Migration Reform and Fertility: Causal Evidence from Rural China*

Wenchao Jin, Zhangfeng Jin

May 6, 2025

Abstract

This study explores the long-term effects of removing migration barriers on fertility. We analyse a major reform that significantly reduced institutional barriers faced by rural-to-urban migrants. As the reform was staggered across over 200 cities in China, we used an event study approach to estimate its impact on surrounding rural areas. We find that removing these barriers led to a significant increase in fertility among rural households over the next decade. The average treatment effect on the treated is 0.011 newborns per household per year, which is about one third of the sample mean. The effect on fertility is particularly strong in extended families and in households where prospective parents were already temporary migrant workers.

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 declare no relevant or material financial interests related to the research described in this article.

1 Introduction

Historically, industrialisation and sustained economic growth have been closely intertwined with declining fertility and increasing education (Galor, 2011). But some countries including China are 'becoming old before getting rich' (Johnston, 2019), raising urgent policy questions about fertility dynamics in developing and emerging economies. In such settings, internal migration is widespread and may shape fertility decisions through both economic and social channels. Yet, despite its importance, the role of migration policy in influencing fertility remains understudied. This paper provides new evidence from China, showing that policies which reduce internal migration barriers can have large, unintended effects on fertility among rural households.

We study a major reform reducing institutional barriers to migration in China and found a large, significantly positive impact on fertility among *rural* households. While existing literature has explored how migration barriers influence fertility among migrants in urban destinations, to the best of our knowledge, this is the first paper to examine their effects on fertility in *sending rural regions*. This fills a significant gap in the literature, because in many contexts (including rural China) the majority of rural-born population remain in rural areas. According to the National Bureau of Statistics of China, 960 million people had Agricultural Hukou in 2010 and 373 million had non-agricultural Hukou[1]; while 674 million lived in rural areas and 666 million people lived in urban areas.[2]

China provides an ideal setting to study this question due to its long-standing Hukou System, which has tightly regulated rural-urban migration since the late 1950*s*. While the early 1980*s* saw initial reforms allowing rural residents to work temporarily in urban areas, these migrants were denied equal access to local public services unless they obtained a local Hukou. Access to local Hukou was strictly limited by a rigid annual quota system until the early 2000*s*, when reforms began replacing quotas with a more flexible system. These reforms removed strict caps on migrant numbers and eliminated restrictions based on age and education, significantly increasing migrants' likelihood of obtaining local Hukou and reducing uncertainty for prospective migrants. The staggered implementation of these reforms across 283 Chinese cities over 2002–2015 creates valuable variation for studying their causal effects.[3]

[1]When registering their birth, Chinese children are assigned a Hukou. It determines their rights and access to public services and social benefits. See institutional background in subsection 2.1.

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

[3]In China, a prefecture-level city is an administrative division that includes both urban and rural areas. Under China's administrative hierarchy, each province typically governs several prefecture-level cities. These cities, in turn, oversee a number of districts and counties, which further administer various subdistricts (or streets) and villages. Within a prefecture-level city, districts are generally regarded as urban centres, while counties typically encompass suburban and rural areas, including most towns and villages. In this paper, we define urban areas to include both

This paper evaluates the impact of Hukou reform using an event study approach in a large panel of rural households. To ensure the robustness of our method to heterogeneous effects and negative weights, we apply the estimator proposed by de Chaisemartin and D'Haultfoeuille (2020). We find that removing migration barriers led to a substantial increase in fertility among rural residents over the subsequent decade. Specifically, the reform increased the likelihood of rural households having a newborn by 1.1 percentage point per year, which is about one third of the sample mean. This effect persists and is broadly increasing over time. We also found the effect to be more pronounced in extended families (with at least 2 generations of adults) and where prospective parents had prior experience of migrant work.

A theoretical model is set up to interpret these findings and clarify a multitude of competing mechanisms through which migration reform can affect fertility. They including income and substitution effects, child-rearing costs, preference changes, and returns to educational investment. Finding of a positive impact on household consumption supports the view of a large positive income effect (from improved access to social benefits such as urban schools). The availability of grandparents for childcare provision limits the substitution effect.

This paper's main contribution is to provide the first causal evidence on how increased migration opportunities affect fertility decisions of households who remain in rural areas. Many studies have explored the effects of pro-migration policies on migrants' fertility behaviour, with mixed findings. Some studies suggest that pro-migration policies increase fertility (while migration barriers decrease it) among migrants (Amuedo-Dorantes *et al.*, 2023a; Lanari *et al.*, 2020; Dong *et al.*, 2023; Amuedo-Dorantes and Arenas-Arroyo, 2021), while others show the opposite effect (Amuedo-Dorantes *et al.*, 2023b; Avitabile *et al.*, 2014; Jensen and Ahlburg, 2004). However, the literature is silent on how migration opportunities affect fertility among households who remain in rural areas. In our sample of rural households, about three quarters of households observed in 2003 are still observed in their rural village in 2015[4]. Thus, understanding the impact on the remaining rural population is crucial for gaining a more comprehensive picture of how migration policies influence aggregate fertility.

This paper relates to the broader literature of how policies affect fertility (Bergsvik *et al.*, 2021; Hart *et al.*, 2024). We show that a non-fertility-targeted policy can increase fertility through channels such as higher incomes or the quantity-quality trade-off. Our finding of a much stronger impact when grandparents were present suggests informal childcare plays a critical role in mediating fertility responses to economic policy. This has important implications for other developing countries where multigenerational households are common and formal childcare remains lim-

urban centres and suburban areas. Unless otherwise specified, the term city refers to a prefecture-level city.

[4]The attrition is not necessarily due to permanent migration to urban areas. Section 6.2 will examine what factors are correlated with sample attrition and check robustness of our main finding in a balanced sample.

ited.

This paper also adds to the literature on Hukou reforms in China by evaluating its impact on rural households, complementing a number of studies that have investigated its impact in urban destinations such as migration inflows (Jin and Zhang, 2023), migrants' wages (An *et al.*, 2024), urban natives' employment or income (Zhang *et al.*, 2024), entrepreneurial activities (Liu *et al.*, 2024) and migrants' fertility in urban destinations (Dong *et al.*, 2023).

The rest of this article is organised as follows. Section 2 explains the Hukou System and the reform under investigation, followed by the empirical and institutional background of declining fertility. Section 3 sets up a stylised model of fertility to illustrate five mechanisms through which the reform could affect fertility. Section 4 describes data sources and summary statistics for key variables in the chosen sample. Section 5 details the model specification and challenges to identification. Section 6 presents our empirical findings. The final section concludes the study.

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 and villages) 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*.[5] We compiled data on the adoption dates of these reforms across 283 prefecture-level cities[6] 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 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 1 illustrates the timing of migration reform adoption across cities. 30% cities adopted the policy by 2004. The share increased gradually to 61% by 2014, and in 2015 the policy was rolled out nationally.

Cities' decision about when and how to implement the Hukou reform may be affected by the local economic and social conditions. In subsection 4.2, we will examine a wide range of economic and demographic conditions in 2001 and see which are correlated with cities' timing of policy adoption. We found later adoption is associated with four factors and we will address the endogeneity issue in two ways in the analysis.

[5] 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.

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

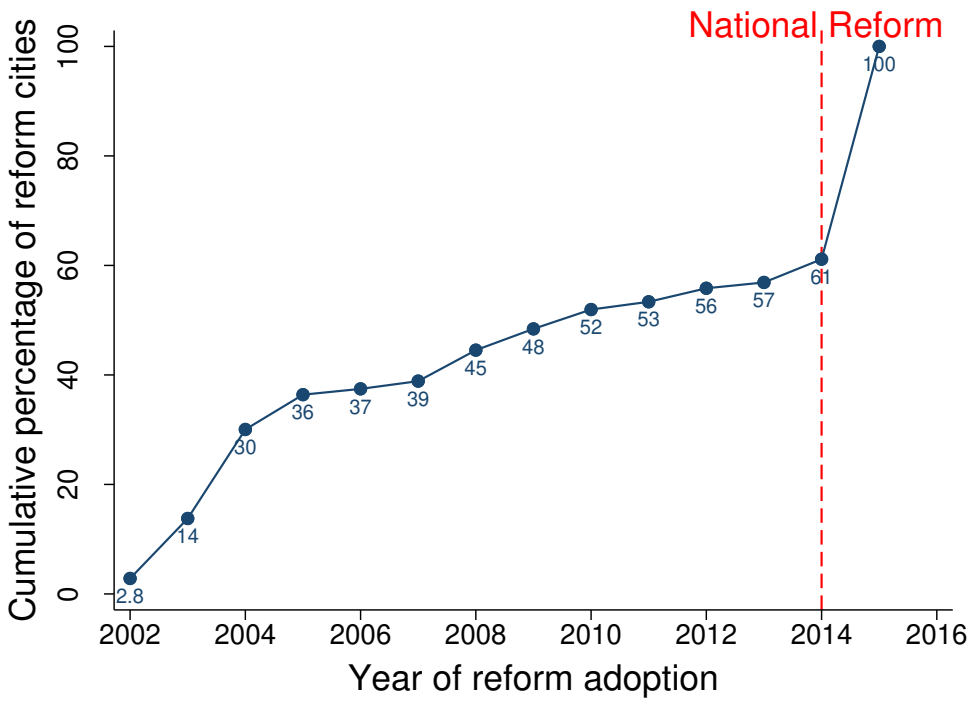


Figure 1: The Adoption of Migration Reform Over Time

Notes: This graph shows the adoption of migration reform over time.

