in urban areas, city-level abolition of agricultural taxes.
*** p<0.01, ** p<0.05, * p<0.10.

Table 7: Robustness to Household-Level Differential Trends

Specification	Sample size	Mean of outcome	Estimated ATT	Standard error
Benchmark	154909	0.028	0.011***	0.004
Mother characteristics	154909	0.028	0.014**	0.006
Mother and father characteristics	154909	0.028	0.015**	0.007
Mother, father, and caregiver characteristics	154909	0.028	0.015**	0.007

Notes: All rows in this table use the same sample. 'ATT' is the Average Treatment Effect averaged across all Treated (g, t) cells. ATT and standard errors are produced by the *did_multiplegt* package in STATA. Sample size is the number of observations at the household-year level. 'Mean of outcome' is the mean of the outcome variable in the sample. All household characteristics are measured in 2003 before migration reform exposure and are interacted with year fixed effects. Column (1) reports the benchmark specification. Column (2) additionally controls for baseline characteristics of potential mothers, including female presence, age, education, and migration status. Column (3) further adds baseline characteristics of potential fathers. Column (4) additionally controls for family care availability measured by the presence of grandparents. *Significance levels*: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

driven by differential underlying demographic dynamics.

Table 8: Placebo Test: Pre-reform Fertility Changes and Reform Timing

	(1) Rural TFR	(2) Rural childlessness
Year of Hukou reform	-0.005	0.000
	(0.007)	(0.000)
Observations	256	256
R-squared	0.002	0.000

Notes: This table reports placebo regressions examining whether the timing of Hukou reform is correlated with pre-existing fertility trends. The dependent variables are changes in rural total fertility rates and rural childlessness rates between 1990 and 2000. The key explanatory variable is the prefecture-level year of Hukou reform adoption (between 2002 and 2015). Robust standard errors are reported in parentheses.

Fifth, to explicitly address the concern that individuals may be exposed to migration opportunities arising from reforms in other cities, we construct a distance-weighted measure of external reform exposure, excluding the individual's own city. This variable captures the extent to which migration barriers are relaxed in other locations, with greater weight assigned to geographically closer cities. After controlling for this external exposure, the estimated effect of local Hukou reform remains positive and quantitatively stable (Table 9).

Sixth, to address the potential confounding effect of the relaxation of the one-child policy (OCP) around 2013–2014, we estimate our main treatment effects by restricting the sample to years up to 2012. We find that the estimated effects are significant and positive for early adopters (Table 9).

Table 9: Dynamic effects of Hukou reform on fertility: robustness checks

Outcome	sample size	mean of outcome	estimated ATT	standard error
Control for reform exposure in other cities	187880	0.027	0.009*	0.005
Restrict sample to year up to 2012	144467	0.029	0.011**	0.005

Note: All rows in this table use the same sample. 'ATT' is the Average Treatment Effect averaged across all Treated (g, t) cells. ATT and standard errors are produced by the did_multiplegt package in STATA. Sample size is number of observations at household-year level. 'Mean of outcome' is the mean of the outcome variable in the sample. The first specification controls for distance-weighted reform exposure in other cities (excluding the city itself). The second specification restricts the sample to years up to 2012 to account for potential confounding from subsequent relaxations of the one-child policy. *** p<0.01, ** p<0.05, * p<0.10.

Finally, we will examine subgroups who are less affected by the Hukou reform directly as a falsification test. In particular, we will look at high-skilled individuals and those with non-

agricultural Hukou households. Both of them faced lower barrier to migrate before the reform and were therefore less affected by it. Therefore, we expect the same event study to find no significant effects on them. If the estimates on these groups were significantly positive, that would indicate our main results were biased positively due to some omitted variable. Tables 25 and 26 in Appendix show that the reform had no significant impact on these groups.

5.3 Attrition from sample

Attrition in our sample is nontrivial but moderate. On a year-to-year basis, 93.5% of households observed in year t continue to be observed in $t+1$.

Attrition is often temporary and households return to the sample in later years. There are 13,787 households in 2003, but 20,513 households appeared at least once over 2003-2015. Among all households observed in 2003, 55% were observed every year from 2003 to 2015, 18% were observed in 12 out of 13 years. At the household-year level over 2003-2015, 86% of observations participated in the survey (we observe them) and 14% are missing (i.e. unobserved).

Our main analysis (Figure 5) includes all households, regardless of how many years they are observed. As a result, the composition of the sample varies over time. This opens the door to potential bias if attrition is systematically related to treatment exposure or fertility behavior.

First, attrition biases the estimated treatment effect if households who drop out of the sample differ systematically in fertility preferences. If households with low fertility preferences (who are more career oriented) are more likely to leave the sample due to the Hukou reform, then our observed sample shifts to be more child-loving and so our estimate would be biased upwards. The opposite could be true if child-loving households are most likely to move due to the improved education provision in urban areas. Fortunately, Table 21 in Appendix 9.3 suggests household attrition is not correlated with number of young children.[15]

Second, the reform could have differential treatment effects across households and their treatment effect may be correlated with their probability of being observed. For example, if always migrants and marginal migrants — who are more likely to benefit from reform — are both more likely to leave the sample and have more positive fertility responses than always stayers, then our sample would put too much weight on stayers and our ATT estimate will understate the true ATT.

To address the bias from selective attrition, we have two approaches. First, we re-estimate the ATT using a balanced sample of households observed in every year from 2003 to 2015. This balanced panel contains nearly 100,000 household-year observations, or just under half of the

[15] Appendix 9.3 shows attrition (or being unobserved) is correlated with some other characteristics with statistical significance, but the difference is small.

original sample. Figure 7 shows that the estimated ATT in the balanced sample is 1.96 ppts, even larger than the 1.10 ppts in the main analysis. While this balanced sample is not representative of the original sample, it does show that the qualitative finding of a strong positive fertility impact is not driven by selective attrition.

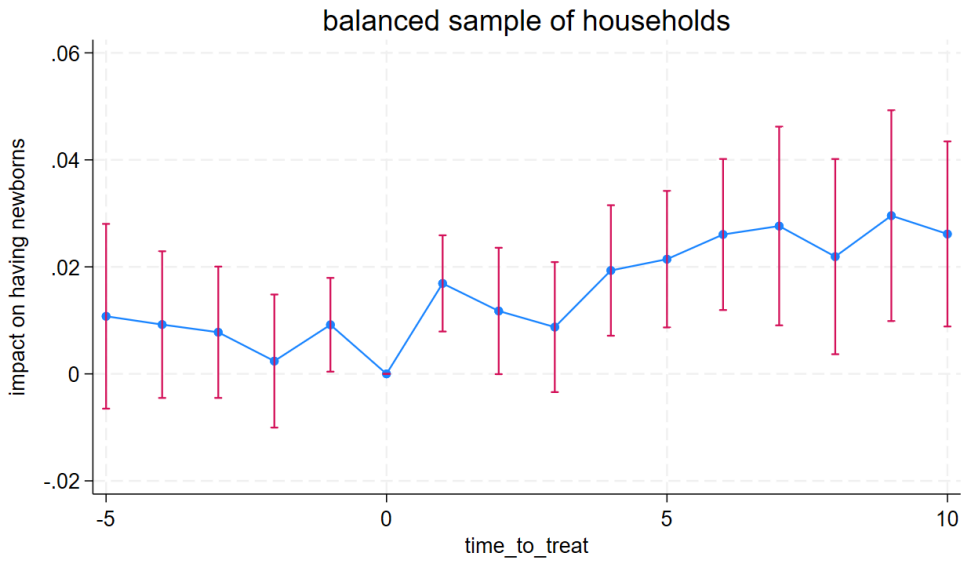


Figure 7: Effect of reform on fertility in the balanced sample of households

Note: Sample restricted to households observed in every year over 2003–2015. Estimation uses the method from de Chaisemartin and D'Haultfoeuille (2020).

Second, we apply a simplified version of the framework from Rathnayake *et al.* (2024) to put bounds on the bias. We assume that the reform has a weakly negative impact on the probability of being observed. This implies that the estimator identifies the treatment effect for always-observed households, plus a bias term.

The bias term is the product of two factors: 1. The share of control-group households whose observability would be affected by the reform, and 2. The difference in outcome trends between always-observed and selectively-missing households in the control group. The derivation and explanation are in Appendix 9.3. Our estimation on how the reform affects the probability of being observed suggests the first factor is likely smaller than 0.1 in absolute magnitude. Moreover, it cannot exceed 0.25, because three quarters of households observed in 2003 remained in 2015.

For the second factor, we look at historical trends and reason that 0.2ppts would be a plausible magnitude and 1ppt would be an extreme case. Taking these two factors together, our central estimate of the bias in our ATT is around –0.02 percentage points, which is very small relative to our main estimate of 1.1 ppts. Even under extreme assumptions about selection and trend

differences, the estimated ATT remains within a plausible range of [0.85, 1.35]. Therefore, our main finding of strong positive fertility response is not due to selection bias.

6 The model

Having established the empirical result that the reform increased fertility, we now develop a theoretical model to illustrate the mechanisms behind this response. While the model will not be structurally estimated, its predictions will be tested empirically. Moreover, the model represents a novel contribution by integrating fertility decision and partial migration choices into a single household framework — a gap in existing theoretical literature. To the best of our knowledge, no model has treated fertility and rural-to-urban migration as decisions in the context of a developing country. The closest is Tumen (2012), which incorporates endogenous fertility into a model of residential sorting, illustrating how fertility and neighborhood composition co-evolve. However, this analysis focuses on intra-urban sorting in settings with frictionless mobility and without migration barriers. In many developing-country contexts, by contrast, institutional barriers to migration are substantial and evolving. Moreover, we explicitly model the decision of family separation together with migration, as extended families are common and an important factor in partial migration in China. This makes our model more applicable to other developing countries.

Consider a unitary household with a father and a mother who make decisions. The couple choose number of children n, consumption c, mother's labor supply h_w, educational expenditure E, and migration configuration m. Grandparents may be present which could affect the parents' decisions, but grandparents' presence is assumed exogenous.

We assume there are 4 migration configurations to choose from: $m = 0$: No one migrates; $m = 1$: Father-only migrates; $m = 2$: Both parents migrate, any child remains rural; $m = 3$: Both parents and children migrate. These four configurations cover about 85% households in the applicable sample.[16] Each migration option has its trade-offs, as listed in table 10.

[16] In our data, among rural households containing two opposite-sex adults (with or without older adults or children), the proportions of different migration configurations were as follows: "no one migrates" (52.5%), "father only migrates" (12%), "both parents migrate while at least one child remains rural" (15.6%), and "both parents and children migrate" (5.2%). Together, these four categories account for about 85% of such households. The non-conforming ones are mostly due to "children aged 15+ migrate" (6.2%), "mother migrates while father stays rural" (4.1%), and "both parents stay while child migrate" (3.5%).

Table 10: Migration Configurations

m	Migration Type	Benefits and Costs
0	All stay rural	Low cost of children due to grandparental care; poor school quality; low rural wages and transfers.
1	Father-only migrates	Higher household income; spousal separation cost; grandparental care preserved; child remains in rural schools.
2	Both parents migrate, child stays	Even higher income; parent-child separation costs; grandparents' availability is required; child remains in rural schools.
3	Parents and children migrate	High income and access to social benefits, urban school (δ_s^U); higher cost of raising children p^n; grandparents' help less available

In theory, the model is totally static and all choices are made simultaneously. For ease of derivation, we will consider the migration decision as the first stage, and in the second stage all other choices are made conditional on migration decision. Conditional on the migration decision m, the household solves the following optimization problem, where subscript m denotes the migration choice.

$$\max_{c,n,E} U(c,n,q) = (1-\gamma)\log c + \gamma(\log n + \beta \log q) - C^m \tag{2}$$

$$s.t.$$

$$\log q = \delta_s^m + \delta_e^m \log \frac{E}{n} \tag{3}$$

$$\tau_w + h_w = 1 \tag{4}$$

$$\tau_w = n\tau(1 - G^m) \tag{5}$$

$$c + E + p^m n = w_m^m + w_w^m h_w + y_0^m \tag{6}$$

$$\tau_w, h_w \geq 0 \tag{7}$$

The utility function is log-linear in its three components, following Galor (2012). Preference parameters γ and β are assumed fixed and independent of migration choice. Each type of migration incurs a disutility $C^m > 0, \forall m > 0$.

Average quality of children q depends on the social environment and educational investments per child, as in equation (3). E is total parental expenditure on education, δ_s is the social environment, which can be affected by policies such as quality of state-funded education and health services for children, and δ_e is the returns to parents' educational expenditure. The model assumes away the possibility of parental time investments to improve child quality.[17]

Each potential parent has one unit of time endowment. The woman's time can be spent on either childcare τ_w or labour supply h_w. The man is assumed to spend all his time on work and zero on childcare. Raising each child takes a fixed amount of time τ. Childcare can be done by either the mother or grandparents, as in equation (5). The assumption that men do not provide childcare is not far from the reality in rural China: our calculation based on the 2010 Survey on the Social Status of Women in China shows that among female respondents born with a rural Hukou, only 1.5% reported that their husbands were the primary provider of daytime care for their children before age three, compared to 70.7% who did so themselves and 25.3% who relied on the grandparents. [18]

Note that we assume grandparents' presence G affects how much childcare time is required of the mother. If a grandparent is fully present $(G = 1)$, parents do not need to spend any time on childcare. Grandparents availability depends on the migration choice: G^m has subscript m. For parents to both migrate and leave children behind $m = 2$, grandparents (or other relatives) must be available to take care of the children $G^2 = 1$. If children live with parents in urban areas $(m = 3)$, which is necessary for utilising better urban schools, then G will likely be lower, because it would be costly to bring grandparents to urban areas. If grandparents are not available $(G = 0)$ or only partially available $(0 < G < 1)$ due to poor health or physical distance, then the parents will have to spend some time on childcare. This model design echoes empirical finding in Fan *et al.* (2011) that the main factors determining the decision of leaving children behind are child's age and whether grandparents are available to help. Conditional on m, grandparents' provision of childcare is not a choice in the model, but a given. In our sample the older generation are much lower-educated than the younger generation, so it's often efficient for the grandparents to provide childcare instead of the mother/father.

Finally, equation (6) is the budget constraint. p^m is the cost of raising children (e.g. housing and food) and is independent of child's quality. y_0^m is unearned income. All environmental parameters—including the man's wage w_m^m, woman's wage w_w^m, costs of having children p^m, and

[17] This assumption is also made in the model about quantity-quality tradeoff in Doepke *et al.* (2023). Parental time inputs may play a small role in child quality in our context because children's education tend to far exceed their rural parents.

[18] Interestingly, among male respondents with a rural Hukou at birth, 10% reported that they themselves were the primary daytime childcare provider before their child(ren) reached age of three, while 59% indicated it was their spouse, and 28.6% said it was the grandparents.

unearned income y_0^m —are location-specific and hence differentiated by subscript m.

We assume the economic conditions vary by m in the following ways. Wages are higher in urban areas. The respectively adult needs to migrate if they want to obtain the urban wage: $w_m^0 < m_m^1 = w_m^2 = w_m^3$. $w_w^0 = w_w^1 < w_m^2 = w_m^3$. Grandparents availability varies across households for exogenous reasons, denoted as G_h. If a household has $G_h < 1$, then the mother cannot migrate without the child ($m = 2$ is not an option). For a given household with available grandparents, taking children to urban areas reduces grandparents availability: $G^3 < G_h = G^2 = G^1 = G^0$. Education quality is higher in urban areas and can only be enjoyed if children migrate: $\delta_s^3 > \delta_s^0 = \delta_s^1 = \delta_s^2$. Returns to educational investment is higher in urban areas because there are more courses available: $\delta_e^3 > \delta_e^0 = \delta_e^1 = \delta_e^2$. Direct cost per child is higher in urban areas (food and housing): $p_n^3 > p_n^m, \forall m < 3$. We can also consider compulsory tuition fees as part of p_n^m, because every child is obliged to attend school and parents are obliged to pay fees per child. This fee was typically higher in urban areas than in rural areas, hence adding to the difference $p_n^3 > p_n^m, \forall m < 3$. The fixed cost of migration is positive $C^3, C^2, C^1 > C^0 = 0$. We are agnostic about which form of migration is most costly, because partial migration involves the psychological cost of separation while less logistic cost than whole-family migration.

Conditional on m, the optimal choice of n, E, c are:

$$n^m = \frac{\bar{y}^m(\gamma - \gamma\beta\delta_e^m)}{p^{*m}} \tag{8}$$

$$E^m = \gamma\beta\delta_e^m \bar{y}^m \tag{9}$$

$$c^m = (1-\gamma)\bar{y}^m \tag{10}$$

where $p^{*m} = w_w\tau(1-G) + p^m$, and $\bar{y}^m = w_m^m + w_w^m + y_0^m$. p^{*m} is the shadow cost of children, which is affected by mother's wage if childcare is not entirely done by grandparents. $\bar{y}^m$ is the full income conditional on migration choice. Equations (8)-(10) show that full income are allocated by fixed proportions across three areas : quantity of children, education expenditure, and consumption. Their derivations are in appendix 9.4.

The different economic conditions by m lead to different levels of fertility, according to equation (8). First, we have $p^{*0} = p^{*1} = p^{*2} << p^{*3}$. p^{*3} is particularly high for three reasons: children need to be housed and urban housing cost is much higher (high p_n^3); taking children to urban areas reduce access to grandparents' help (G^3 is lower); and the wife's higher urban wage adds to higher p^{*3}. Interestingly, $w_w^2 > w_w^1$ does not translate into $p^{*2} > p^{*1}$ for any given household, because the choice $m = 2$ requires grandparents to be available $G = 1$, and in such households the wife's wage does not enter childcare cost. Between $m = 1$ and $m = 0$, there is no difference in the shadow cost of children because the migrant father plays no role in childcare.

Second, we have $\bar{y}^0 < \bar{y}^1 < \bar{y}^2 \leq \bar{y}^3$. The first two inequalities are due to higher wages in urban areas. The last inequality would be true if the urban destination gives more welfare transfers for families with children.

Given equation (8), we can compare fertility across migration choice m:

- $n^0 < n^1$, male migration will increase family income and will not increase the cost of childcare or direct cost of feeding/housing children.
- $n^1 < n^2$. Female migration has a pure income effect. But female migration is only feasible if $G = 1$
- $n^2 > n^3$ is more likely than not. This is because the proportional difference in p^* is likely to exceed the proportional difference in full income. It's ambiguous whether n^3 is higher than n^0 (or n^1), as there are substantial difference in both p^{*m} and $\bar{y}^m$.

These differences across m are conditional on all other parameters being the same.[19] These differences will be useful for thinking about how fertility would change if a household changes their m choice, perhaps in response to the Hukou reform.

6.1 Effects of Hukou Reform

The reform can affect several exogenous variables, and we will discuss how each of them affects fertility and migration. We will first look at fertility response among households whose migration choice m did not respond to the reform in Table 11, and then we will look at switchers.

For 'always migrants', Table 11 shows how their fertility conditional on migration choice (n^m) is affected by each change in exogenous variables. First, migrant households' full income will increase ($\uparrow \bar{y}^m, m > 0$). This is due to improved access to access the social services (e.g. pensions, unemployment insurance, medical insurance). Some of the benefits were related to employment status, while others just require obtaining the local Hukou. This means the reform improved both the expected labour income (inc. pensions) and unearned income for migrants. Hence, the reform will have a positive impact on both fertility and consumption, according to equations (8) and (10).

Second, a higher wage for the mother may have a negative substitution effect on fertility, due to the increased cost of childcare time. This channel is only present in the case of $m = 3$ and only if the grandparent does not provide all the childcare $G < 1$.[20] Men's wage has no substitution

[19] In the cross section, households that choose certain migration configurations may have different preferences regarding the quantity and quality of children, as well as different wages and grandparents availability.

[20] If $m < 2$, the mother stays rural and her wage is unaffected by the reform. If $m = 2$, the mother migrates and the child is left behind with grandparents, then the mother's wage does not affect childcare cost.

effect because they do not provide childcare.

Third, the monetary cost of raising children in the urban destination p^3 will fall because the reform improved children's access to public services (e.g. education, health). The mechanism only applies if $m = 3$. If this mechanism is the primary driver of the overall positive fertility effect, we would expect the fertility increase to be concentrated among households who would always choose $m = 3$ (whole-family migration).

In summary, the model predicts positive effects on fertility and consumption among most households who were going to migrate regardless of the reform ($m = 1, 2, 3$). The only potential exception is the small subgroup where $m = 3$ and $G < 1$. Even for this group, the substitution effect is likely to be dominated by the income effect, because the reform's impact on female wages is likely smaller than its impact on full income. A secondary prediction of the model is that grandparents availability $G = 1$ is associated with a more positive fertility response to the reform, because it helps keep the childcare time cost low.

Meanwhile, migration choice m is likely to respond to the reform. Table 12 lists how the reform moves all exogenous variables, and their impact on the relative payoffs of various migration options. Increases in unearned income and male wage in urban areas will increase the payoffs of all three types of migration $m = 1, 2, 3$. Higher female wage will increase the attraction of $m = 2, 3$, compared to $m = 0, 1$. The Hukou reform improved access to state schools and health services in urban areas, which were better resourced than in rural areas. In other words, δ_s is likely to increase significantly for rural households who migrate and obtain urban Hukou. As migrant parents are no longer forced to pay tuition fees, that is also a reduction in p_n^3. Both $\uparrow \delta_s$ and $\downarrow p_n^3$ clearly increase the attraction of taking children to urban destinations. In short, the reform increases the payoffs of all migration options relative to staying rural $m = 0$, and especially for taking children along with migrant parents $m = 3$.

Recall that the model predicted $n^0 < n^1 < n^2$, this means shifts from $m = 0$ to $m = 1$ and to $m = 2$ will come with higher levels of fertility. The model cannot ascertain the level of n^3 relative to n^0, n^1, so it's ambiguous whether switchers who switch to $m = 3$ will have higher or lower fertility. Because $m = 2$ is associated with highest fertility and because $m = 2$ requires grandparents availability, the model predicts that households with grandparents availability are more likely to increase fertility in response to the reform.

Finally, the reform does not directly affect 'rural stayers': those who would choose $m = 0$ regardless of the reform. But their expectations about their own prospects and their children's prospects are likely to improve due to the reform. This could be considered an increase in full income for the stayers y_0^0, which should increase their fertility and consumption as well. As this has to work through expectations of future, we expect the fertility response to be weaker among

Table 11: Model predictions on the effects of Hukou Reform

If migration choice m does not respond to reform				
exog. variable affected	migration choice m	mechanism	effect on fertility	other effects / notes
$\uparrow y_0^m$	$\forall m \geq 1$	full income $\uparrow$	$n^m \uparrow$	$c^m \uparrow$
$\uparrow w_m^m$	$\forall m \geq 1$	full income $\uparrow$	$n^m \uparrow$	$c^m \uparrow$
$\uparrow w_w^m$	$\forall m \geq 2$	full income $\uparrow$	$n^m \uparrow$	$c^m \uparrow$
$\uparrow w_w^3$	$m = 3$	substitution effect	$n^3 \downarrow$	$h_w \uparrow$ only if $G < 1$
$\downarrow p_n^3$	$m = 3$	substitution effect	$n^3 \uparrow$	none

Table 12: Reform's impact on payoffs from migration

exog. variable affected	Impact on indirect utility from option m		
	$m = 1$ man only migrate	$m = 2$ parents only migrate	$m = 3$ parents&kids
$\uparrow y_0^m, \uparrow w_m^m$	(+)	(+) if $G = 1$	(+)
$\uparrow w_w^m$	o	(+) if $G = 1$	(+)
$\uparrow \delta_s^3, \downarrow p_n^3$	o	o	(+)

'rural stayers' than among 'always migrants'.

In summary, this model explains the various channels of fertility response to the Hukou reform and elaborates the heterogeneity in fertility responses. Specifically, it predicts that

1. consumption should increase among 'always migrants' due to higher full income. It may also increase among 'always stayers' due to improved expectations about the future.

2. the fertility impact is more positive on 'always migrants' than 'rural stayers'; and arguably stronger among those who would bring children to urban areas than other 'always migrants';

3. households where the prospective parent(s) migrate and children stay behind have higher fertility than non-migrating households (unless their preferences/constraints are very different)

4. migration with children should increase after the reform, especially if grandparents are unavailable;

5. the reform would also cause some households to switch from no-migration to parent(s)-only type of migration. And these switchers' fertility should increase at the same time.

6. the fertility response to the reform is more positive when grandparents are available

The next section verifies the above predictions. While we cannot know what m each household would choose in the counterfactual of no reform, we can use pre-reform migration history as a proxy for 'always migrants'.

7 Heterogeneous effects and evidence on mechanisms

Having established that the Hukou reform led to a significant increase in rural fertility, we now turn to the mechanisms driving this result. The theoretical model in Section 6 suggests that the reform operates through multiple channels: it affects migration types, raises expected full income, and alters the shadow price of children depending on childcare availability. To disentangle these forces, we first ask *whose* fertility increased in response to the reform. By identifying the specific subgroups that drove the aggregate fertility boom, we can better understand the mechanisms.

7.1 Heterogeneous effects

We first examine whether the fertility response varies by the household's capacity to benefit from urban opportunities. The model predicts a more positive fertility impact on households that were already oriented towards migration ('always migrants') compared to rural stayers. These households are best positioned to capture the increased full income ($\bar{y}^m$) and improved access to urban public services without facing increased living costs.

As we cannot observe counterfactual migration choices, we use individuals' work and residence status in 2003 as proxy measures for 'always migrants' ($m > 0$). Table 13 reports the results of this heterogeneity analysis. Consistent with the model's prediction, the point estimates are larger for households in which potential parents had migration experience prior to the reform, particularly for those with fathers who had prior migration experience. These patterns are consistent with the idea that 'always migrants' are better positioned to benefit from urban services and higher lifetime incomes, while the higher costs of raising children in urban areas are less relevant to their migration decisions because they would have migrated regardless of the Hukou reform. We also formally assess whether these differences are statistically distinguishable across groups, with the results reported in Appendix Table 29.

The stronger result among households with migration experience is also consistent with the declining price of children as the main cause of increasing fertility. As discussed in Table 11, a decline in the price of raising children in urban destinations p_n^3 will cause an increase in fertility among those migrating with children ($m = 3$) and not other types of migration pattern ($m =$

0, 1, 2.) We don't have a proxy for households who would always choose $m = 3$, and we cannot rule out the declining price of children as a main cause.

It is important to distinguish between having *past* migration experience and being a *current* migrant. The subgroup result in Table 13 does not necessarily mean that the reform only had a positive impact on migrating households or the main channel of fertility increase is through migration responses. Subsection 7.3 will examine how migration is affected and its role in the overall fertility impact. Table 24 (in Appendix) shows that the treatment effect differs little by whether the household *currently* has a migrant member. Both households with and without current migrants show a positive response of approximately 0.01. This suggests that the fertility increase is driven by the capacity to access future opportunities rather than immediate remittance income or current urban residence.

Table 13: Average treatment effect by 2003 characteristics of potential parents

Sub-sample if the potential father was	sample size	mean of outcome	estimated ATT	standard error
present in 2003	104060	0.035	0.014***	0.005
does not work non-locally for 180+ days	64437	0.030	0.009	0.006
works non-locally for 180+ days	39623	0.042	0.022**	0.009
does not work non-locally for 30+ days	52839	0.030	0.008	0.006
works non-locally for 30+ days	51221	0.040	0.021***	0.008

Sub-sample if the potential mother was	sample size	mean of outcome	estimated ATT	standard error
present in 2003	93561	0.031	0.017***	0.006
does not work non-locally for 180+ days	70120	0.029	0.014**	0.006
works non-locally for 180+ days	23441	0.036	0.017*	0.010
does not work non-locally for 30+ days	65008	0.028	0.016**	0.006
works non-locally for 30+ days	28553	0.036	0.013	0.009

Note: each row uses the same event study specification as in Figure 5 and Table 4. Each row selects a subsample as labeled.'Potential mother' is the youngest adult female in the household. Households without any adult female under 40 are excluded from the top panel. 'Potential father' is the youngest adult male in the household. Households without any adult male under 40 in 2003 are excluded from the bottom panel. 'present in 2003' just means the prospective parent was reported as a member of the household in 2003; they did not have to be physically present or living at home.
*** p<0.01, ** p<0.05, * p<0.10.

While economic capacity determines the potential gains from the reform, household constraints—specifically regarding childcare—determine the cost of realizing them. The model predicts that grandparent availability ($G = 1$) leads to a more positive fertility response by relaxing the time constraint on the mother, thereby neutralizing the substitution effect of rising female

wages.

'Grandparents' can be any adult in the household other than the two parents, who can potentially care for the child. Here we examine a few proxies as potential grandparents available for childcare. First, we split households by whether there is one or multiple generations of adults. About two thirds of the entire sample were households with multiple generations of adults, and they have a much higher mean of the outcome variable (0.037 vs 0.006 in Table 14). The ATT is 0.014 in multi-generational households, which is statistically significant and more than one third of the subsample mean. By contrast, the ATT is near zero in single-generation households.

Table 14: Average treatment effect by presence/absence of potential carers

Sub-sample	sample size	mean of outcome	estimated ATT	standard error
households with healthy people aged≥47	125385	0.033	0.014**	0.005
households with healthy women aged≥47	98951	0.035	0.014*	0.007
households with multiple generations of adults	137069	0.037	0.014**	0.006
households without healthy people aged≥47	77421	0.019	0.008**	0.004
households without healthy women aged≥47	103855	0.021	0.009***	0.004
households without multiple generations of adults	65737	0.006	0.003	0.003

Note: each row uses the same event study specification as in Figure 5 and Table 4. Each row selects a subsample as labelled.
*** p<0.01, ** p<0.05, * p<0.10.