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 persists and actually increases over the first 6 years (the end of the window examined in Jin and Zhang (2023)). Meanwhile, the reform has significantly increased the agricultural TFP of rural households, increased the level of agricultural mechanisation and facilitated the subcontracting of land by low productivity farmers (Ji and Nie, 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.

Fertility rate has been substantially higher in rural China than urban China since the mid

1960*s*. In rural China, fertility fell dramatically from above six births per woman in the 1970 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 if they were already in poverty.

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-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).[7] 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 over time. The decline might be partly a result of Hukou reforms since the early 2000*s*. 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. If prospective parents from rural areas are forward-looking, they may be encouraged by this policy to have more children, or have children sooner. Next section will illustrate this intuition in a formal model, alongside other factors relevant for fertility choice.

[7] 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.

3 Conceptual framework

This section develops a stylised model of fertility incorporating investments on children and childcare requirement. We discuss how the Hukou reform was likely to affect various exogenous parameters in the model and thereby deduce how fertility is likely to respond. The objective is to clarify the competing mechanisms through which the reform could affect fertility, and what evidence could be used to support or reject different mechanisms.

In total, the model captures 5 channels: income effect, substitution effect, changes to the costs of raising children regardless of child quality, changes to the benefits of having more and higher-quality children, and changes to the return to investment in child quality. The model does not consider child labor or child mortality, because both were already rare in China in the sample period.

Consider a household that generates utility from number of children n, average quality of children q, and consumption c. They maximise utility subject to a set of constraints about budget and time. We will make functional-form assumptions in order to explicitly derive the formula for fertility choice.

The utility function is log-linear in those three components, following Galor (2012).

$$U(c, n, q) = (1 - \gamma) \log c + \gamma(\log n + \beta \log q) \tag{1}$$

where $0 < \gamma < 1$ and $0 < \beta <= 1$ are preference parameters. It's unclear whether the Hukou reform should affect preference parameters β, γ. If these preferences are 'warm glows', then the reform is unlikely to change them. If preferences are shaped by one's social surroundings, then moving to urban areas may increase β and reduce γ. This is because compared to rural areas, urban areas have a social norm of having only one child (partly due to decades of the One Child Policy) and of investing heavily in children's education.

On the other hand, if parental preferences are based on expected support from adult children during parents old age[8], then expected future incomes of children matter. If the reform increases expected future incomes of all children (independent of their quality), and thereby expected old-age support from adult children, then γ should go up. At the same time, the reform may also improve parents' access to more generous pension schemes, thus making old-age support

[8]Old-age support was an important motivation for child-rearing in rural China, as captured by the old Chinese saying "raise children to protect against old age". China's laws stipulate that adult children are responsible for supporting their parents. According to China's 2015 Population Census Survey (National Bureau of Statistics of China, 2018), among individuals aged 60 and above, 49% and 7% of rural residents cited "family members" and "pensions", respectively, as their primary sources of old-age support, compared to 15% and 75% in urban areas. The difference is partly due to much higher provision of pensions for urban residents.

from adult children less important. One can argue this should reduce γ. If the reform increased the gradient of future child's income with regard to child quality, then β should go up. We will interpret child quality as human capital. It is debatable whether the reform increased the gradient of future income to child quality because of two competing effects: the urban labor market has a higher return to human capital; child's ability to settle in a city during adulthood used to strongly depend on their human capital and would be less so after the reform.

Quality of children depends on the social environment and educational investments per child.

$$\log q = \delta_s + \delta_e \log \frac{E}{n} \tag{2}$$

where 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 educational investment. The model assumes away the possibility of parental time investments to improve child quality.[9] Independence between parental time inputs and child quality is not crazy in our context for two reasons: children's education tend to far exceed their rural parents; younger children were often left behind while parents work non-locally.

The Hukou reform improved access to urban schools, which were better than rural schools and both were largely free. The reform also improved access to 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.

The impact of the reform on δ_e is ambiguous. There are multiple types of education investments available or unavailable to parents, based on their Hukou status. For rural migrant parents without urban Hukou, they could send their children to rural schools for virtually free, or pay a high fee to send them to urban schools designated for rural migrants' children, which could be higher quality than the rural ones. With urban Hukou, their children would be able to attend urban schools designated for urban natives, which tend to be significantly higher quality than the ones designated to rural migrants' children. While there was no formal tuition fee, the quality of urban school had a big impact on local house prices and so house purchase near popular schools could be interpreted as a kind of educational investment. Moreover, private tuition classes were much more available in urban areas than in rural ones, which is another way to invest in children. Finally, tertiary education was subsidised by the state but still a large expense for lower-income parents. Because these investments have vastly different monetary costs and their effects on child quality are hard to ascertain, we cannot ascertain the impact of the reform on δ_e.[10] It is certain

[9] This assumption is also made in the model about quantity-quality tradeoff in Doepke *et al.* (2023).

[10] If one believes that the schools for migrants' children are much better than rural schools and the value added of private tuition and school catchment is low relative to their costs, then δ_e goes down. If the opposite is true, then δ_e

that δ_s improves because a child in an urban school and receiving zero top-up investment would be getting a better quality of education than a child in a rural school.

The household contains two potential parents: a man and a woman, who have market wages w_m for men and w_w for women. 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.

$$h_m = 1 \tag{3}$$

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

Raising each child takes a fixed amount of time τ. This can be done by either the mother or grandparents.

$$t_w = n\tau(1 - G) \tag{5}$$

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. [11] Another common assumption in household models of home production (encompassing childcare) is that man's time and women's time are perfectly substitutable. This means the household will choose to specialise: the higher-wage parent would spend zero time on childcare. As the husband's wage is usually higher than the wife's[12], we would still get the result that the man does not provide childcare.

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. This reflects the reality of many rural children who were left-behind by their migrant parents and taken care of by their rural grandparents.[13] If a grandparent is not present $(G = 0)$ or only partially present $(0 < G < 1)$ due to poor health or physical distance, then the parents will need to spend some time on childcare. The reform has been shown to increase migration

goes up.

[11] 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.

[12] In our sample, the potential father's wage is higher than the potential mother's in 69% of households, conditional on both being observed. Unfortunately we only observe wages from migration, and not wage from local work. This means for all potential mothers [fathers], only a third [half] has wage observations

[13] Fan *et al.* (2011) found that the main factors determining the decision of leaving children behind are child's age and whether grandparents are available to help.

of younger people more than older ones.[14] For the years before schooling, if children are left behind with rural grandparents, conceptually G in the model would still be 1. But if children live with parents in urban areas (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. Thus, the reform has the potential to reduce G through migration. G is not a choice in the model, because in our sample the older generation are much low-educated than the younger generation, so it's almost always optimal for the grandparents to provide childcare instead of the mother/father, as long as the grandparents are present and healthy enough.

Finally, there is a budget constraint:

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

where p^n is the cost of raising children (e.g. housing and food) and is independent of child's quality, and y_0 is unearned income. The Hukou reform increased the expected future y_0, because urban Hukou holders could enjoy more generous pensions and health insurance. p^n is also higher because housing cost and general living costs are higher in urban areas. Wages of men and women w_m, w_w may increase as a consequence of the reform, due to access to better jobs in urban areas or lower supply of workers remaining in rural areas.

To summarise, the household chooses $n, q, c, h_w, \tau_m, \tau_w, E$ to maximise their utility 1, subject to the child quality production function (2), woman's time constraint 4, a childcare time constraint 5, and a budget constraint (6). All the choice variables have to be positive or non-negative.

We can work out the optimal choices of n, c, q. The derivation process is in Appendix 8.1. Here it's worth to first describe the costs and benefits of an additional child.

The First-Order-Condition with regard to the number of children (n) is:

$$\frac{\gamma}{n} - \frac{\lambda_q \delta_e}{n} - \lambda_\tau \tau (1 - G) - \lambda_c p^n = 0 \tag{7}$$

and it simplifies to

$$\frac{\gamma}{n} - \frac{\lambda_q \delta_e}{n} = \lambda_c [w_w \tau (1 - G) + p^n] \tag{8}$$

where $\lambda_c, \lambda_q, \lambda_\tau$ are the Lagrangian multipliers for the budget constraint, the log quality production function, and the childcare time constraint. (18) shows that while an additional child increases utility directly, it also has three costs: 1) by spreading investments more thinly, it tends to reduce average quality of children; 2) there is a time cost through childcare time constraint; 3)

[14] Jin and Zhang (2023) found that the reform significantly increased rural-to-urban migration among individuals aged 30 to 44, but had no significant effect on those aged 60 and above half by 2010.

there is a direct cost through p^n as the additional child requires food and housing. The time cost depends on female wage and grandparents availability.