Second, we split households by whether there were healthy women above age 47. We choose 47 as an age threshold because the vast majority of new grandparents are aged 47+.[21] Furthermore, there are very few new mothers above age 47. We define 'healthy' based on self-reports, classifying respondents as healthy if they categorize their health status as excellent or good.

We find that households with a healthy woman aged 47 or above exhibit a larger estimated treatment effect compared with households without such potential caregivers. A similar pattern is observed when we define potential caregivers more broadly by including healthy older men. More generally, households with multiple generations of adults exhibit larger treatment effects, consistent with the idea that extended family support can facilitate fertility responses to migration reform. We formally test these differences across subgroups and report the results in Appendix Table 29.

These differential results by proxies of grandparents availability support the 6th prediction that the fertility response to the reform is more positive if grandparents are available. As grandparents' availability is potentially endogenous, we test whether the reform has any impact on

[21] Among 3-generation households with newborns, the median age of grandparents is 54, with the 10th percentile at 47.

grandparents availability by the same event study approach. Figure 16 in the Appendix finds there is no significant impact on having multiple generations of adults; there is a negative and marginally significant impact on having healthy women over age 47.

There is another potential mechanisms through which grandparents could amplify the fertility response: they may facilitate parent-only migration by caring for left-behind children, a strategy associated with higher fertility. If the reform shifted households' migration choice from staying ($m = 0$) to parent-only migration ($m = 1, 2$), then these switchers' fertility would increase as a result of the switch and so we would see increased fertility especially among households with grandparents. As we detail in Section 7.3 (Figure 18), the reform did not increase the propensity for parent-only migration more for households with grandparents than for those without in the short term.[22] Since grandparents did not significantly shift the *migration* decision, their strong impact on fertility is best explained by their role in mitigating the substitution effect—allowing households to enjoy the income gains of the reform without the prohibitive time costs of childcare.

These patterns align with broader evidence of rising parental co-residence in both developed and developing countries (Ruggles and Heggeness, 2008; Kaplan, 2012; Dettling and Hsu, 2018; Bau, 2021; Chen *et al.*, 2023). In high-income economies such as the United States, young adults are increasingly likely to return to the parental home in response to adverse labour market conditions (Kaplan, 2012; Dettling and Hsu, 2018). By contrast, in developing countries like China, older adults continue to rely heavily on their adult children for intergenerational care and old-age support (Chen, 2017; Cheng *et al.*, 2018; Chen *et al.*, 2023). These institutional and economic realities generate strong incentives for partial rather than whole-household migration, particularly in settings with limited formal social security for the elderly. Accordingly, our model—which jointly endogenizes partial migration and fertility decisions—can be readily applied to other low- and middle-income contexts where extended families play a central role in shaping migration and demographic behaviour.

7.2 The economic drivers: Income and expectations

Having established that the fertility response is concentrated among households with the capacity to migrate and available childcare, we now examine the economic forces driving this decision. The model posits that the reform stimulates fertility through a positive income effect—stemming from both immediate wage gains and improved long-term expectations of full income ($\bar{y}$)—that dominates any substitution effect from the opportunity cost of time. We exam-

[22] In the longer term, the reform increases the propensity for one-parent migration, but it does not differentially raise the likelihood of two-parents migration for households with grandparents relative to those without.

ine three strands of evidence to support this channel: migrant wages, household consumption, and subjective well-being.

First, we examine the impact of the reform on daily wages by gender. Existing studies have documented the positive effects of Hukou reform on wages.[23] In our model, male and female wages have different implications for fertility. An increase in the male wage produces a pure income effect, whereas an increase in the female wage generates both an income effect and a countervailing substitution effect (via the opportunity cost of childcare).

In the rural survey, the wage measure is one's earnings from 'non-local work' divided by the number of days they work non-locally.[24] In addition, we cannot observe individuals wages many years after urban settlement, which may increase more than in the short-term. Thus, the wage measure in our data may under-state the real increase in potential wage. We apply the same event study approach to estimate the impact of Hukou reform on wages of different groups of individuals. Table 15 reports the event study estimates on daily wages from non-local work. We find a significant positive ATT on men's wages of 5.2% for the full sample, rising to 7.4% for lower-educated men. In contrast, the effect on women's wages is smaller and statistically insignificant. These differential results support the hypothesis that the fertility boom is driven by a male-specific income shock that is not dampened by a simultaneous rise in the opportunity cost of women's time.

Table 15: Average treatment effect on daily wage from non-local work

Outcome	gender	sample size	mean of outcome	estimated ATT	standard error
log daily wage from non-local work	men	332505	3.808	0.052	0.036
log daily wage from non-local work	men below 40	146937	3.750	0.035	0.041
log daily wage from non-local work	lower-educated men	57190	3.667	0.074**	0.030
log daily wage from non-local work	women	314004	3.521	0.026	0.041
log daily wage from non-local work	women below 40	130376	3.496	0.009	0.044
log daily wage from non-local work	lower-educated women	57884	3.387	0.041	0.044

Note: The dependent variable is the daily wage from non-local work. Hence adults who do not work non-locally or do not report their earnings and days from non-local work are dropped. This table shows the Average Treatment Effect averaged across all Treated (g,t) cells (ATT) and its standard error, as produced by the did_multiplegt package in STATA. Sample size is number of observations at adult-year level. 'Mean of outcome' is the mean of the outcome variable across adults of the labelled group. 'Lower educated' means having ≤ 9 years of education.
*** p<0.01, ** p<0.05, * p<0.10.

[23] Jin and Zhang (2023) find that the reform significantly increased average wages in the manufacturing sector of urban destinations, while Song and Li (2014) show that it expanded occupation opportunity for local rural workers.

[24] The survey does not contain individuals' wages or earnings except for 'non-local work'.

Second, we verify that this income shock translates into household welfare. Table 16 shows a significant positive ATT on household annual expenditure of 6.0%. This confirms the model's first prediction that the reform increases consumption through higher full income. While the estimated impact on current annual income is positive but insignificant, this is consistent with the nature of the reform: the "full income" shock encompasses not just immediate cash earnings, but also entitlements to social benefits and future urban earnings, which are captured more accurately by consumption behavior than by volatile short-term income measures.

Table 16: Average treatment effect on other household outcomes

Outcome	sample size	mean of outcome	estimated ATT	standard error
total earnings from non-local work	202806	12794.792	586.291	756.669
total log earnings from non-local work	109892	9.591	0.053	0.048
log household total income	202244	10.230	0.019	0.027
log household net annual income	201425	9.968	0.037	0.026
log household annual expenditure	201561	9.900	0.060**	0.029
log educational expenditure	66223	7.124	-0.092	0.067
remittance from non-permanent family members	16031	2593.872	-752.965	475.691

Note: each row uses the same event study specification as in Figure 5 and Table 4. Each row uses a different outcome variable. 'Mean of outcome' is the mean of the respective outcome variable in the sample. 'Net annual income' is households total income minus costs of family business (agriculture), taxes and levies. 'Number of people' by age group counts all household members regardless of where they work or live, as long as they are reported as permanent household members by the survey respondent. See details in section 3. Sample size is number of observations at household-year level. Missing values are common for some financial outcomes.
*** p<0.01, ** p<0.05, * p<0.10.

We also investigate educational expenditure. While the pure income effect suggests this should rise, the reform may also alter access to education and returns to education. Empirically, we find a statistically insignificant negative impact. However, given that educational expenditure data is missing for nearly two-thirds of the sample, we interpret this null result with caution; it simply suggests that there is no massive reallocation of resources away from education that would contradict the income channel.

Third, the reform may influence fertility decisions—particularly among rural stayers—by shaping expectations of future status and security. Since income expectations cannot be directly measured, we draw on alternative indicators of subjective well-being from the Chinese General Social Survey (2010–2015) to proxy rural stayers' overall expectations about their future. Specifically, we employ two measures. First, based on the survey question "Overall, would you say you are happy with your life?"—ranging from 1 (very unhappy) to 5 (very happy)—we construct

an index of life satisfaction. Second, using the question "Overall, what do you think about your current social class identity?"—ranging from 1 (lowest) to 10 (highest)—we construct an index of perceived social status.

Applying an event-study framework, we find that the reform significantly enhanced subjective well-being in the short term (Figures 8 and 9). Importantly, we do not observe comparable effects among individuals born with non-agricultural Hukou, confirming that this is a specific response to the removal of migration barriers. This suggests that even for households that do not migrate immediately, the reform improves their perceived social standing and expected lifetime utility, thereby encouraging fertility through an expectations channel.

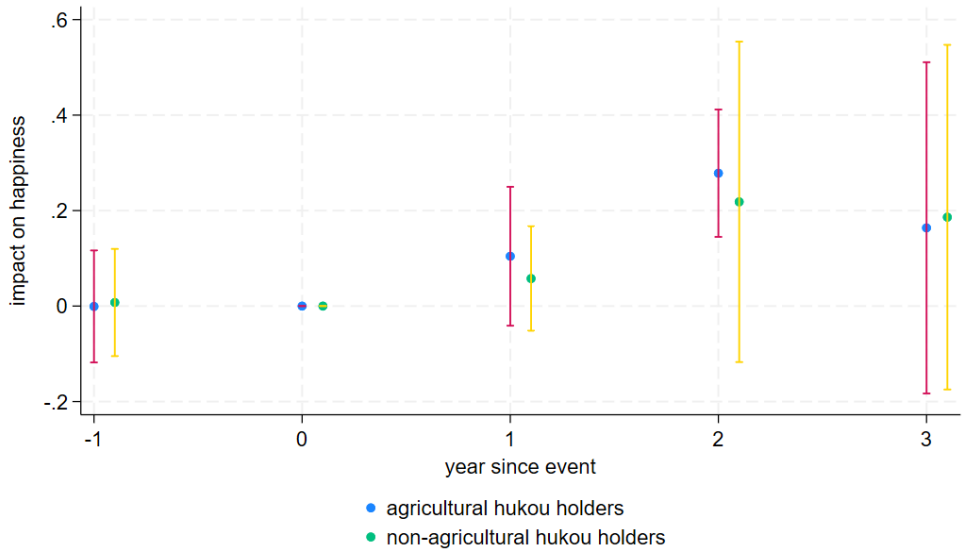


Figure 8: Effect of reform on current life satisfaction by Hukou type

Note: We use the Chinese General Social Survey (2010–2015) for this exercise and apply the same event-study method as in Figure 5. Individual characteristics, including gender, age, and education, are controlled for.

In summary, these corroborating evidence support the interpretation of the fertility increase being a result of positive income effect. The reform plausibly also reduced the effective price of children by improving access to public services in destination cities. We view the price-of-children channel as theoretically consistent with the setting and present in our model; but it's harder to prove empirically.

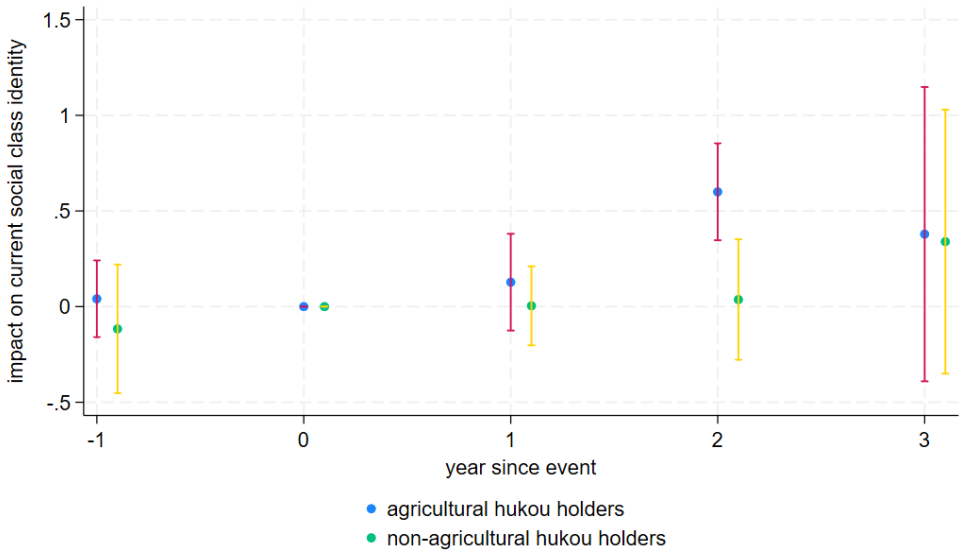


Figure 9: Effect of reform on current social class identity by Hukou type

Note: We use the Chinese General Social Survey (2010–2015) for this exercise and apply the same event-study method as in Figure 5. Individual characteristics, including gender, age, and education, are controlled for.

7.3 Impact of reform on types of migration

Finally, we examine migration behavior. This serves two purposes: first, to confirm that the Hukou reform successfully lowered mobility barriers (the "first stage"); and second, to test whether the observed fertility increase is a result of households switching into high-fertility migration configurations.

We first analyze the impact of the reform on aggregate migration inflows with Census data. Census data is preferred over the rural household survey for examining migration as the outcome variable, because migration is associated with attribution from the rural survey.[25] Using the 2005, 2010, and 2015 population censuses, we impute annual migration inflows from rural areas to prefecture-level cities.[26]

Figure 10 presents the event-study estimates. Consistent with Jin and Zhang (2023), the reform significantly increased adult migration, peaking five years post-implementation. Notably,

[25]For example, if a family moves to the urban area without leaving any family members behind, then we stop observing the household but we cannot distinguish that from other causes of survey non-response.

[26]While Jin and Zhang (2023) construct migration inflows for 1999–2009, we extend their approach to 1999–2015. We define inflow as the number of agricultural *Hukou* holders moving from registered rural locations to city s in year t. As a result, this measure does not count children born to rural parents in urban areas as inflows, since such individuals do not migrate from a registered rural location. Measurement error considerations are detailed in Jin and Zhang (2023).

the reform generated an even stronger response for children: the ATT over 10 years is 37.3% for those aged 0–17, compared to 28.7% for adults. The effect is particularly pronounced for pre-school children (40.8%). This confirms the model's prediction that reducing urban barriers incentivizes whole-family migration ($m = 3$)[27] alongside labor migration.

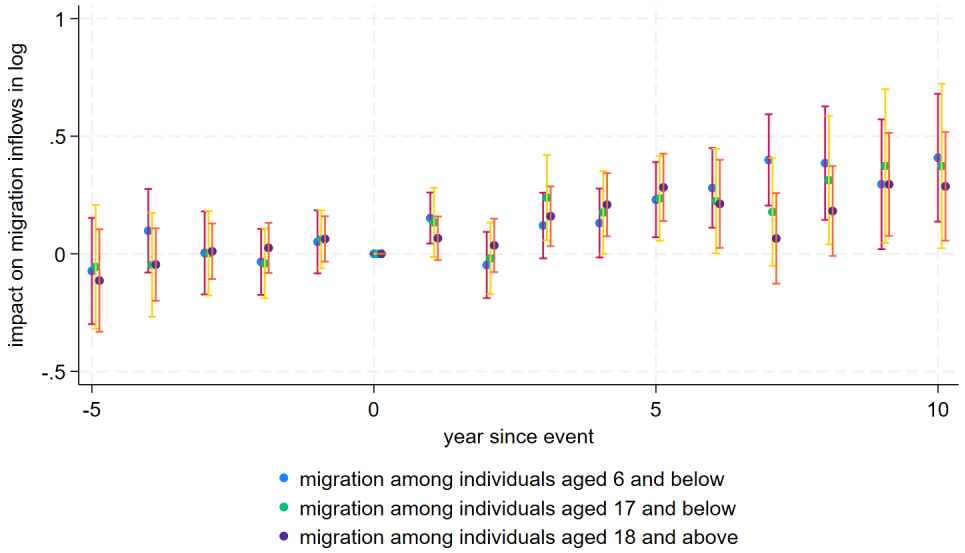


Figure 10: Effect of reform on migration inflows by age

Note: Estimates based on Chinese Population Censuses (2005, 2010, 2015). The outcome is the log of annual migration inflows. The same event study method is applied as in Figure 5.

We corroborate these inflow patterns by exploring the impact of the reform on the number of household members by age group using the rural survey data. Table 17 (in Appendix) checks which age groups are more likely to disappear from the rural household, conditional on the household still appearing in the survey. We find that the reform increases the number of children under school age, which is consistent with our main finding of increased newborns. Moreover, there is a significant negative impact on the number of school-aged children (7–17), and a much smaller and insignificant impact on the number of adults.

This pattern is consistent with the hypothesis that, attracted by access to better urban schools, rural migrants have more babies, let them be cared for in rural areas before school age, and take them to urban areas once they reach school age. Crucially, this response does not necessarily require an increase in the parents' probability of migration; parents who would have migrated regardless of the reform are now incentivized to bring their school-aged children along, while the fertility increase manifests in the babies who temporarily remain rural.

[27]Throughout the paper, we use "parents and children migrate" interchangeably with "whole-family migration."

Table 17: Average treatment effect on number of adults and children in the household

Outcome	sample size	mean of outcome	estimated ATT	standard error
number of people aged 0-6	202806	0.224	0.020	0.015
number of people aged 7-17	202806	0.446	-0.049**	0.022
number of people aged 18-40	202806	1.368	-0.014	0.027
number of people aged 41-60	202806	1.253	0.002	0.023
number of people aged 61+	202806	0.570	0.001	0.015
number of adults working away for 180+days	202806	0.749	0.040	0.029
whether anyone works away for 180+days	202806	0.456	0.004	0.014

Note: each row uses the same event study specification as in Figure 5 and Table 4. Each row uses a different outcome variable. 'Mean of outcome' is the mean of the respective outcome variable in the sample. 'Net annual income' is households total income minus costs of family business (agriculture), taxes and levies. 'Number of people' by age group counts all household members regardless of where they work or live, as long as they are reported as permanent household members by the survey respondent. See details in section 3. Sample size is number of observations at household-year level. Missing values are common for some financial outcomes.
*** p<0.01, ** p<0.05, * p<0.10.

Next, we examine household migration configurations in Census data. We consider each wave of census as a single period, and define the outcome variable as the share of specific type of migration by household, in a classification that is similar to the theoretical model.[28] Using the 2000, 2005, 2010, and 2015 population censuses (Figure 17), we find that the share of households in which "All stay rural" declined from 88% in 2000 to 59% in 2015. By contrast, the share of households in which "parents and children migrate" and "parent-only migration" increased from 6.6% to 29% and from 4.5% to 10%, respectively, over the same period.

Applying the same event-study framework (Figure 11), we further find that the reform reduced the prevalence of "all stay rural" (3.0 percentage points in the first wave of census after the reform) while raising the prevalence of "parents and children migrate" (2.2 percentage points in the first wave of census after the reform) among rural households with two adults and children. The estimated effects on one-parent and two-parents migration are both 0.3 percentage points and statistically insignificant. The sharp decline in the share of households in which all members remain rural, together with the rise in whole-family migration, echoes the age-specific migration inflow patterns documented in Figure 10.

Overall, population census data analyzed within the event-study framework provide robust

[28] We classify five types of migration among rural households consisting of two parents and children: "all stay rural", "one-parent migrates", "parents migrate while at least one child remains rural", parents and children migrate", "other". We restrict samples to those prospective parents aged 35 and below to mitigate the concern of labor migration of adult children.

evidence that the reform substantially increased rural-to-urban migration among both adults and children, and the pattern of migration configuration shifted primarily from "all stay rural" to "parents migrate with children". These findings support the 4th model prediction that migration with children should increase after the reform.

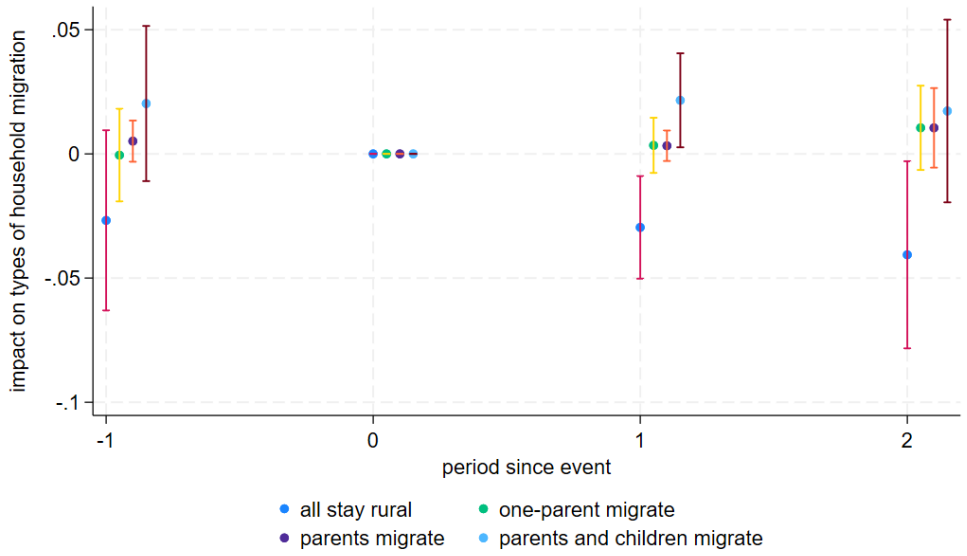


Figure 11: Effect of reform on types of household migration

Note: Estimates based on Chinese Population Censuses (2000–2015). A census wave is assigned to event time 0 if the reform occurred between the previous census and that census wave. For example, if a reform took place in $t \in (2000, 2005]$, the 2005 Census is coded as event time 0.

These shifts in migration configurations may serve as a mediating channel for the observed fertility increase. To evaluate this, we must first understand the baseline relationship between household arrangements and fertility. Figure 12 illustrates the correlation between specific migration types and fertility. Consistent with the model's third prediction, parent-only migration ($m = 2$)—where children remain in rural areas, often with grandparents—is associated with significantly higher fertility than non-migration. Meanwhile, whole-family migration ($m = 3$) is associated with fertility levels quite similar to "all stay rural" households. As the shift in migration configuration is primarily from "all stay rural" to "parents and children migrate", and the change in "parent-only migrate" is insignificant, the main finding of a fertility increase is NOT a mechanical result of households sorting into higher-fertility migration types. In other words, shifts in migration configurations are not a major mechanism for the finding of a positive fertility effect.

To corroborate, we re-estimate the main event study for fertility while explicitly controlling

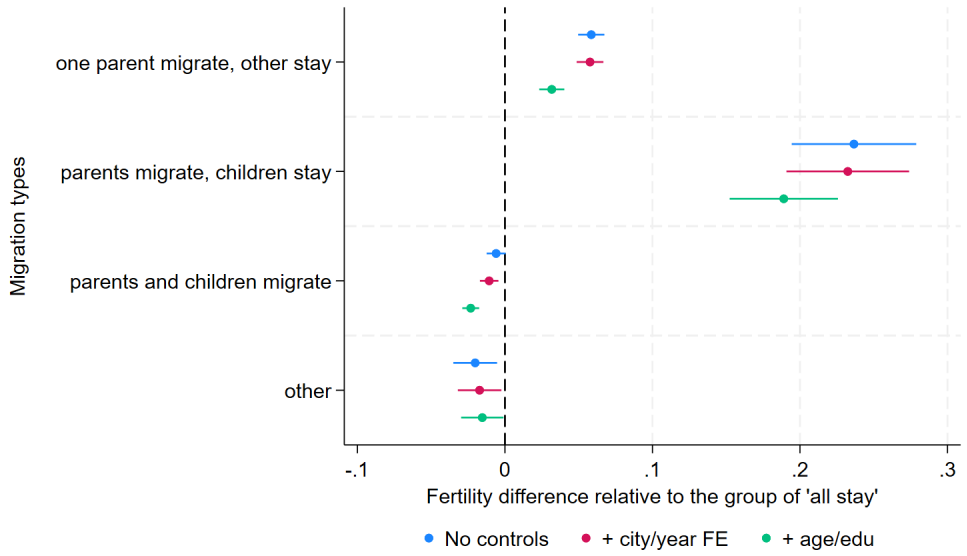


Figure 12: Fertility differences across different types of household migration

Note: We classify five types of migration among rural households consisting of two parents and children: "all stay rural", "one-parent migrates", "parents migrate while at least one child remains rural", parents and children migrate", "other". To mitigate concerns regarding labor migration of adult children, we restrict the sample to prospective parents aged 35 and below, and children aged 17 and below. Using data from the Chinese Population Censuses (2000–2015), the mean of the outcome variable—having any birth in the household in the past year—is 0.098 on average across the four census waves (0.087 in 2000, 0.101 in 2005, 0.109 in 2010, and 0.104 in 2015), which is notably higher than that in the main rural sample due to sample selection for this analysis. Coefficients represent the likelihood of having a newborn relative to the "all stay rural" benchmark group, controlling for city and year fixed effects.

for the current migration status of the parents. As shown in Table 18, rather than shrinking, the estimated fertility impact actually increases from 1.1 to 2.8 percentage points when we account for the non-local work and residence status of the prospective parents. This indicates that the fertility boom is not a compositional artifact of migration. Once the migration channel is accounted for, the underlying stimulus provided by the reform to the population with rural origins as a whole becomes even more pronounced.

8 Conclusion

This paper provides new causal evidence that reducing internal migration barriers can raise fertility in *sending* rural regions. Exploiting the staggered roll-out of China's Hukou reform across 283 cities, we find that the reform led to a significant and persistent increase in fertility among rural households. The average treatment effect on the treated is 0.011 newborns per household

Table 18: Effect of reform on fertility, controlling for current migration status

Outcome	sample size	mean of outcome	estimated ATT	standard error
No controls	202806	0.027	0.011***	0.004
Control for parent non-local residence	202806	0.027	0.028***	0.008
Control for parent non-local work	202806	0.027	0.028***	0.008

Note: All rows in this table are computed using the same sample. 'ATT' is the Average Treatment Effect averaged across all Treated (g, t) cells. ATT and standard errors are produced by the did_multiplegt package in STATA. Sample size is number of observations at household-year level. 'Mean of outcome' is the mean of the outcome variable in the sample. Parent non-local residence or work refers to dummy variable controls indicating whether the prospective mother or father lived or worked away for at least 180 days in the current year.
*** p<0.01, ** p<0.05, * p<0.10.

per year—a substantial effect representing one-third of the baseline mean. By focusing on the source population rather than just urban migrants, we capture the effect of the reform on the vast majority of rural-origin households.

We interpret these findings through a unified household framework that endogenizes both fertility and partial migration. Our empirical results suggest that the fertility boom was driven by a dominant *income effect* that outweighed the substitution effect of higher opportunity costs. A concurrent reduction in the effective price of children—through improved access to urban education and health services as settlement becomes feasible—may have complemented these fertility gains, although we do not have definitive evidence for the price channel. Consistent with the income channel, we document significant post-reform increases in household consumption and male wages. Crucially, we find that the fertility response is highly heterogeneous: it is concentrated in households with prior migration experience (who had the capacity to capture urban income gains) and those with available grandparents (who faced lower childcare costs).

The role of extended family structures proves critical in mediating the policy's impact. While grandparent availability did not significantly alter the propensity to migrate, it strongly predicted the fertility response. This indicates that informal childcare provision allowed households to realize the benefits of the positive income shock without being penalized by the time costs of child-rearing.

These results challenge the simple view that exposure to urbanization necessarily depresses fertility. By reducing the uncertainty and cost of future settlement, the Hukou reform raised the option value of migration, releasing suppressed demand for children. This highlights that migration policy is effectively demographic policy: in developing settings where extended family support is strong, reducing the frictions of spatial mobility can create space for more optimistic

demographic behaviour.

Looking ahead, our findings underscore the importance of looking beyond the destination city when evaluating migration policy. The reforms altered expectations and economic dynamics for the majority of the rural-origin population who remained in the countryside. Future research could usefully explore how these expanded opportunities affect not just the quantity of children, but the long-term human capital investments and educational attainment of the next generation in rural areas.

References

Adamopoulos, T., Brandt, L., Chen, C., Restuccia, D. and Wei, X. (2024). Land security and mobility frictions. *The Quarterly Journal of Economics*, **139** (3), 1941–1987.

Amuedo-Dorantes, C. and Arenas-Arroyo, E. (2021). Immigration policy and fertility: Evidence from undocumented migrants in the US. *Journal of Economic Behavior & Organization*, **189**, 274–297.

—, Borra, C. and Rivera-Garrido, N. (2023a). Fertility implications of family-based regularizations. *Journal of Economic Geography*, **23** (2), 449–484.

—, Ibanez, A., Rozo, S. and Traettino, S. (2023b). More Benefits, Fewer Children: How Regularization Affects Immigrant Fertility.

Anelli, M. and Balbo, N. (2021). Fertility drain or fertility gain? emigration and fertility during the great recession in italy. *Demography*, **58** (2), 631–654.

Athey, S. and Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, **226** (1), 62–79.

Avitabile, C., Clots-Figueras, I. and Masella, P. (2014). Citizenship, fertility, and parental investments. *American Economic Journal: Applied Economics*, **6** (4), 35–65.

Bau, N. (2021). Can policy change culture? government pension plans and traditional kinship practices. *American Economic Review*, **111** (6), 1880–1917.

Benjamin, D., Brandt, L. and Giles, J. (2005). The evolution of income inequality in rural china. *Economic Development and Cultural Change*, **53** (4), 769–824.

Borusyak, K., Jaravel, X. and Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, **91** (6), 3253–3285.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, **225** (2), 200–230.

Casarico, A., Facchini, G. and Frattini, T. (2018). What drives the legalization of immigrants? evidence from irca. *Regional Science and Urban Economics*, **70**, 258–273.

Chen, S., Jin, Z. and Prettner, K. (2023). Can i live with you after i retire? retirement, old age support and internal migration in a developing country. *Canadian Journal of Economics/Revue canadienne d'économique*, **56** (3), 964–988.

— and Wang, J. (2014). Rural tax reform: agricultural productivity and local public finance. *Available at SSRN 2374923*.