Denote

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

p^* can be considered the shadow price of children incorporating the time cost of childcare. The latter depends on the female wage and whether a grandparent is available to help with childcare.

Denote full income $\bar{y} = w_m + w_w + y_0$.

We can express the optimal choice of c, n, E explicitly as functions of exogenous variables. The derivation process is in Appendix 8.1.

$$c = (1 - \gamma)\bar{y} \tag{9}$$

$$n = \frac{\bar{y}(\gamma - \gamma\beta\delta_e)}{p^*} \tag{10}$$

$$E = \gamma\beta\delta_e\bar{y} \tag{11}$$

(9) shows that consumption will be a fixed proportion of full income, which is unsurprising given the log-linear specification of utility function. (11) implies educational expenditure E will also be a fixed share of full income $\bar{y}$, and the share depends positively on preference for child quality β and the return to investment in child quality δ_e. (10) shows that fertility n times its shadow price p^* will be a proportion of full income $\bar{y}$, and the proportion is $(\gamma - \gamma\beta\delta_e)$. This proportion is increasing in γ and decreasing in β, δ_e. This proportion captures the trade-offs parents face: β captures the motivation for trading-off child quantity vs child quality; γ captures the tradeoff between consumption and children. In summary, the fertility choice will be increasing in $\gamma, \bar{y}$ and decreasing in β, δ_e, p^*.

Now we can analyse how the Hukou reform should affect the fertility choice n. Table 1 lists how each exogenous variable in the model is likely to be affected by the reform, and thereby what's the impact on fertility n.

First, there is a positive income effect, from both the increase in y_0 and w_m. The effect on y_0 is hard to capture empirically because y_0 encompasses social benefits such as health insurance and pensions which materialise in the distant future. Chen *et al.* (2019) estimated that the willingness to pay for an urban Hukou in JiNan was approximately 90,000 to 126,000 yuan in 2017, which was 5.4 to 7.6 times the city's rural per capita annual income. In other words, urban Hukou was highly valuable to rural migrants. One testable prediction of the positive income effect is that the reform should increase household consumption and education expenditure at the same

Table 1: Mechanisms through which the reform impacts fertility

exog. var. affected	why it's affected	mechanism	impact on fertility
y_0 increases	access to better social security	income effect via $\bar{y}$	$+$
w_m increases	access to better jobs	income effect via $\bar{y}$	$+$
w_w increases	access to better jobs	income effect via $\bar{y}$	$+$
		substitution effect via p^*	$-$ if $G < 1$
p^n increases	higher housing and food costs	increase p^*	$-$
G falls	grandparents left behind	increase p^*	$-$
γ may increase	children will be richer	preference for children $+$	$+$
γ may fall	assimilate to urban culture	preference for children $-$	$-$
	pensions reduce reliance on children	preference for children $-$	$-$
β may increase	q less critical for child's income	quantity-quality tradeoff	$-$
	urban culture prizes child quality	quantity-quality tradeoff	$-$
β may fall	q less critical for child's migration	quantity-quality tradeoff	$+$
δ_e ambiguous	returns to educational investment	quantity-quality tradeoff	$+/-$
δ_s increases	better schools for free	impact q only, not n	0

time. But education is also affected by the perceived returns to educational investments (the fifth channel), so the total impact on education is ambiguous. Section 6 will check how the reform affected household consumption as well as educational expenditure.

Higher male wage will not have a substitution effect because men do not contribute to child-care in this model. By contrast, higher female wage w_w has both an income effect and a substitution effect. The model's prediction that male wage has a more positive impact on fertility than female wage echoes the historical finding that the decline in gender wage gap during the process of industrialisation contributed to the decline in fertility (Galor, 2011; Galor and Weil, 1996). Empirically, section 6 will show that the reform has positive effects on both male and female wages in rural households. While they are not always significant, those estimated effects are likely to under-estimate the average impact on wages because we cannot observe individuals' formal wages in the longer run (after they permanently settle in urban areas).[15]

The second channel (a negative substitution effect) depends on grandparents' availability (G). Suppose G is fixed and varies across households for exogenous reasons, then households with higher G will have a smaller shadow price p^* and have more children, holding other exogenous variables equal. The same proportional increase in full income (due to y_0 or w_m) will have the same proportional impact on fertility, but the absolute magnitude of the effect would be larger in household with higher G. If the reform also increases w_w, then the impact on fertility will be more positive in households with higher G due to a weaker substitution effect of female

[15]The sample is collected from rural areas, and we only observe wages of migrants who were still considered part of the rural household by left-behind family members.

wage. These predictions are supported in section 6.3, where we find households with multiple generations of adults or healthy grandparents have a higher fertility rate on average and also a more positive impact from the reform.

Third, the reform may increase the direct cost of raising each child p^n (e.g. food and housing) and make grandparents less available (G falls) through physical separation. Both of these will increase p^* and reduce fertility. It can be somewhat mitigated by leaving children behind in rural areas before school age. Note that this channel works through the physical process of migration and separation. For households who were going to migrate regardless of the Hukou policy, they were already planning to endure the higher cost of raising children and lack of informal childcare in urban destinations. For them, the reform has the afore-mentioned income and substitution effects, but not this third effect. As this third effect is negative, the model predicts that for households who were going to migrate regardless of the reform, the impact of the reform on fertility should be more positive than for other households. This prediction will be verified in subsection 6.3.

Fourth, the reform may change parental preferences for the number and quality of children. The direction of change depends on what determines preferences. For example, if preferences for more and higher-quality children come from the desire to receive more old-age support, then the reform could increase the demand for children γ because in expectation children will be richer. But it's still ambiguous how β should be affected, because the reform has ambiguous effect on the future returns of child's income to child's human capital. We do not have data to tell how preferences regarding children have changed.

Fifth, it's unclear how the reform affected the returns to parental investment in child quality. If it improves access to more effective ways of investing in children's quality (δ_e increases), then it will reduce fertility through the quality-quantity tradeoff. If urban settlement removes the need for the pre-reform type of investment (e.g. paying fees to urban schools designated for migrants) and other types of investments are still too expensive to be relevant to newcomers in urban areas, then δ_e will fall and the impact on fertility will be positive. Later sections will examine educational expenditure of rural households, but we will not see the whole picture as we cannot observe education expenditure after permanent settlement in urban areas, nor do we know how effective different investments were in improving quality of children. Hence there isn't enough evidence to suggest which way this channel will go.

Finally, the quality of children in the absence of parents investments δ_s has zero impact on fertility in this model. This is because the utility function is log-linear: the marginal utility of an additional child does not depend on their quality. If parents' preferences for children are purely based on old-age support and higher-income children contribute more, then the marginal utility

of a child should depend positively on their quality. Then, intuitively, the reform should increase fertility through improving expectation of children's prospects. This potential mechanism is outside the model presented here.

To summarise, the theoretical model presented here incorporates several common determinants of fertility. The Hukou reform has the potential to simultaneously affect many exogenous variables in the model. This means five competing mechanisms may be at work at the same time, and the total effect of the reform on fertility is in theory ambiguous.

The model does have three other predictions that we can verify in the rural survey. First, the reform should increase household consumption.[16] Second, the impact on fertility should be more positive if grandparents were available for childcare. Third, the impact should be more positive if the household was going to migrate regardless of the reform. These predictions will be confirmed in subgroup analysis in sub-sections 6.3 and 6.4.

4 Data and Summary Statistics

4.1 National Fixed Point Survey (NFPS)

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 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 survey within NFPS collects demographic data and details of each household member's 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 labor and human capital 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

[16] We think it's very unlikely for γ to increase so much as to offset the increase in $\bar{y}$. This is because there are reasons for the reform to reduce γ.

household-level information only, and not individual-level information such as age.

Our data contains 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 their 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 adult children leaving rural home to settle permanently elsewhere, would be omitted. In our sample, about half of households had at least one person living at home for under 180 days.

Our analysis focuses on household-year observations from 2003-2015. Table 2 presents the distribution of key variables across 202,806 observations at household-year level. 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. So 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 2). This reflects the prevalence of multigenerational living arrangements, which we will explore shortly in Table 4.

The lower panel of Table 2 shows summary statistics for education and labour supply averaged across all adults within the same household-year. Table 3 shows summary statistics for the same variables in the individual distribution 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.[17] Table 3 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. The lower half of Table 2 shows that the mean number of days worked per year per adult is 175. Among that, 67 days were spent on local agricultural work, 67 days on non-local work, and 42 days on local non-agricultural work. Table 3 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 of work differs a lot between young and old and by

[17] 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 2: Summary statistics at household-year level

variable	observations	mean	median	min	max
has newborn	202806	.0274055	0	0	1
number of adults	202806	3.182066	3	0	11
number of children	202806	.6697237	0	0	9
whether head of household is male	201404	.9339487	1	0	1
age of head	201223	53.61613	54	18	97
years of education of the head	195874	6.809271	7	0	22
any person not living at home	202806	.4954587	0	0	1
mean outcome among adults in the household					
male	202780	.5167009	.5	0	1
age	202636	45.86307	43.5	18	96
years of education	200229	7.049918	7	0	22
total number of days worked	202806	175.497	176.6667	0	730
days worked in local agriculture	202803	67.07424	43.33333	0	365
days worked in local non-agricul	202804	41.63838	0	0	365
days worked non locally	202794	66.78977	30	0	365
number of days lived at home	199364	294.569	341.25	0	365

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 away from home was 180+.

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 3 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 (Table 2).

Notably, household structure is closely tied to fertility patterns. About two-thirds of households contain multiple generations of adults, and these households have significantly higher newborn rates (3.7% annually) compared to single-generation households (0.6%). Table 4 also breaks the sample by more detailed household type. The most common type of household

Table 3: Summary statistics for adults at individual-year level

	male age under 45			female age under 45		
variable	median	mean	std dev	median	mean	std dev
male=1;female=0	1.00	1.00	0.00	0.00	0.00	0.00
age	30.00	30.67	8.13	30.00	31.00	8.31
years of education	9.00	8.85	2.65	8.00	8.27	2.84
number of days lived at home	300.00	221.95	147.12	360.00	258.61	142.06
total number of days worked	270.00	218.28	126.22	200.00	178.84	134.87
days worked in local agriculture	0.00	43.72	81.71	0.00	55.42	92.08
days worked in local non-agricul	0.00	45.63	99.07	0.00	37.05	90.04
days worked non locally	60.00	134.38	143.11	0.00	90.42	135.17
health status:excellent=1;good=2	1.00	1.29	0.57	1.00	1.34	0.58
	male age over 45			female age over 45		
male=1;female=0	1.00	1.00	0.00	0.00	0.00	0.00
age	58.00	58.94	9.09	57.00	58.64	9.50
years of education	6.00	6.52	2.67	5.00	4.92	2.81
number of days lived at home	365.00	329.87	87.09	365.00	345.05	67.52
total number of days worked	200.00	175.24	128.45	70.00	116.93	123.71
days worked in local agriculture	40.00	83.77	100.87	30.00	75.50	98.69
days worked in local non-agricul	0.00	52.15	99.09	0.00	28.41	77.13
days worked non locally	0.00	43.08	96.69	0.00	14.78	62.37
health status:excellent=1;good=2	2.00	1.97	1.07	2.00	2.15	1.12

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

is parents 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. 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 means prospective grandparents could be an important source of childcare provision. Later in subsection 6.3 we will examine the impact of the reform by the presence of healthy prospective grandparents. The survey asks respondents to categorise health status into 5 categories, indexed 1 to 5 from the best to the worst. As shown in Table 3, among both genders over 45, the median health status is 'good', the 2nd category; and the mean is close to 2.

In recent years, the share of households having newborns has declined in our sample. Figure 2 shows that it used to be around 3% in the 2000s, and has started declining around 2010, to

Table 4: Summary statistics by household type

household type	observations	mean of num. of adults	num. of kids	hasnewborn	age of head
by whether adults are in the same or different generations					
single generation of adults	65739	1.893746	.5521684	.0064497	53.89773
multiple generation of adults	137067	3.799959	.7261047	.0374561	53.48241
by detailed household type					
single or couple only	41910	1.854903	.004796	0	61.70899
parents+ children<18	22546	1.988024	1.547281	.0173423	39.33398
parents+adult children	68070	3.423667	.2365212	.000808	51.06626
three generations	66400	4.098645	1.204639	.0701506	56.04056
4 or 5 generations	3253	5.165078	1.264064	.1223486	53.45042
other	627	4.320574	.8580542	.0893142	52.82353

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. '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.

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 outmigration of young people. Figure 3 shows that the decline in the share of households having newborns after 2010 coincides with decline in the presence of woman under 40. The right sub-figure shows that was due to ageing of the rural population rather than gender-biased migration.

The NFPS tracks households over time. For most of our analysis, we do not restrict the sample to be balanced. Table 5 shows the attrition rate over time. Year on year, between 1.5% and 16% households drop out of the sample. Overall the annual attrition rate of 4.5%. Some households come back to survey in later years. Figures 4 shows what proportion of households observed in 2003 have remained in the sample over time. By the end of the sample of period, about three quarters remain. Moreover, this rate of attrition did not differ much by initial migration status.

4.2 Timing of reform

Table 6 shows pairwise correlation between initial economic conditions and the year of reform at the city level. The strongest correlation is seen between urban unemployment rate and reform year, with a correlation coefficient of 0.22. It's unsurprising that cities with higher urban unemployment rate adopted the reform later. Higher urban unemployment would make the city

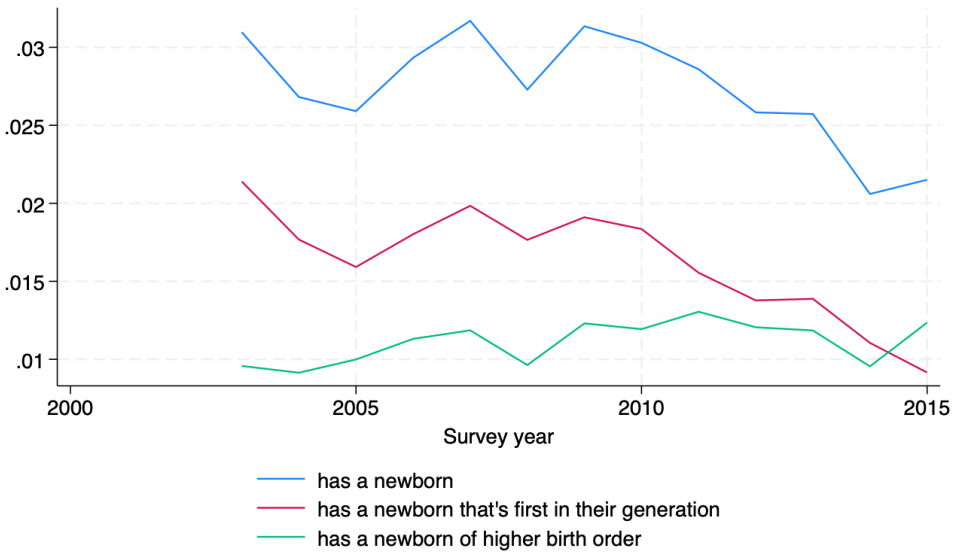


Figure 2: Share of households having a newborn over time

Notes: the sample is all households in cleaned data as described in Table 2.

officials more cautious about allowing more rural migrants to come. Reform year is also positively correlated with the share of fiscal expenditure on education (correlation coefficient= 0.18), the prefecture fertility rate (0.17), and the proportion of state-owned firms in employment (0.16). All the other correlation coefficients are below 0.1. The non-random timing of reform causes an identification challenge, which will discussed in section 5.

5 Model 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:

$$Y_{hst} = \alpha_s + \beta_t + \sum_{k=-5}^{10} \gamma_k 1[t - policy_s = k] + \varepsilon_{ht} \tag{12}$$

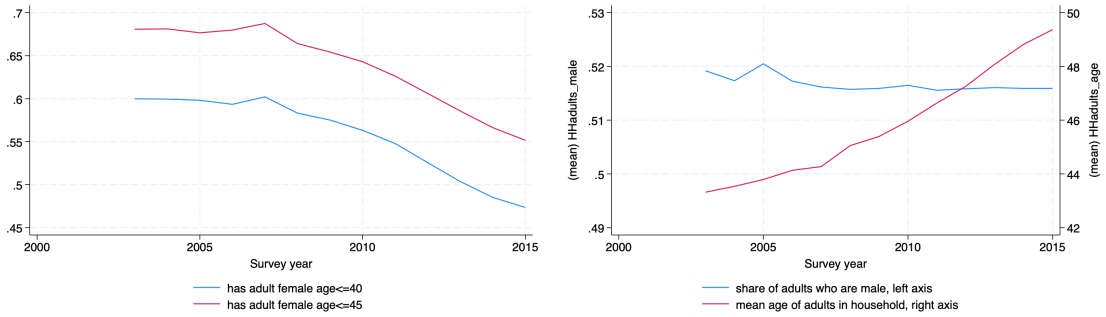


Figure 3: Gender and age composition of rural adults over time

Notes: the sample is all households in cleaned data as described in Table 2.

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).