Chen, X. (2017). Old age pension and intergenerational living arrangements: a regression discontinuity design. *Review of Economics of the Household*, **15** (2), 455–476.

— and Zhang, X. (2009). The distribution of income and well-being in rural china: a survey of panel data sets, studies and new directions.

Chen, Y. and Fang, H. (2021). The long-term consequences of china's "later, longer, fewer" campaign in old age. *Journal of Development Economics*, **151**, 102664.

— and Feng, S. (2013). Access to public schools and the education of migrant children in china. *China Economic Review*, **26**, 75–88.

— and — (2017). Quality of migrant schools in china: Evidence from a longitudinal study in shanghai. *Journal of Population Economics*, **30** (3), 1007–1034.

Chen, Z., Zhang, J. and Zhou, K. (2025). Endogenous migration restrictions and population aging in china. *Journal of Development Economics*, p. 103674.

Cheng, L., Liu, H., Zhang, Y. and Zhao, Z. (2018). The heterogeneous impact of pension income on elderly living arrangements: evidence from china's new rural pension scheme. *Journal of Population Economics*, **31** (1), 155–192.

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

Dettling, L. J. and Hsu, J. W. (2018). Returning to the nest: Debt and parental co-residence among young adults. *Labour Economics*, **54**, 225–236.

Doepke, M., Hannusch, A., Kindermann, F. and Tertilt, M. (2023). The economics of fertility: A new era. In *Handbook of the Economics of the Family*, vol. 1, Elsevier, pp. 151–254.

Dong, X., Liang, Y. and Zhang, J. (2023). Fertility responses to the relaxation of migration restrictions: evidence from the Hukou reform in China. *China Economic Review*, **81**, 102040.

Facchini, G., Liu, M. Y., Mayda, A. M. and Zhou, M. (2019). China's "great migration": The impact of the reduction in trade policy uncertainty. *Journal of International Economics*, **120**, 126–144.

Fan, C. C., Sun, M. and Zheng, S. (2011). Migration and split households: A comparison of sole, couple, and family migrants in beijing, china. *Environment and Planning A: Economy and Space*, **43** (9), 2164–2185.

Fang, T. and Lin, C. (2015). Minimum wages and employment in china. *IZA Journal of Labor Policy*, **4** (1), 22.

Fang, Y. and Shi, Z. (2018). Children of migrant parents: Migrating together or left behind. *Habitat international*, **76**, 62–68.

Fitzgerald, J., Gottschalk, P. and Moffitt, R. A. (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics.

Gai, Q., Guo, N., Li, B., Shi, Q. and Zhu, X. (2024). Rural pensions, labor reallocation, and aggregate income: An empirical and quantitative analysis of china. *HKU Jockey Club Enterprise Sustainability Global Research Institute Paper*, (2025/002).

Galor, O. (2012). The demographic transition: causes and consequences. *Cliometrica*, **6** (1), 1–28.

Gao, X., Liang, W., Mobarak, A. M. and Song, R. (2023). *Migration Restrictions Can Create Gender Inequality: The Story of China's Left-Behind Children*. Working Paper 30990, National Bureau of Economic Research.

Goodman-Bacon, A. (2018). *Difference-in-differences with variation in treatment timing*. Tech. rep.

Guo, R., Yi, J., Zhang, J. and Zhang, N. (2025). Rationed fertility: Treatment effect heterogeneity in the child quantity–quality tradeoff.

Guo, Z., Wu, Z., Schimmele, C. M. and Li, S. (2012). The effect of urbanization on china's fertility. *Population Research and Policy Review*, **31** (3), 417–434.

Imai, K. and Kim, I. S. (2018). On the use of two-way fixed effects regression models for causal inference with panel data.

Jensen, E. and Ahlburg, D. (2004). Why does migration decrease fertility? evidence from the philippines. *Population studies*, **58** (2), 219–231.

Ji, Q. and Nie, S. (2024). Impact of relaxing internal-migration restrictions on agricultural tfp of rural households: evidence from china. *Applied Economics Letters*, pp. 1–5.

Jin, Z. (2024). State retreats and labor market advances? evidence from china. *Evidence from China*.

— and Zhang, J. (2023). Access to local citizenship and internal migration in a developing country: Evidence from a Hukou reform in China. *Journal of Comparative Economics*, **51** (1), 181–215.

Kaplan, G. (2012). Moving back home: Insurance against labor market risk. *Journal of Political Economy*, **120** (3), 446–512.

Knight, J. and Song, L. (1999). Employment constraints and sub-optimality in chinese enterprises. *Oxford Economic Papers*, **51** (2), 284–299.

Lanari, D., Pieroni, L. and Salmasi, L. (2020). Regularization of immigrants and fertility in Italy.

Lee, S. (2025). Fertility and urbanization. *Cities*, **163**, 106002.

Liu, Z. (2005). Institution and inequality: the hukou system in china. *Journal of Comparative Economics*, **33** (1), 133–157.

Mookherjee, D., Prina, S. and Ray, D. (2012). A theory of occupational choice with endogenous fertility. *American Economic Journal: Microeconomics*, **4** (4), 1–34.

Naufal, G. S. (2015). Impact of remittances on fertility. *IZA World of Labor*.

Pierce, J. R. and Schott, P. K. (2016). The surprisingly swift decline of us manufacturing employment. *American Economic Review*, **106** (7), 1632–1662.

Qian, X. and Smyth, R. (2008). Measuring regional inequality of education in china: widening coast–inland gap or widening rural–urban gap? *Journal of International Development: The Journal of the Development Studies Association*, **20** (2), 132–144.

Rathnayake, G., Negi, A., Bartalotti, O. and Zhao, X. (2024). Difference-in-differences with sample selection. *arXiv preprint arXiv:2411.09221*.

Ruggles, S. and Heggeness, M. (2008). Intergenerational coresidence in developing countries. *Population and development review*, **34** (2), 253–281.

Song, J. and Li, S. (2014). Hukou's impact on labor occupation segmentation. *China Agricultural Economic Review*, **6** (3), 506–522.

Sun, L. and Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.

Tian, Y. (2024). International trade liberalization and domestic institutional reform: Effects of wto accession on chinese internal migration policy. *Review of Economics and Statistics*, **106** (3), 794–813.

Tumen, S. (2012). Fertility decisions and endogenous residential sorting. *Regional Science and Urban Economics*, **42** (1-2), 78–87.

White, M. J., Muhidin, S., Andrzejewski, C., Tagoe, E., Knight, R. and Reed, H. (2008). Urbanization and fertility: An event-history analysis of coastal ghana. *Demography*, **45** (4), 803–816.

Yang, S., Jiang, Q. and Sánchez-Barricarte, J. J. (2022). China's fertility change: an analysis with multiple measures. *Population Health Metrics*, **20** (1), 12.

Yue, A., Tang, B., Shi, Y., Tang, J., Shang, G., Medina, A. and Rozelle, S. (2018). Rural education across china's 40 years of reform: past successes and future challenges. *China Agricultural Economic Review*, **10** (1), 93–118.

Zhang, J. (2017). The evolution of china's one-child policy and its effects on family outcomes. *Journal of Economic Perspectives*, **31** (1), 141–160.

9 Appendix

9.1 Calculation of population shares

To determine the relative size of the four groups illustrated in Figure 1, we will use information from our data and census statistics. Denote the shares of the four groups as $x1, x2, x3, x4$, and $x1 + x2 + x3 + x4 = 1$.

Our survey provides the relative sizes of the first three groups: non-migrants, left-behind family members, and connected migrants. We count an individual as a migrant if they lived at home for under 180 days of the years. In 2010, the percentage of these three groups in our data are $47\%, 31\%, 22\%$. These percentages correspond to $x1/(x1 + x2 + x3), x2/(x1 + x2 + x3), x3/(x1 + x2 + x3)$. In 2005, the percentages are $52\%, 29\%, 19\%$. It may be easier to work with 2005 numbers because there were fewer migrants who got non-agricultural Hukou.

Meanwhile, we can get an estimate of $x3 + x4$ from Census statistics. The share is not precisely observed because Census did not ask about individual's place of birth or where they lived in 2000 or just before the reform. Ideally, the 'originally-rural population is defined as everyone who was born in a rural village, or everyone who was living in a rural village in 2000 (before the reform). And $x3 + x4$ is the share of them who in 2010 did not live in their rural village for the majority of

the time (days at home < 180). Now our question is simply what statistic/measure from census is the best proxy? We want to use as few assumptions as possible.

According to the National Bureau of Statistics of China, in 2010, 960 million people had Agricultural Hukou and 373 million had non-agricultural Hukou. Meanwhile 674 million lived in rural areas and 666 million people lived in urban areas.[29] This means at least $1 - 674/960 = 29.8\%$ of Agriculture-Hukou holders were migrants. We take this as a proxy for the share of rural-origin population who were migrants.

Suppose n million people were born with Agricultural Hukou and obtained non-agricultural Hukou before 2010, then the true share of migrants is $n + 960 - 674/(n + 960)$. This is increasing in n. To give a high estimate, $n = 20$ would translate to 31.2% migrant share.

So we have four unknowns and four equations:

$$\begin{aligned} x1/(x1 + x2 + x3) &= 47\% \\ x2/(x1 + x2 + x3) &= 31\% \\ x3 + x4 &= 30\% \\ x1 + x2 + x3 + x4 &= 1 \end{aligned}$$

The solution is $x1, x2, x3, x4 = 42\%, 28\%, 20\%, 10\%$. If $n = 20$, the solution for x4 would be 11.3%.

9.2 Additional descriptives

Table 19 shows summary statistics for adults' characteristics and labor supply, by gender and broad age group (split at age 45). For education, there are large differences across generations. In China, the first national law on compulsory education was introduced in 1986. But the goal of universal 9-year compulsory education took many years to materialise in rural areas.[30] Table 19 shows younger adults (below 45) have 8-9 years of education on average, whereas the older ones have 5-6 years of education only.

In terms of labor supply, the NFPS data contains the number of days each household member spends in three distinct categories: local agricultural work, local non-farm work, and non-local work. Table 19 shows these labor supply measures by gender and two broad age groups. It's worth noting that young men (age below 45) have the highest total number of working days, followed by young women and older men, and the last is older women. Moreover, the pattern

[29] Source: the 2010 National Population Census, https://www.stats.gov.cn/sj/pcsj/rkpc/6rp/indexch.htm

[30] Yue *et al.* (2018) documented a rise in the enrollment rate of 13-15-year old in secondary schools, from a low of 60 percent in 1990, to around 90 percent by 2000, and 100 percent by 2010.

Table 19: Summary statistics for adults at individual-year level

	male age under 45		female age under 45	
variable	mean	std dev	mean	std dev
age	30.668	8.130	31.005	8.308
years of education	8.855	2.650	8.265	2.840
number of days lived at home	221.950	147.122	258.613	142.061
live away for 180+ days	0.423	0.494	0.313	0.464
total number of days worked	218.278	126.220	178.844	134.865
days worked in local agriculture	43.719	81.713	55.424	92.076
days worked in local non-agricul	45.630	99.065	37.052	90.041
days worked non locally	134.382	143.105	90.419	135.168
health status:excellent=1;good=2	1.289	0.567	1.336	0.580
	male age over 45		female age over 45	
age	58.941	9.095	58.640	9.498
years of education	6.515	2.671	4.917	2.812
number of days lived at home	330.023	86.892	345.232	67.201
live away for 180+ days	0.092	0.290	0.047	0.212
total number of days worked	175.798	128.258	117.274	123.721
days worked in local agriculture	84.075	100.929	75.755	98.768
days worked in local non-agricul	52.314	99.201	28.486	77.206
days worked non locally	43.166	96.769	14.783	62.399
health status:excellent=1;good=2	1.969	1.074	2.148	1.123

Note: This table is based on our cleaned sample over 2003-2015, at the individual-year level, by gender and age group.

of work differs a lot between young and old and by gender. Working non-locally is much rarer among older people than younger ones. Young men on average work the most days away from home.

The survey asks the number of days each person lived at home. Table 19 shows that the median number of days at home is 365 for both older men and older women (age over 45), 360 for younger women, and 300 for younger men. But the mean is much lower than the median, especially among the younger groups. In other words, there are a small group of adults who live away from home for long periods, but the majority of adults do live at home for most of the year. Because these rural households are large, at the household level about half of households had at least one person who lived at home for under 180 days, which was shown in Table 1.

Table 20 presents the total number of households, new entrants, stayers, and the annual exit rate. The year-to-year attrition rate ranges from 1.5% to 16%.

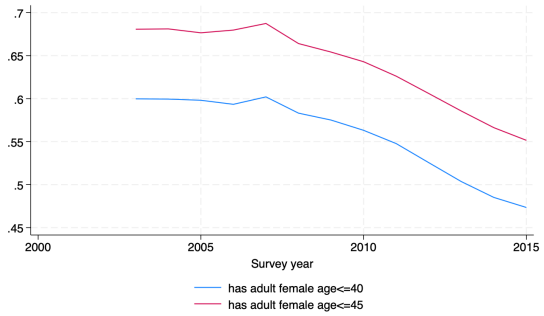


Figure 13: Share of households with younger woman

9.3 Selection bias

Table 21 examines household characteristics between those that stay versus those that disappear in the next year. While some differences are statistically significant, most are small. For instance, drop-out households had slightly more adults under age 40 (1.45 vs. 1.36 among stayers), and fewer older household members. Crucially, the average number of children aged 0–6 is 0.22 among stayers and 0.23 among drop-outs—a difference that is not statistically significant. This suggests that attrition is unlikely to be systematically driven by fertility preferences.

Next, we will use the same terminology as in Rathnayake *et al.* (2024), to illustrate the bias caused by endogenous sample selection. The aim here is to gauge the direction and magnitude of the bias. It is beyond the scope of this paper to formally develop the bounds for treatment effects in the staggered DiD case with heterogenous treatment effects. That is also beyond the scope of the econometric paper Rathnayake *et al.* (2024). Here we will apply their framework to a 2-period DiD context with negative selection. In practice, the ATT from our event study analysis (using de Chaisemartin and D'Haultfoeuille (2020) is a weighted average of hundreds of ATTs. The method averages across ATT only for (g, t) cells whose treatment status changes from $t-1$ to t. This means over 200 g, t cells in our context because each prefecture city experienced the reform only once. If each of those ATTs is biased in the same direction due to endogenous selection, then their weighted average would be biased in the same direction. The rest of the section works out the likely sign and magnitude of bias caused by endogenous selection in a 2-period DiD in the context of endogenous migration and fertility.