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.[18]

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

[18]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.

Table 5: Household entry and exit rates

Year	Number of HHs	New entry	Stayers	Exit rate
2003	13787	0	0	.0581707
2004	14880	1895	12985	.0580645
2005	15016	1000	14016	.060602
2006	14766	660	14106	.0522823
2007	17036	3042	13994	.1608946
2008	14330	35	14295	.0152826
2009	16906	2795	14111	.0266769
2010	16674	219	16455	.0593139
2011	16402	717	15685	.0859042
2012	15680	687	14993	.0496811
2013	15782	881	14901	.0582309
2014	15924	1061	14863	.0833961
2015	15623	1027	14596	0

Note: This table shows the number of households in our cleaned sample from 2003 to 2015. The 'New entry' 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.

(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, if cities that had a lower urban unemployment rate chose to adopt earlier and had less restrictive rules, and 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 and not its growth, then this would be absorbed by the city fixed effect α_s in (12). One way to check the endogeneity problem is to look at pre-trends. If an omitted factor causes both earlier adoption of the reform and an upward trend of fertility, we should see an upward-sloping pre-trend. The second way to address this non-random timing of reform is to control for interactions between initial city-level conditions and year. This will absorb potential correlation between any initial urban condition

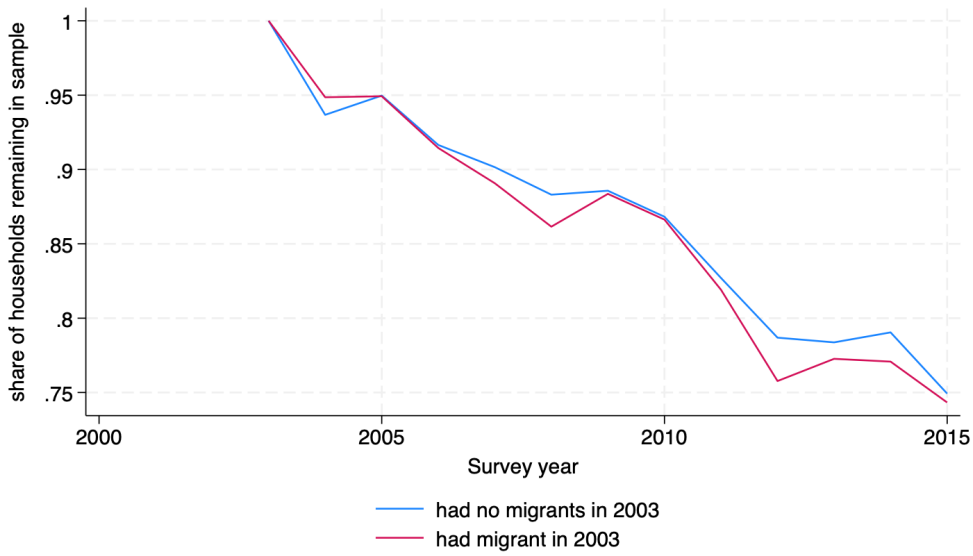


Figure 4: Sample attrition by initial status of migrant work

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.

and the slope of rural fertility rate within the prefecture.

Third, the fertility behaviour we observe among remaining rural households may be complicated by selective migration. Specifically, there are three types of migration. 1) the whole family moves and hence drops out from the rural sample, 2) households remain observed but have migrating members that we observe as living or working away from home, and 3) the household remains observed but individual migrants are unobserved. First, if child-loving households are less likely to migrate together as a family after the policy change, then in the unbalanced sample the average fertility would increase after the policy change just due to this changing composition. This selection bias will be address in subsection 6.2. Second, if a young couple migrates while their elderly parents stay in our rural sample, it's not certain whether the couple and their baby 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.[19] Due to the lack of affordable childcare in urban destinations, it's highly likely for the baby to stay behind with grandparents in the rural village and therefore most of newborns should be observed. If the reform induces the younger

[19]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.

Table 6: Correlation between reform year and initial demographic and economic conditions across cities

city condition in 2001	correlation coefficient
Urban unemployment rate	0.2158
Share of fiscal expenditure on education	0.1849
Fertility rate	0.1689
Employment share of state sector	0.1570
Migration ratio (migrant/population)	0.0998
Ratio of fiscal expenditures to fiscal revenue	0.0680
Total urban population (10,000)	0.0044
Labor productivity gap (urban vs rural)	-0.0745
Logarithim of Per capita GDP (RMB)	-0.0966
NTR gap	-0.1365

Note: This table of correlation is at the city level across 194 prefecture-level cities. The demographic and economic information is collected from China City Statistical Yearbooks and the 2000 census data. Our main sample of the NFPS contains 219 prefecture-level cities, but we don't have initial conditions for 25 of them because they underwent significant administrative changes between 2000 and 2010. At the aggregate level, the total number of prefectures in China increased from 236 at the end of 1999 to 283 by the end of 2010. Note that the fertility rate is measured at the city level, as the relaxation of migration barriers affects not only rural-to-urban migration within the same city but also migration across cities. Additional details on the initial demographic and economic conditions are provided in the appendix.

generation to migrate and all babies are left behind, our estimate will not be unbiased. If some babies migrated with the parents and are unobserved by us, then our estimate 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 particular, we will look at high-skilled individuals and those with non-agricultural hukou households. Both of them faced lower barrier to migrate and were therefore less affected by the Hukou reform.

6 Empirical Results

6.1 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 vertical axis denotes changes in the

outcome of interest (whether the household had a newborn baby). The points represent the estimated treatment effects in the treatment periods and placebo effects in the pretreatment periods, while the error bars 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. The Average Treatment Effect on the Treated is .0109954, with a standard error of .0039658.[20] As the sample mean is 0.027, the ATT is more than one third of it.

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

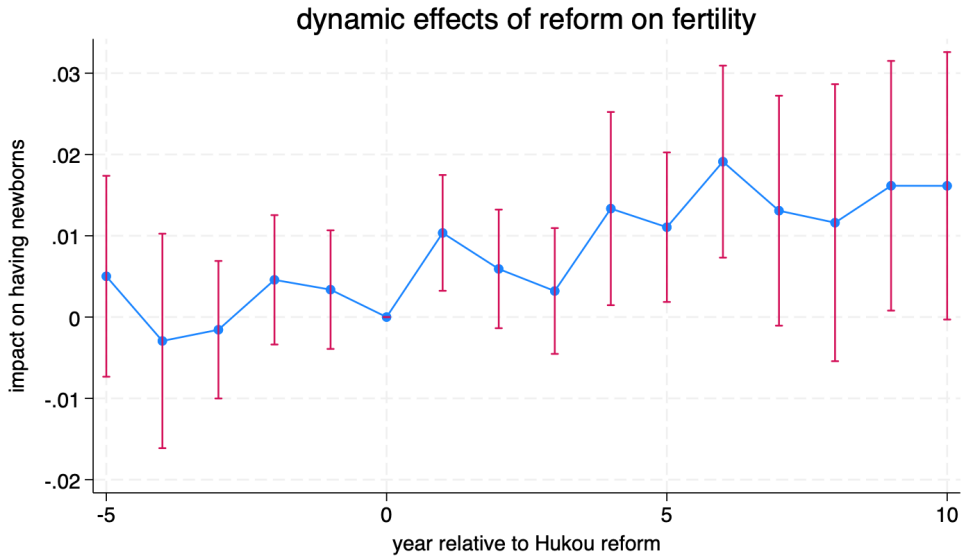


Figure 5: Effect of removing migration barriers on the probability of having a newborn

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).

Figure 5 shows that the pre-treatment effects are close to zero and not statistically significant at the conventional level.[21] Recall that 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.

[20] The ATT is computed by the STATA package *did_multiplegt*. It is an average of γ_k across all (g,t) cells which were treated.

[21] Test of joint nullity of pre-trends has $p - value = .24968$.

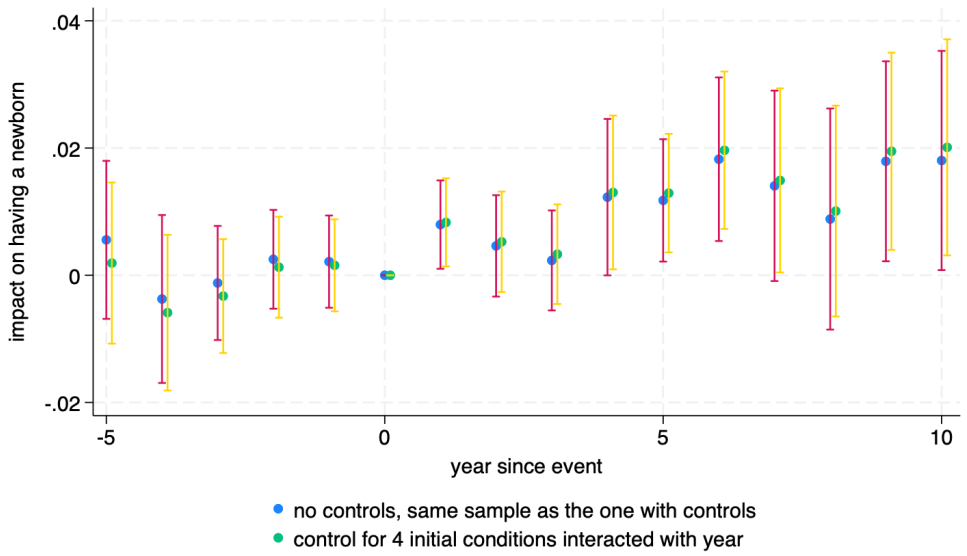


Figure 6: Effect of removing migration barriers on the probability of having a newborn, with and without city controls

Notes: Same method as figure 5. Both series shown here are based on the same sample of 187,880 observations. The controls are interactions of year and four city's initial conditions: urban unemployment rate, the share of fiscal expenditure on education, prefecture fertility rate, and the proportion of state-owned firms in employment

As a robustness check, we control for interactions between city's initial conditions and year. About 7% of the original sample is lost due to the initial city conditions being missing for them. The estimated effects with and without controls are shown in Figure 6. With controls, the ATT is .0112813, very similar to both the main estimate of .0109954 and the estimate in the smaller sample. Because adding these controls makes very little difference to estimated effects and reduces the sample size, subsequent analysis will not use these city-level controls.

We apply different estimators from the recent literature on staggered diff-in-diff and heterogeneous treatment effects. Figure 7 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 (showed in section 5). These results highlight the importance of correcting for negative weights in event study analysis with heterogeneous treatment effects.

For robustness, Table 7 also look at the number of newborns, rather than the binary outcome

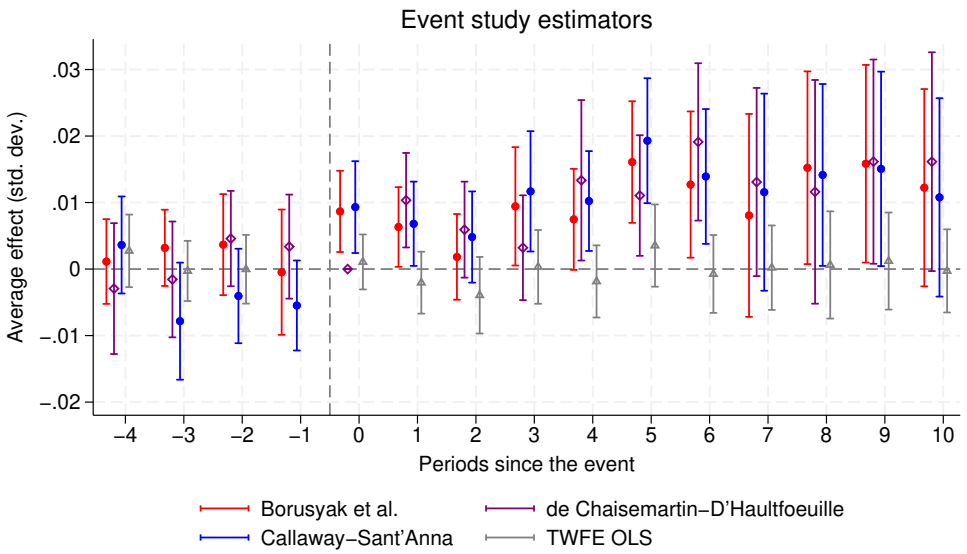


Figure 7: Effect of reform on the probability of having a newborn across different event study estimators

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 just one newborn in that year. In 2003, there was on average 0.69 women under age 40 per household. Thus, the average impact of the reform was to increase fertility by 0.016 per woman under age 40 per year. This is very sizeable: it would accumulate to an extra 0.32 child per woman if the effect persists over 20 years. We could only look at effects in the first 10 years and it shows no sign of weakening.

Table 7: Average treatment effect on fertility

Outcome	sample size	mean of outcome	estimated ATT	standard error
has newborn	202806	0.0274	0.0110***	0.0040
number of newborns	202806	0.0279	0.0112***	0.0041
has a newborn that's 1st in that generat	202806	0.0162	0.0065**	0.0033
has a newborn that's 2nd or higher birth	202806	0.0112	0.0045**	0.0020

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.

Our finding of 0.016 extra newborn per young woman per year 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 were at their urban destinations, whereas our sample is a panel of *rural* households.[22] Looking at current migrants alone misses the impact on the large numbers of people remaining in rural areas. As we argued in section 3, rural households could be affected by the reform through expectations of future migration as well as changes to current economic conditions. Our robustness check (Table 14 in the Appendix) showed the positive impact on fertility was not concentrated in households with current migrants. Moreover, the longitudinal nature of our data has allowed us to test for selective migration (in subsection 6.2) and therefore be more confident of the findings.

The last two rows in Table 7 distinguish newborns by whether they are the first in their generation. This is a proxy to distinguish 1st-borns and higher birth orders, because the survey only identifies relationship between each individual and the head. The reform increased the birth of 1st-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 1st-borns suggests the reform may have increased marriage rate or accelerated the arrival of the first child.

6.2 Attrition from sample

Our main analysis (Figure 5) has used all households regardless of how many times they have appeared. The number and types of households in the sample vary over time. If there was selective attrition, in particular if the reform caused households with low-fertility-preferences to migrate permanently (and therefore drop out of the sample), then our estimate would be biased upwards.

In the cleaned sample of households (used in the main analysis), the annual attrition rate is not high. On average, 93.5% observations in year t continue to be observed in $t+1$. Of all households observed in 2003 (start year of the sample), 75% were observed in 2015 (last year of the sample).

Whether attrition depends on fertility preferences matters for the consistency of our main result. To test this indirectly, we look at attrition rate by existing number of children. For every household type listed in Table 4, 93.1-94.2% households remain in the sample for the next year. Whether the observed number of children is 0,1 or 2, the proportion is similar and all within 93.1-93.7% range. Households with 3 or more children have a slightly higher attrition rate: 91.3% of them remain for the next year. But they account for a relatively small share of all households. In short, there is no evidence to suggest that households with lower fertility preferences are more

[22] There are also differences in the timing of policy and sample period. 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.

likely to leave the sample.

Household attrition can be due to reasons other migration, such as death or survey non-response. We have examined the pairwise correlation between attrition and a wide range of household circumstances including the share of adults who work non-locally, average work days, average age and so on. All the correlation coefficient are below 0.03.

However, attrition from the sample is correlated with the Hukou reform. In Figure 8, we apply the event study approach to the outcome of 'remaining' in the sample in the next year. This outcome has a mean of 93.5%, and we found the reform has a significant negative impact on it especially in years 5-10. The ATT is $-.0546155$. There was no significant pre-trend. The negative effects 5-10 years after the reform could be because it took years for rural migrants to settle in urban destinations (there could be frictions in job search and satisfying other requirements of Hukou application in urban destinations); or it could be that after a young migrant or couple moves permanently to the urban destination, it took years for them to save money before bringing their parents and children over.

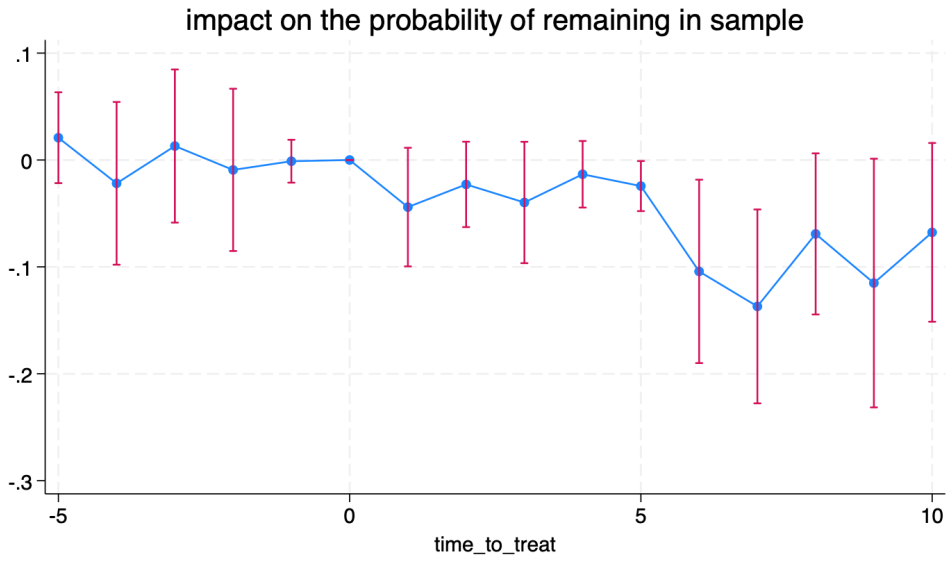


Figure 8: Effect of reform on the probability of households remaining from t to $t+1$

Note: the sample is restricted to households who appeared every year over the 2003-2015 period. The estimation uses the same event study method from de Chaisemartin and D'Haultfoeuille (2020).