In Rathnayake *et al.* (2024), the theoretical framework for selection considers 8 latent groups: OOO,OON,ONO,ONN, NOO,NON,NNO,NNN. The first letter denotes whether the household is observed in the before period, the 2nd letter is whether it's observed in the after period if untreated, the 3rd letter is whether it's observed in the after period if treated. For example, the OON latent group are households Observed before the reform, and Observed in the after period

Table 20: Household entry and exit rates

Year	Number of HHs	New entrants	Stayers	Exit rate
2003	13787	0	0	0.058
2004	14880	1895	12985	0.058
2005	15016	1000	14016	0.061
2006	14766	660	14106	0.052
2007	17036	3042	13994	0.161
2008	14330	35	14295	0.015
2009	16906	2795	14111	0.027
2010	16674	219	16455	0.059
2011	16402	717	15685	0.086
2012	15680	687	14993	0.050
2013	15782	881	14901	0.058
2014	15924	1061	14863	0.083
2015	15623	1027	14596	0.000

Note: This table shows the number of households in our cleaned sample from 2003 to 2015. The 'New entrants' are the number of households who did not appear in preceding year. 'Stayers' are households that were observed in both preceding and current year. The 'Exit' rate indicates the proportion of households in current year that will drop out of the sample next year.

in the absence of the reform, and Unobserved in the after period if treated. They are the people who disappear after the reform because of the reform. In our context, we will assume the reform has a non-positive impact on the probability of being observed. This rules out the ONO latent group does not exist.

We start with the first expression for the DiD estimand in their Lemma 1. Because there is no ONO latent group in our case, we have $p_{OOO1} = 1$ and this simplies the DiD estimand to

$$
\begin{aligned}
\tau_{DiD} &= \tau_{OOO} \\
&\quad + (\mathbb{E}[Y_1^*(0) - Y_0^*(0) \mid D = 0, \mathrm{OOO}]) \\
&\quad - \mathbb{E}[Y_1^*(0) - Y_0^*(0) \mid D = 0, \mathrm{OON}] \\
&\quad + (\mathbb{E}[Y_1^*(0) - Y_0^*(0) \mid D = 0, \mathrm{OON}] - \mathbb{E}[Y_1^*(0) - Y_0^*(0) \mid D = 0, \mathrm{OOO}]) \cdot p^0_{\mathrm{OOO}} \\
&\quad = \tau_{OOO} + (\mathbb{E}[Y_1^*(0) - Y_0^*(0) \mid D = 0, \mathrm{OON}] - \mathbb{E}[Y_1^*(0) - Y_0^*(0) \mid D = 0, \mathrm{OOO}]) \cdot (p_{\mathrm{OOO0}} - 1)
\end{aligned}
$$

where $p_{OOO0} = \pi_{OOO0}/(\pi_{OOO0} + \pi_{OON0})$. π_{OON0} is the frequency of the OON latent group in the control group and π_{OOO0} is the frequency of the OOO latent group in the control group. Thus, $1 - p_{OOO0}$ is the share of the control group (observed both before and after the reform) that would disappear if it were treated.

The equation says the DiD estimand is essentially the ATT for the OOO latent group plus

Table 21: Balance tables for year-on-year attrition

	Stayers	Drop outs	P-value
number of generations in the household	2.148	2.189	0.000
how many generations of adults are present in the household	1.760	1.779	0.001
number of people aged 0-6	0.224	0.228	0.358
number of people aged 7-17	0.443	0.490	0.000
number of people aged 18-40	1.363	1.447	0.000
number of people aged 41-60	1.256	1.206	0.000
number of people aged 61+	0.569	0.593	0.001
age of head	53.614	53.653	0.715
male of head	0.935	0.915	0.000
eduyear of head	6.808	6.834	0.298
homeduration of head	319.378	320.171	0.398
whether anyone in the household has liveaway180==1	0.496	0.481	0.001
whether adult in the household has non-agri hukou	0.096	0.112	0.000
mean of male among adults in households	0.517	0.516	0.448
year of reform adoption	2009.899	2009.482	0.000
Post-reform	0.469	0.430	0.000
year_since_reform	-0.882	-0.837	0.439
Observations	266669		

a bias term. The bias term captures the difference in trends between OON and OOO groups among the untreated population, and multiplied by $(p_{OOO0} - 1)$.

$1 - p_{OOO0}$ is the share of the control group (observed both before and after the reform) that would disappear if it were treated. We believe $1 - p_{OOO0}$ is small. Recall that about one quarter of 2003 observations were not observed in 2015. In the extreme case if all of this one quarter attrition were caused by the reform, we would get $1 - p_{OOO0} = 0.25$. In practice, attrition has many causes and is also prevalent in untreated cases. To gauge the magnitude of $1 - p_{OOO0}$, we estimate the reform's effect on the probability of being observed in the sample. Figure 14 presents the event study results. There is a negative pre-trend before the reform, a small insignificant increase in the first few years after the reform, and an insignificant decline in years 8-10. The estimated effects post-reform are all smaller than 0.1.

In terms of differential trends between latent groups in untreated population, we do not have a strong prior whether the OON or OOO group should have a more positive trend. Intuitively, the OOO group contains always-stayers and some households with individual migrants. The OON group are households who leave the sample due to the reform, and hence they are mostly marginal migrants. Some of the OON group may be 'always migrants', who would be reported by remaining rural parents as members of family in the absence of reform and not reported after

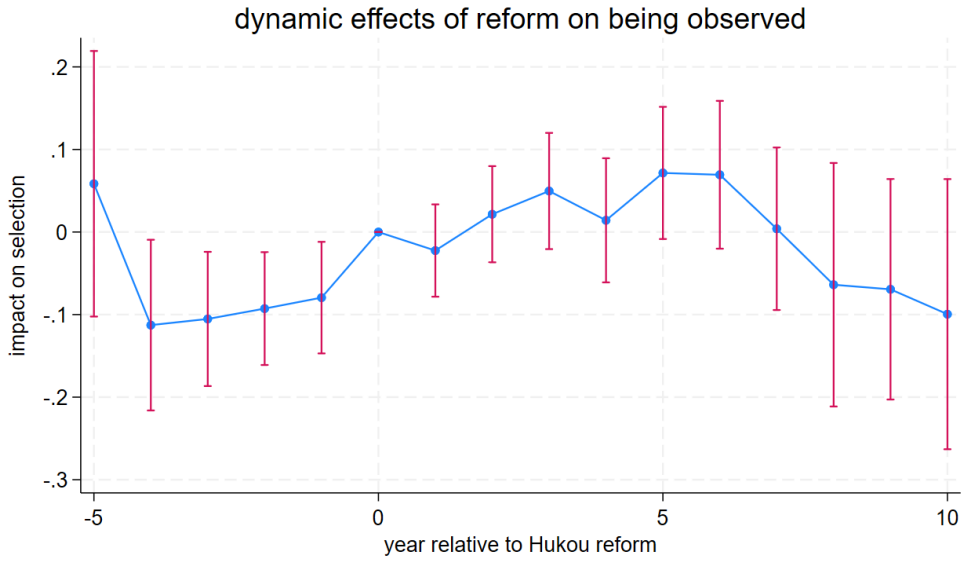


Figure 14: Effect of reform on the probability of households being observed

Note: the outcome variable is whether the household is observed in that year. For households observed in a subset of years, we fill the unobserved years setting the outcome variable to 0. This gives a balanced sample. The graph shows the treatment effects on the outcome of household being observed in the current year. Estimation uses the method from de Chaisemartin and D'Haultfoeuille (2020).

the reform, because the reform makes them more indepedent from rural parents. And some marginal migrants would be in the OOO group, if they have family members stay behind and report them family members. While we cannot identify which households are OOO and which ones are OON, we can look at trends in fertility before reform across subgroups of households. In table 22, we examine households whose reform year is 2012 or later and look at their observations over 2003-08. These households were clearly not yet treated; they were actually years away from being treated. We check whether fertility trend differs by 1) how often they appear in this pre2008 sample, and 2) whether the prospective dad was working or living away in 2003. The first two columns of table 22 looks at households who were observed every year over 2003-08 vs households who were missing some of the years. Neither subsample has a significant time trend. The next four columns distinguish households by whether the prospective father was working non-locally or living at home. They show no trend for the migrant subgroup, and a negative trend for the stayer subgroup. The negative trend is up to -0.2ppts per year. This decline may be due to population aging in rural China, which was discussed in section 3. If the OOO latent group has a negative trend with slope -0.2ppts per year and the OON latent group has a flat trend, then the bias term within the DiD estimand would be $0.002 * (p_{OOO0} - 1)$. If the amount of

attrition caused by reform is 10%,[31] then the bias would be about -0.02ppts. This is a very small bias relative to our main estimate of 1.1ppts. In a more extreme scenario, suppose the difference in trend between OON and OOO groups is 1ppt. This is incredibly big given the sample mean of the outcome variable is 2.7ppts. Even if all of the selection (about 25%) if caused by the reform, 1ppt times -25% gives a bias of -0.25. As our main estimate is 1.1ppts, this means the true ATT on the OOO group (τ_{OOO}) is in the [0.85,1.35] range.

Table 22: Trend in fertility, 2003-2008

	(1) always obs b/se	(2) sometimes missing b/se	(3) not work away b/se	(4) work away b/se	(5) live at home b/se	(6) live away b/se
year	-0.001	0.000	-0.002	-0.001	-0.003	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
_cons	1.224	-0.366	3.345	2.077	5.171	-1.144
	(1.111)	(2.701)	(1.744)	(2.576)	(1.799)	(2.487)
N	33120	4663	15798	9702	16278	9222

Note: we select households who were observed in 2003, and whose reform year was 2012 or later, and look at their observations for years 2003-2008. "Always obs" means household was observed every year over 2003-08; sometimes missing means they were missing for some of those years. The next four columns condition on there being a prospective father observed in 2003. "Not work away" and "work away" refer to whether the prospective father worked non locally for 180+ days. "live away" and "live at home" refer to whether he lived at home for 180+ days.

9.4 Derivation of model predictions

This section derives the solution to the household's optimisation problem described in section 6. The household maximises utility subject to a few constraints:

[31] Based on Figure 14, the impact of the reform on the probability of being observed is no greater than 0.1 in absolute terms, and not always positive or negative, and mostly insignificant.

$$
\begin{aligned}
\max U(c, n, q) &= (1-\gamma)\log c + \gamma(\log n + \beta \log q) \\
s.t. & \\
\log q &= \delta_s + \delta_e \log \frac{E}{n} \\
\tau_w + h_w &= 1 \\
\tau_w &= n\tau(1-G) \\
c + E + p^n n &= w_m + w_w h_w + y_0 \\
\tau_w, h_w, h_m &\geq 0
\end{aligned}
$$

The First-Order-Conditions with regard to $c, \log q, E, h_w, \tau_w, n$ are, in turn, :

$$\frac{1-\gamma}{c} = \lambda_c \tag{11}$$

$$\gamma\beta = \lambda_q \tag{12}$$

$$\lambda_q \delta_e / E = \lambda_c \tag{13}$$

$$\lambda_w = \lambda_c w_w \tag{14}$$

$$\lambda_\tau = \lambda_w \tag{15}$$

$$\frac{\gamma}{n} - \frac{\lambda_q \delta_e}{n} - \lambda_\tau \tau(1-G) - \lambda_c p^n = 0 \tag{16}$$

where $\lambda_c, \lambda_q, \lambda_w, \lambda_\tau$ are the Lagrangian multipliers for the budget constraint, the log quality production function, woman's time constraint, and the childcare time constraint.

Substituting for λ_c, λ_q in the third F.O.C. $\Rightarrow$:

$$
\begin{aligned}
\gamma\beta\delta_e / E &= \frac{1-\gamma}{c} \\
\Rightarrow E &= \frac{\gamma\beta\delta_e c}{(1-\gamma)}
\end{aligned} \tag{17}
$$

This means educational expenditure E will have a fixed ratio to consumption c, and the ratio is product of preference for child quality relative to preference for consumption $\frac{\gamma\beta}{(1-\gamma)}$ and the return to investment in child quality δ_e.

The F.O.C. wrt n can be simplified into:

$$\begin{aligned}
\frac{\gamma}{n} - \frac{\gamma\beta\delta_e}{n} &= \lambda_\tau \tau(1-G) + \lambda_c p^n \\
\frac{\gamma}{n} - \frac{\gamma\beta\delta_e}{n} &= \lambda_c w_w \tau(1-G) + \lambda_c p^n \\
\frac{\gamma - \gamma\beta\delta_e}{n} &= \lambda_c [w_w \tau(1-G) + p^n] \\
\frac{\gamma - \gamma\beta\delta_e}{n} &= \frac{1-\gamma}{c} [w_w \tau(1-G) + p^n]
\end{aligned}$$

Denote

$$p^* = w_w \tau(1-G) + p^n$$

This is the shadow price of each additional child incorporating the time cost of childcare.

The F.O.C. wrt n further simplifies to

$$\frac{\gamma - \gamma\beta\delta_e}{n} = \frac{(1-\gamma)p^*}{c} \tag{18}$$

$$\Rightarrow \frac{c}{n} = \frac{(1-\gamma)p^*}{\gamma - \gamma\beta\delta_e} \tag{19}$$

This gives the ratio c/n as a function of exogenous variables only.

Denote full income

$$\bar{y} = w_m + w_w + y_0$$

The budget constraint is:

$$c + E + p^n n + w_m(1-h_m) + w_w(1-h_w) = w_m + w_w + y_0 = \bar{y} \tag{20}$$

$$c + E + p^n n + w_m \tau_m + w_w \tau_w = \bar{y} \tag{21}$$

$$\tag{22}$$

Substitute for τ_w, τ_m in the last equation, we get

$$c + E + p^n n + 0 + w_m n \tau(1-G) = \bar{y} \tag{23}$$

$$c + E + (p^n + w_m \tau(1-G))n = \bar{y} \tag{24}$$

$$c + E + p^* n = \bar{y} \tag{25}$$

This shows full income has three destinations: consumption, educational expenditure, and quantity of children times its shadow price.

Substitute for E in the above equation:

$$c + c\frac{\gamma\beta\delta_e}{(1-\gamma)} + p^* n = \bar{y} \tag{26}$$

$$c[1 + \frac{\gamma\beta\delta_e}{(1-\gamma)}] + p^* n = \bar{y} \tag{27}$$

$$c[\frac{1-\gamma+\gamma\beta\delta_e}{(1-\gamma)}] + p^* n = \bar{y} \tag{28}$$

Use (19) to substitute for c in the last equation:

$$\begin{aligned}
n\frac{(1-\gamma)p^*}{\gamma-\gamma\beta\delta_e}[\frac{1-\gamma+\gamma\beta\delta_e}{(1-\gamma)}] + p^* n &= \bar{y} \\
np^*\frac{1-\gamma+\gamma\beta\delta_e}{\gamma-\gamma\beta\delta_e} + p^* n &= \bar{y} \\
np^*[\frac{1-\gamma+\gamma\beta\delta_e}{\gamma-\gamma\beta\delta_e} + 1] &= \bar{y} \\
np^*\frac{1}{\gamma-\gamma\beta\delta_e} &= \bar{y} \\
np^* &= (\gamma-\gamma\beta\delta_e)\bar{y} \\
n &= (\gamma-\gamma\beta\delta_e)\bar{y}/p^*
\end{aligned}$$

Thus, the chosen number of children is a fraction of full income divided by its shadow price.