As a robustness check for the non-random attrition, we look at a balanced sample: households who appeared every year throughout the sample period. This balanced sample has nearly 100,000 observations at household-year level, just under half of the original sample. The mean of the outcome variable is 0.0268 in the balanced sample, very similar to the mean of 0.0274 in the

original sample. Figure 9 shows that the estimated effect in the balanced sample is also increasing gradually over time, and averages at .019648. This ATT is even bigger than the .0109954 estimate in the original sample.

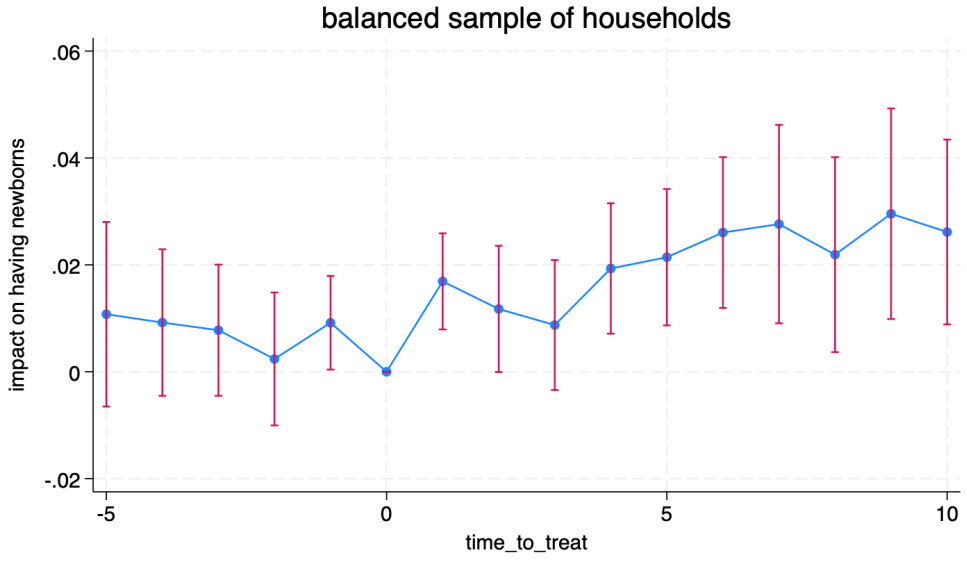


Figure 9: Effect of reform on fertility in the balanced sample of households

Note: the sample is restricted to households who appeared every year over the 2003-2015 period. The estimation uses the same event study method from de Chaisemartin and D'Haultfoeuille (2020).

6.3 by subgroup

This subsection explores how the impact of the reform varies across demographic groups, as suggestive evidence for different mechanisms.

Figure 10 shows differential effects by the presence of an adult woman under 40. 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. The blue line in Figure 10 shows stronger effects of the policy: the effect goes up to around 3% 10 years after for households with a woman under 40. By contrast, the red line shows that 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 8 reports a mean of 0.2% in this subsample.

Note the presence of a young woman 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 11 in the Appendix shows the presence of women had a declining trend in the treated regions before the reform. Thus, it's a valid concern that the presence of women of child-bearing age is endogenous. Therefore, our results and most of the following analysis will not restrict sample by the contemporaneous presence of women under any age.

As our main sample contains a lot of households for whom fertility is extremely likely to be zero, our main estimate of .0109954 per household per year has been diluted. At the start of our sample period (2003), 60% households contained a woman under age 40. Thus, per household of fertility age and per year, the average impact of the reform was $.0109954/0.6 = 0.018$.

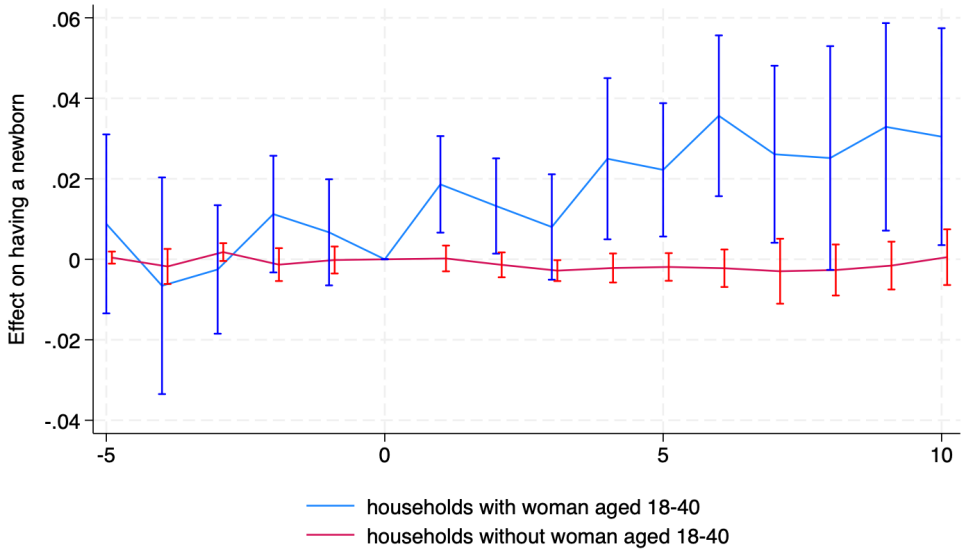


Figure 10: Effect of reform on fertility among households by the presence of women under 40

Note: the blue line shows estimation from the subsample of households who had a woman under 40. The red one is for households who did not. The same event study method is applied as in Figure 5.

Table 8 reports the Average Treatment Effect on Treated (ATT) for a range of sub-samples. Conditional on the presence of a woman under 40 , the ATT is 0.0216 and highly significant. This is nearly half of the mean of outcome in this subsample. By contrast, in the absence of such a woman, less than 0.2% of households had a newborn and the ATT is close to zero. If we draw the line at age 45 instead of 40, we would get similar patterns.

The model presented in section 3 has predictions related to grandparents' availability G. The model predicted both a higher level of fertility and a more positive impact from the reform in households with higher grandparents availability. Here we examine a few proxies for having

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 (0.0375 vs 0.0064 in Table 8). The ATT is 0.0142 in multi-generational households, which is statistically significant and more than one third of the subsample's mean. By contrast, the ATT is near zero in single-generation households.

Table 8: Average treatment effect by presence/absence of potential carers

Sub-sample	sample size	mean of outcome	estimated ATT	standard error
households with woman aged 18-40	112949	0.0475	0.0216***	0.0064
households without woman aged 18-40	89857	0.0021	-0.0016	0.0017
households with woman aged 18-45	129369	0.0419	0.0179***	0.0058
households without woman aged 18-45	73437	0.0019	-0.0012	0.0015
households with multiple generations of adults	137078	0.0375	0.0142**	0.0056
households without multiple generations of adults	65728	0.0064	0.0028	0.0027
households with healthy women aged>=47	98951	0.0346	0.0135*	0.0070
households without healthy women aged>=47	103855	0.0205	0.0094***	0.0035
households with healthy people aged>=47	125385	0.0327	0.0138**	0.0054
households without healthy people aged>=47	77421	0.0189	0.0083**	0.0041

Note: each row uses the same event study specification as in Figure 5 and Table 7. 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. 47 is chosen as an age threshold because vast majority of new grandparents are aged 47+.[23] And there would be very few new mothers above 47. 'healthy' is self-reported by the respondent, if they categorise their health status as excellent or good, rather than the lower three categories. Households with a healthy woman over 47 has a much higher chance of having a newborn than the others. Their ATT is also more positive than the others, and it's statistically significant. The same pattern is observed if we count older men as well as women.

These differential results by proxies of grandparents availability support the hypothesis that childcare support from extended family lowers the opportunity cost of childcare and thereby allowing the reform have a more pronounced positive impact on fertility.

Next, Table 9 explores the policy impact by characteristics of the potential parents observed in 2003. 'Potential mother' refers to the youngest adult female in the household, and 'potential

[23] Among 3 generation households with newborns, the age of grandparents has median 54 and the 10th percentile at 47.

father' refers to the youngest adult male. This means we lose households who did not appear in 2003 or did not have a man/woman under 40.

Conditional on having a woman under 40 observed in 2003, we find the ATT is over 2% if the potential mother had less than 9 years of education. For comparison, conditional on the household having a woman under 40 observed in 2003, the ATT is 0.016.[24] Recall that lower-skilled people are exactly the types of people whose probability of urban settlement was most affected by the reform. Higher probability of urban settlement improves their expected lifetime incomes and expected educational opportunities for their children. Hence one would expect the impact on fertility to be more positive. This is true for less-educated fathers as well. Similarly, if we split households by the education of the household head or the most educated person in the household, we would still find the impact of the reform to be more positive in the less-educated subgroup. This is shown in table 16 in Appendix.