We use (19) to work out c:

$$\begin{aligned}
c &= n\frac{(1-\gamma)p^*}{\gamma-\gamma\beta\delta_e} \\
&= \frac{(\gamma-\gamma\beta\delta_e)\bar{y}}{p^*}\frac{(1-\gamma)p^*}{\gamma-\gamma\beta\delta_e} \\
&= (1-\gamma)\bar{y}
\end{aligned}$$

Finally, we use (17) to get E:

$$\begin{aligned}
E &= \frac{\gamma\beta\delta_e c}{(1-\gamma)} \\
&= \frac{\gamma\beta\delta_e(1-\gamma)\bar{y}}{(1-\gamma)} \\
&= \gamma\beta\delta_e\bar{y}
\end{aligned}$$

9.5 Additional event-study analysis

9.5.1 Impact on fertility

Table 23 shows differential effects by the presence of an adult woman under 40. The results show stronger effects of the policy for households with a woman under 40. By contrast, the effect is consistently close to zero for households without a woman below 40. This is expected since the outcome variable is almost always zero in such households: Table 3 reports a mean of 0.2% in this subsample.

Table 23: Effect of reform on fertility among households by the presence of women under 40

Outcome	sample size	mean of outcome	estimated ATT	standard error
Households without woman aged 18-40	89857	0.002	-0.002	0.002
Households with woman aged 18-40	112949	0.048	0.022***	0.006

Note: All rows in this table are computed using the same sample (households with/without woman aged 18-40). 'ATT' is the Average Treatment Effect averaged across all Treated (g, t) cells. ATT and standard errors are produced by the did_multiplegt package in STATA. Sample size is number of observations at household-year level. 'Mean of outcome' is the mean of the outcome variable in the sample.
*** p<0.01, ** p<0.05, * p<0.10.

Figure 15 examines the impact of the reform on the presence of a woman under age 40. The results show the presence of women had a weakly declining trend in the treated regions both before and after the reform. Thus, we cannot eliminate the concern that the presence of women of child-bearing age is endogenous. Figure 16 examines the impact of the reform on grandparents' availability in rural households, using two proxies. The results show that there is no significant impact on having multiple generations of adults but there is a negative and marginally significant impact on having healthy women over age 47.

Table 24: Average treatment effect by current presence of migrants

Outcome	sample size	mean of outcome	estimated ATT	standard error
households without migrant workers	110413	0.020	0.011***	0.004
households with migrant workers	92393	0.037	0.012*	0.006
households without migrants	102324	0.019	0.010***	0.004
households with migrants	100482	0.036	0.012	0.007

Note: This table shows the Average Treatment Effect averaged across all Treated (g, t) cells (ATT) and its standard error, as produced by the did_multiplegt package in STATA. Sample size is number of observations at household-year level. 'Mean of outcome' is the mean of the outcome variable in the sample. 'Migrant' refers to any person that did not live at home for at least 180 days. 'Migrant worker' means any person that worked non-locally for 180+ days.

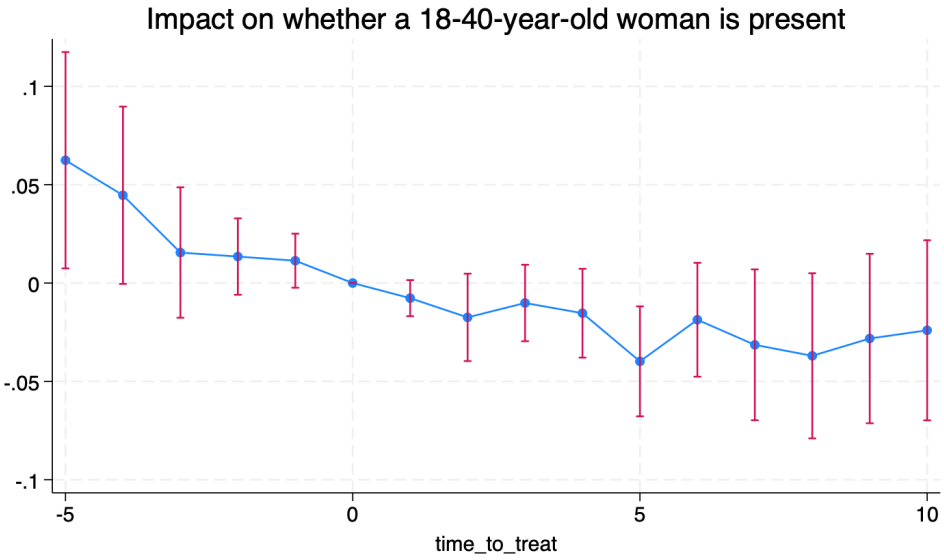


Figure 15: Effect of reform on the presence of women under 40

The same event study method is applied as in Figure 5.

Table 24 shows that the treatment effect differs little by whether the household currently has a migrant member. Both households with and without current migrants show a positive response of approximately 0.01. This suggests that the fertility increase is driven by the capacity to access future opportunities rather than immediate remittance income or current urban residence.

Table 25 splits households by the education level of either the head or the most educated person in the household. Previous research by Jin and Zhang (2023) showed that the reform mainly increased rural to urban migration for lower-skilled adults. As a result, we expect the impact on fertility to be larger in households with lower-skilled adults. The ATT is 0.012 in households where everyone has less than 12 years of education (which corresponds to High School completion), and 0.003 in the rest. It's worth noting that the former subsample has a lower mean of the outcome variable than the latter subsample, so in a proportional sense, the ATT is much bigger in household with less-educated adults.

Table 26 reports results separately by household Hukou type. Prior research by Jin and Zhang (2023) shows that the reform primarily increased rural–urban migration among individuals with agricultural Hukou. Consistent with this pattern, we find larger fertility responses among agricultural-Hukou households. The estimated ATT is 0.012 for households with agricultural Hukou, compared with 0.007 for other households. It is worth noting, however, that only 3.6% of households hold non-agricultural Hukou, so the latter estimates should be interpreted

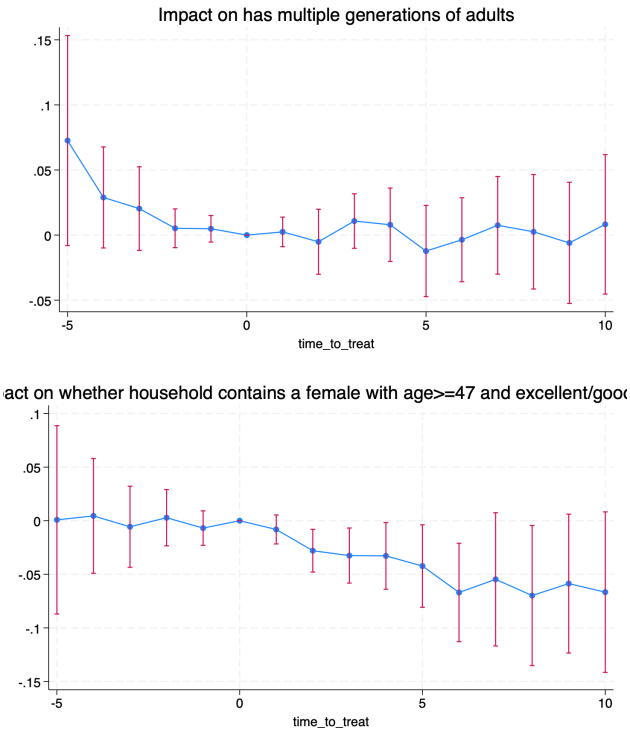


Figure 16: Effect of reform on grandparents availability

The same event study method is applied as in Figure 5. The left graph uses the binary outcome of whether the household has more than 1 generation of adults. The right graph uses the binary outcome of whether household contains a healthy woman aged 47+.

with greater caution.

9.5.2 Impact on migration

Using Chinese Population Censuses (2000, 2005, 2010 and 2015), we provide additional evidence on the impact of the reform on migration patterns. Figure 17 shows the trends of household migration from 2000 to 2015. We find that the share of households in which "All stay rural" declined from 88% in 2000 to 59% in 2015. By contrast, the share of households in which "parents and children migrate" and "parent-only migration" increased from 6.6% to 29% and from 4.5% to 10%, respectively, over the same period.

To further explain the mechanism behind the heterogeneous response identified in Section 7.1—where households with grandparents showed a significantly higher fertility response —we examine how grandparent availability shapes migration choices in response to the reform. While Section 7.1 established that grandparents are a strong predictor of the fertility response, our anal-

Table 25: Average treatment effect by education groups

Sub-sample	sample size	mean of outcome	estimated ATT	standard error
if head has <7 years of education	99624	0.027	0.012**	0.005
if head has 7+ years of education	103182	0.028	0.008*	0.005
if the most educated person has <12 years of education	159207	0.026	0.012***	0.004
if the most educated person has 12+ years of education	43599	0.033	0.003	0.010

Note: This table shows the Average Treatment Effect averaged across all Treated (g, t) cells (ATT) and its standard error, as produced by the did_multiplegt package in STATA. Sample size is number of observations at household-year level. 'Mean of outcome' is the share of observations having a newborn. The median education level for adults in the sample is 7 years. 12 years correspond to high-school graduation, which is 3 years beyond compulsory education.
*** p<0.01, ** p<0.05, * p<0.10.

Table 26: Average treatment effect by Hukou types

Sub-sample	sample size	mean of outcome	estimated ATT	standard error
NON-Agricultural hukou	7338	0.018	0.007	0.007
Agricultural hukou	194940	0.028	0.013***	0.004

Note: This table shows the Average Treatment Effect averaged across all Treated (g, t) cells (ATT) and its standard error, as produced by the did_multiplegt package in STATA. Sample size is number of observations at household-year level. 'Mean of outcome' is the share of observations having a newborn. Only 3.6% of households have nonagricultural Hukou.
*** p<0.01, ** p<0.05, * p<0.10.

ysis in Figure 18 in the Appendix reveals that the reform had no significant effects on migration configuration in the short run; and in the longer run the only statistically significant effect is an increase in the share of households where one-parent migrate and grandparents are available. Even that effect is less than 5% and only marginally significant. Figure 19 further indicates that the reform reduces the share of households in which all members remain in rural areas without grandparent support, while increasing the share of whole-family migration households (parents and children migrate) without grandparent availability. Taken together, grandparents facilitate fertility primarily by lowering childcare costs (mitigating the substitution channel), rather than by inducing a specific shift toward migrant status.

9.6 A description for city-level variables

One challenge to identification is the non-randomness in the timing of reform. To assess this, we collect and construct a range of economic and social statistics for prefecture-level cities. This

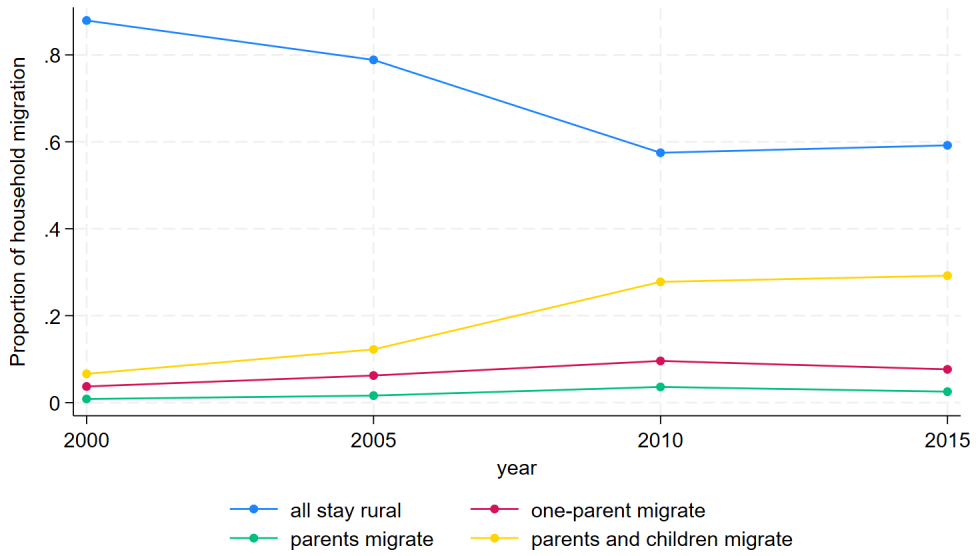


Figure 17: Proportion of household migration

Note: We resort to the Chinese Population Censuses (2000, 2005, 2010 and 2015) to do this exercise.

part of appendix describes the data sources, the definition of city-level controls, and reports their summary statistics.

This study draws on prefecture-level data for empirical analysis. The primary data come from the China City Statistical Yearbooks, covering the period from 2001 to 2007. To complement this, we use data from the 2000 population census to construct several key variables, including over-education, migrant share, fertility rate, and urban unemployment rate. We also use the Chinese Annual Survey of Industrial Firms (CASIF) for the years $2001 - 2007$ to measure the size of the state-owned sector in each prefecture over time.

According to the National Bureau of Statistics, the number of prefecture-level administrative units in China increased from 236 in 1999 to 283 by 2010. To ensure consistency, we restrict our sample to 256 prefectures that did not experience major administrative adjustments between 2000 and 2010. After excluding a few prefectures with missing data, our final sample includes 248 prefectures.

The controls range from local demographics, economic and labour market conditions, to fiscal conditions. They are factors we believe could affect a local government's assessment of the costs and benefits of reducing migration barriers.

The Table below shows the distribution of these variables across prefecture-level cities, and the correlation coefficient between each variable and the city's year of reform.

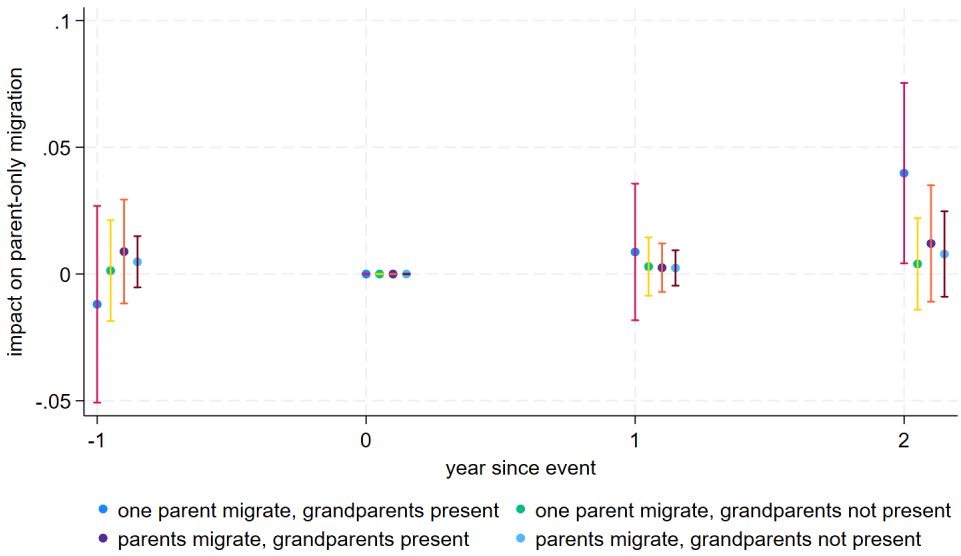


Figure 18: Effect of reform on parent-only migration

Note: We resort to the Chinese Population Censuses (2000, 2005, 2010 and 2015) to do this exercise. We measure grandparent availability using household co-residence information and define it as an indicator equal to one if at least one grandparent or grandparent-in-law is either registered in the household roster or reported as a co-resident member, and zero otherwise. The same event study method is applied as in Figure 5.

Basic demographic conditions

Prefectures with larger urban populations could differ in both the demand for and the capacity to manage migrant inflows, leading to variation in the adoption and intensity of pro-migration policies. We look at the **total urban population** and find it to have a near-zero correlation with the year of reform.

Meanwhile, prefectures differ in the extent of existing rural-to-urban migration. The **share of population who were migrants** was 11.9% at the mean, 3.3% at the 5th percentile, and 27.6% at the 95th percentile. Higher migration rates could make the local government more reluctant to allow even more migrants into the urban centres.

Finally, the prefecture's **fertility rate** serves as a proxy for the expected natural growth of the local labor force. A higher fertility rate implies lower future demand for migrant labor. In our sample, the average fertility rate is 1.21. There is notable heterogeneity across prefectures, with the 5th, 50th, and 95th percentiles at 0.81, 1.19, and 1.74, respectively. The initial fertility rate is positively correlated with the year of reform.

Economic conditions

Economic conditions in the urban areas relative to home villages are likely to affect the migration

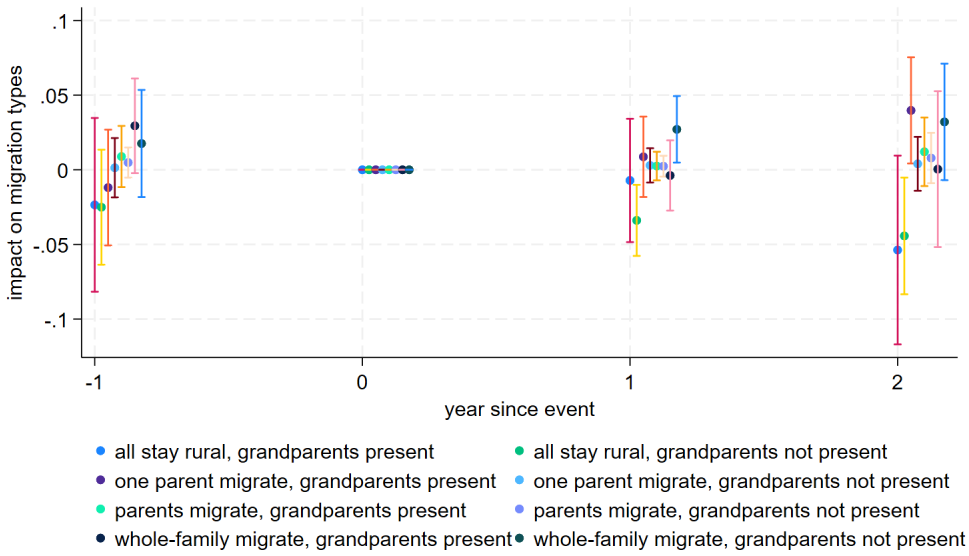


Figure 19: Effect of reform on migration types

Note: We resort to the Chinese Population Censuses (2000, 2005, 2010 and 2015) to do this exercise. We measure grandparent availability using household co-residence information and define it as an indicator equal to one if at least one grandparent or grandparent-in-law is either registered in the household roster or reported as a co-resident member, and zero otherwise. The same event study method is applied as in Figure 5.

decision. Meanwhile, they also affect the local government's willingness to reduce migration barriers, in particular through the perceived capacity of urban areas to absorb more labor.

To capture the level of economic development in each prefecture, we look at **GDP per capita**. On the one hand, prefectures with higher per capita GDP are generally more attractive to migrants and may possess greater fiscal capacity to implement pro-migration policies. On the other hand, the much better employment opportunities and public services in those urban destinations may be so attractive that migrants are drawn to them in the absence of any promising route to settlement. After all, the most developed megacities (Shanghai, Beijing, etc.) have yet to relax their Hukou policy to the same extent of the reform we are investigating (the reform applied to all of mainland China expect megacities in 2014.) Table 27 shows prefectures with higher GDP per capita tend to adopt the reform earlier, with a correlation coefficient of -0.1.

Second, we look at relative labor productivity between urban and rural areas within a prefecture. A larger gap means more potential gain from rural-to-urban migration. Chinese prefectures are composed of both urban 'districts' and rural 'counties'. We calculate labor productivity separately for districts and the entire prefecture. Labor productivity is defined as GDP in the secondary sector (e.g., manufacturing and construction) divided by the number of workers in that

Table 27: Summary Statistics of initial city-level conditions

City-level condition in 2001	mean	standard deviation	correlation coefficient with year of Hukou reform
Rural population share aged below 30	0.478	0.057	0.234
Urban unemployment rate	0.089	0.045	0.216
Share of fiscal expenditure on education	0.300	0.250	0.185
Fertility rate	1.231	0.291	0.169
Output share of state sector	0.371	0.220	0.157
Rural out-migration rate	0.081	0.096	0.135
Rural out-migration rate (ages 20–34)	0.144	0.111	0.113
Migration ratio (migrant/population)	0.105	0.084	0.100
Rural total fertility rate	1.314	0.440	0.075
Ratio of fiscal expenditures to fiscal revenue	1.911	1.419	0.068
Rural employment share in secondary sector	0.115	0.125	0.059
Urban population share aged below 30	0.431	0.042	0.025
Total urban population (10,000)	101.535	92.137	0.004
Rural employment share in tertiary sector	0.089	0.050	-0.018
Labor productivity gap (urban vs rural)	0.365	0.198	-0.075
Logarithim of Per capita GDP (RMB)	8.851	0.592	-0.097
Urban elderly population share (age 65+)	0.069	0.019	-0.100
NTR gap	0.309	0.084	-0.137
Rural elderly population share (65+)	0.074	0.017	-0.213

This table presents summary statistics for prefecture-level variables. All monetary variables have been deflated to account for price changes. The definition and source of each variable is explained in text within this appendix.

sector. According to Table 27, the average relative labor productivity in 2001 was 37%, meaning that, on average, labor productivity in counties was 37% lower than that in districts. I found this to be weakly negatively correlated with the year of reform: prefectures with larger productivity gap adopted the reform earlier.

Third, we look at three measures related to the cities' capacity to absorb rural migrants. The first and most obvious measure is **urban unemployment rate**. Using micro-level census data from 2000, we find that the average urban unemployment rate across prefectures was 8.9%. There is considerable variation in this indicator, with the 5th, 50th, and 95th percentiles at 3.0%, 8.4%, and 16.8%, respectively. When the urban labor market is slack, local governments have often prioritised job security for local residents by restricting the employment of migrants in local firms (Knight and Song, 1999). Indeed, we see that cities with higher urban unemployment rate adopted the reform later: the correlation between unemployment rate and reform year is

Table 28: Predicting Hukou Reform Timing Using Initial City Characteristics

	Year of reform adoption
	b/se
Total urban population (10,000)	0.005
	(0.004)
Logarithim of Per capita GDP (RMB)	-0.557
	(1.043)
Ratio of fiscal expenditures on education to fiscal revenue	8.288***
	(2.157)
Ratio of fiscal expenditures to fiscal revenue	-0.950***
	(0.328)
Rural out-migration rate	6.010
	(5.677)
Rural out-migration rate (ages 20–34)	11.884
	(14.776)
Rural elderly population share (65+)	-34.829
	(32.194)
Rural population share aged below 30	-3.487
	(12.552)
Rural employment share in secondary sector	18.091***
	(6.396)
Rural employment share in tertiary sector	-19.928
	(12.709)
Rural total fertility rate	-0.776
	(1.085)
Urban elderly population share (age 65+)	-56.614**
	(22.220)
Urban population share aged below 30	-24.551**
	(12.072)
Urban unemployment rate	37.607***
	(7.723)
Fertility rate	8.080***
	(2.168)
Proportion of state-owned firms in employment	1.288
	(1.738)
Labor productivity gap (urban vs rural)	-0.215
	(1.789)
Migration ratio (migrant/population)	-11.389
	(18.693)
Observations	194
R-squared	0.299

Notes: The dependent variable is the year of migration reform adoption at the prefecture-city level. The explanatory variables are predetermined city characteristics measured before the reform. A higher value indicates later adoption of migration reform. Robust standard errors are reported in parentheses.

0.22 (Table 27).

We also construct a measure of exposure to trade shock as a second way to capture variation in labour demand. The U.S. granting of Permanent Normal Trade Relations (PNTR) to China was a natural policy shock. Since the 1980*s*, Chinese exports to the U.S. had benefited from the relatively low Normal Trade Relations (NTR) tariff rates granted to WTO members. However, unlike other countries, China's access to these low tariffs required annual renewal by the U.S. Congress, an uncertain and politically sensitive process. Without renewal, Chinese exports would have been subjected to the much higher non-NTR rates stipulated by the Smoot-Hawley Tariff Act of 1930. Following the methodology of Pierce and Schott (2016) and Facchini *et al.* (2019), we use the NTR-gap to quantify the degree of trade policy uncertainty prior to PNTR. The **NTR-gap** is defined as the difference between the relatively low NTR tariff rate (averaging 4% across industries in 1999) and the much higher non-NTR rate that would have applied in the absence of renewal (averaging 36%). The average NTR-gap across prefectures in our sample is 31%, and there was substantial variation in local exposure to trade policy uncertainty. However, we do not find this trade shock to be correlated with the timing of reform.

A third measure related to labour demand is size of the State sector. The Chinese economy was transitioning from a planned economy dominated by large state-owned enterprises to a more market-oriented system. This transition involved mass layoffs from SOEs and rapid economic growth led by private enterprises. Thus, a larger state sector in 2001 may be perceived as a risk factor for the labour market in the near future and cause local governments to be more cautious about loosening migration control. We obtain **the output share of state-owned enterprises (SOEs)** based on data from the Chinese Annual Survey of Industrial Firms (CASIF). Table 27 shows this variable varies considerably across cities and it's positively correlated with the year of Hukou reform (correlation coefficient =0.16).

Fiscal conditions

In China, local governments are primarily responsible for providing public services such as education and healthcare. Among these, access to public education is a key benefit reserved for local residents under the Hukou system, often inaccessible to migrant workers (Liu, 2005; Chen and Feng, 2013).

The fiscal balance of local governments may affect their decision to relax migration. On the one hand, migrants can contribute to fiscal revenues because they tend to be young and working. On the other hand, making it easier for migrants to settle and access local public services will require more public spending. We found that **the ratio of fiscal expenditure to fiscal revenue** does not correlate strongly with the prefecture's year of reform. The ratio has a

mean of 1.9, suggesting that fiscal expenditures frequently exceed revenues - a common pattern among Chinese prefectures.

Finally, we look at the **share of fiscal expenditure on education**. Access to better education in urban areas is a key benefit offered by the reform, and also has the potential to encourage fertility among prospective migrants. On average, prefectures allocated 30% of their spending to education, and there is substantial variation. These disparities mirror the uneven distribution of educational quality and public service provision across China (Qian and Smyth, 2008). Table 27 prefectures that spent more in education adopted the reform later, with a correlation coefficient of 0.18. This could be because for those prefectures, the fiscal cost of widening access to public services would be higher.

Rural/urban population characteristics

We construct a set of additional controls based on the 2000 Population Census, focusing primarily on characteristics of the rural population. These include: (i) the overall out-migration rate, (ii) the out-migration rate among individuals aged 20–34 (the primary reproductive age group), (iii) the share of the population aged 65 and above, (iv) the share of the population aged below 30, (v) the shares of employment in the secondary and tertiary sectors, and (vi) the total fertility rate among rural residents.

To further capture demographic conditions at the city level, we also include measures of urban population structure, specifically the share of the urban population aged below 30 and the share aged 65 and above. These variables allow us to account for broader demographic pressures that may influence both migration patterns and local policy incentives.

Together, these variables allow us to account for potential confounding factors related to selective migration, population aging, and structural transformation, emphasized in the migration and local policy adoption literature (Casarico *et al.*, 2018; Jin, 2024; Tian, 2024; Chen *et al.*, 2025), which may jointly influence both the timing of Hukou reform and fertility outcomes.

Importantly, rural population measures are constructed at the rural level, thereby aligning the level of aggregation with our outcome variable—fertility behavior among rural households. This helps address concerns regarding the mismatch between city-level controls and rural-level outcomes, while the inclusion of urban demographic characteristics captures complementary variation in local population structure.

In summary, Table 27 shows that most city-level initial conditions are only weakly correlated with the year of reform. To mitigate the concern of non-random reform timing, we include the full set of these pre-reform characteristics interacted with year fixed effects, thereby allowing for differential trends across cities with different initial conditions and the results were shown in

Table 5. It did not much noticeable differences to the estimated effects.

9.7 Formal Tests of Differences in Treatment Effects Across Household Subgroups

The heterogeneity analyses in Tables 13 and 14 compare estimated treatment effects across households with different initial characteristics. To formally assess whether these subgroup differences are statistically significant, we conduct bootstrap tests of equality between subgroup treatment effects.

Specifically, for each characteristic, we estimate the difference between the average treatment effect for households with the characteristic and the corresponding effect for households without the characteristic:

$$\Delta ATT = ATT_{with\ characteristic} - ATT_{without\ characteristic}.$$

Standard errors are obtained from 200 bootstrap replications clustered at the prefecture-city level. A positive value indicates that the estimated treatment effect is larger among households with the specified characteristic.

Appendix Table 29 reports the results. The differences associated with potential parents' migration histories are positive and economically sizable, particularly for households with fathers who had prior migration experience, although the formal tests do not reject equality at conventional significance levels. For potential carers, the difference between households with and without multiple generations of adults is statistically significant, while the differences based on the presence of healthy older people or women are not statistically distinguishable from zero.

Table 29: Tests of Differences in Treatment Effects Across Household Characteristics

Subgroup comparison	Difference in ATT	Bootstrap SE	p-value
Panel A: Potential parents' characteristics			
Father: works non-locally for 180+ days	0.013	0.010	0.208
Father: works non-locally for 30+ days	0.013	0.009	0.129
Mother: works non-locally for 180+ days	0.003	0.010	0.790
Mother: works non-locally for 30+ days	-0.002	0.010	0.833
Panel B: Potential carers' characteristics			
Healthy people aged ≥ 47	0.005	0.005	0.310
Healthy women aged ≥ 47	0.004	0.007	0.544
Multiple generations of adults	0.011^{**}	0.006	0.047

Notes: This table reports formal tests of differences in treatment effects across household subgroups. Each row reports the difference between the estimated ATT for households with a given characteristic and the corresponding ATT for households without that characteristic. The reported difference is calculated as $ATT_{with\ characteristic} - ATT_{without\ characteristic}$. Standard errors are obtained from 200 bootstrap replications clustered at the prefecture-city level. A positive coefficient indicates that the treatment effect is larger for households with the specified characteristic. Panel A reports tests based on characteristics of potential parents observed in 2003, while Panel B reports tests based on the availability of potential carers.