The ATT is also more positive if the potential mother lived at home for less than 330 days, if she worked non-locally for 180+ days than if she did not, and if she worked in any way for 30+days than if she did not. There are similar differential effects by men's initial conditions. The ATT is above 2.0% if the potential father worked non-locally (both for 180+ days and 30+ days). The ATT is negative if the potential dad did not work at least 30 days in any category, which is relatively rare in rural China.

One way to interpret the differential impact by past migration history is to consider past migration as a proxy for the intention and ability to move to urban areas in future in the absence of the reform. As described in section 3, the negative impact of higher costs of raising children on fertility (the third channel) is irrelevant to households who were going to migrate regardless of the Hukou reform, and therefore the impact of the reform on these households are less negative - they are much more positive. Another way of interpretation is that people with prior migration history were best positioned to capitalise on the future opportunities offered by the reform to their children. Thus, it's consistent with the hypothesis that the reform increased fertility intention by improving the expected quality of children.

This is not to say that the reform only had a positive impact on households with migrants and zero impact otherwise. There is a difference between having migration experience in the past and actually being a migrant at the time of fertility decision. Table 14 in the Appendix shows the ATT of the reform by whether the household currently has a migrant. There is very little difference in the ATT between household with and without current migrants: both are around 0.01. A plausible explanation is that the reform mainly increases fertility through the expectation

[24]This is lower than the estimate of 0.02 in Table 8 because the previous Table conditioned on a potential mother observed in year t whereas now we condition on such a woman observed in 2003.

Table 9: Average treatment effect by 2003 characteristics of potential parents

Sub-sample if the potential mother was	sample size	mean of outcome	estimated ATT	standard error
present in 2003	93561	0.0306	0.0165***	0.0057
>25 year old	54987	0.0232	0.0149***	0.0051
<=25 year old	38574	0.0410	0.0187*	0.0098
<9 years of education	56743	0.0308	0.0223***	0.0063
>=9 years of education	36818	0.0301	0.0078	0.0089
does not work non-locally for 180+ days	70125	0.0288	0.0139**	0.0063
works non-locally for 180+ days	23436	0.0359	0.0183*	0.0095
does not work non-locally for 30+ days	65013	0.0284	0.0149**	0.0064
works non-locally for 30+ days	28548	0.0355	0.0144	0.0089
does not work for 30+ days	18936	0.0308	0.0073	0.0124
works (anywhere) for 30+ days	74625	0.0305	0.0183***	0.0058
lives at home for <180 days	25737	0.0342	0.0181**	0.0085
lives at home for 180+ days	67824	0.0292	0.0136**	0.0066
lives at home for <330 days	31964	0.0346	0.0217**	0.0088
lives at home for 330+ days	61597	0.0285	0.0124*	0.0068

Sub-sample if the potential father was	sample size	mean of outcome	estimated ATT	standard error
present in 2003	104060	0.0349	0.0141***	0.0054
>25 year old	59615	0.0242	0.0163**	0.0067
<=25 year old	44445	0.0492	0.0121	0.0084
<9 years of education	53829	0.0348	0.0145*	0.0078
>=9 years of education	50231	0.0350	0.0096	0.0062
does not work non-locally for 180+ days	64424	0.0304	0.0092	0.0057
works non-locally for 180+ days	39636	0.0421	0.0221**	0.0094
does not work non-locally for 30+ days	52852	0.0296	0.0075	0.0057
works non-locally for 30+ days	51208	0.0403	0.0208***	0.0079
does not work for 30+ days	16738	0.0301	-0.0187	0.0133
works (anywhere) for 30+ days	87322	0.0358	0.0204***	0.0056
lives at home for <180 days	37398	0.0401	0.0111	0.0091
lives at home for 180+ days	66662	0.0319	0.0144***	0.0054
lives at home for <330 days	49161	0.0389	0.0106	0.0085
lives at home for 330+ days	54899	0.0313	0.0161***	0.0056

Note: each row uses the same event study specification as in Figure 5 and Table 7. Each row selects a subsample as labelled.'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.

of future benefits of a permanent move to urban destinations, rather than through an immediate increase in migration and the associated benefits. As described under the first mechanism of the model in section 3: the expectation of future benefits is a positive income effect that does not require an immediate move to the urban area.

Table 10: Average treatment effect by agriculture and non-agriculture Hukou

Sub-sample	sample size	mean of outcome	estimated ATT	standard error
NON-Agricultural hukou	7338	0.0183	0.0074	0.0065
Agricultural hukou	194940	0.0277	0.0128***	0.0037

Note: each row uses the same event study specification as Table 7. Each row selects a subsample as labelled.
*** p<0.01, ** p<0.05, * p<0.10.

Finally, we examine the minority of rural households who had non-agriculture Hukou. Their Hukou status means they have less to gain from the reform. Table 10 shows that less than 5% of the entire sample are households with non-agriculture Hukou. The ATT on this group is smaller than those with agriculture Hukou, although the difference is insignificant. The standard error is small enough to rule out ATT above 0.013.

6.4 Impact of reform on other outcomes

This subsection applies the same event study method to other outcomes at the household-year level and at the adult-year level.

Table 11 shows a significant positive ATT on household annual expenditure, at 6.0%. This is consistent with the first mechanism described in section 3: the reform increases both fertility and consumption via a positive income effect. Note that the reform's impact on full income encompasses higher future wages as well as entitlements to more generous social benefits. Hence, looking at households actual income in the short term will understate the improvement in full income. Table 11 shows that the estimated impact on household annual income is insignificant and we can rule out effects larger than 10%.

If there's only a positive income effect, the model predicts that both educational expenditure and consumption should increase by the same proportion as full-income. However, the reform may also have affected preferences and the returns to education investment, and therefore the total impact on educational expenditure is ambiguous. The survey contains educational expenditure as reported by rural households, but it is missing in nearly two thirds of the sample. Among the non-missing observations, we find a negative impact that's statistically insignificant.

Table 11: Average treatment effect on other household outcomes

Outcome	sample size	mean of outcome	estimated ATT	standard error
total earnings from non-local work	202806	1.28e+04	586.2906	766.7326
total log earnings from non-local work	109892	9.5910	0.0529	0.0497
log household total income	202244	10.2304	0.0188	0.0276
log household net annual income	201425	9.9682	0.0371	0.0265
log household annual expenditure	201561	9.9004	0.0602**	0.0297
log educational expenditure	66223	7.1240	-0.0916	0.0669
remittance from non-permanent family members	16031	2593.8718	-752.9649	475.2727
number of people aged 0-6	202806	0.2241	0.0200	0.0148
number of people aged 7-17	202806	0.4456	-0.0492**	0.0222
number of people aged 18-40	202806	1.3678	-0.0143	0.0265
number of people aged 41-60	202806	1.2534	0.0021	0.0229
number of people aged 61+	202806	0.5703	0.0013	0.0149
number of adults working away for 180+days	202806	0.7486	0.0399	0.0288
whether anyone works away for 180+days	202806	0.4556	0.0042	0.0143

Note: each row uses the same event study specification as in Figure 5 and Table 7. 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 4. 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.

Another limitation due to survey design is that we cannot observe parents and child if and after they separate both physically and economically from the rural grandparents. This means the the survey cannot capture longer-term increases in educational expenditure, so the estimated impact (an insignificant negative effect) should be treated with caution.

We will also explore the impact of reform on the number of people by broad age group. The objective is to check which age groups are more likely to separate from the rural household; this is conditional on the household still appearing in the data. Table 11 also shows that the reform increases the number of children under school age, which is consistent with the main finding of increased newborns. Moreover, there is a significant negative impact on the number of school-aged children, and much smaller impact on the number of adults. This is consistent with the hypothesis that attracted by access to better schools, rural migrants have more babies, let them be cared for rurally before school age, and take them to urban areas during school age. This does not necessarily require an increase in the parents' probability of migration; the parents may migrate in the absence of the reform, and the reform just incentivised them to bring school-aged children along.

Table 12: 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	3.8083	0.0522	0.0370
log daily wage from non-local work	men below 40	146937	3.7504	0.0352	0.0434
log daily wage from non-local work	lower-educated men	57190	3.6668	0.0740**	0.0301
log daily wage from non-local work	women	314004	3.5213	0.0261	0.0440
log daily wage from non-local work	women below 40	130376	3.4964	0.0094	0.0490
log daily wage from non-local work	lower-educated women	57884	3.3872	0.0406	0.0435

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.

Next, we examine the impact of the reform on daily wages by gender. In the model, increasing male wage will have a positive income effect only, while female wage may also have a substitution effect. The survey asks about earnings from 'non-local work' and the number of days they work non-locally. The survey does not contain individuals' wages or earnings except for 'non-local work'. In addition, we cannot observe individuals wages many years after urban settlement, which may increase more than in the short-term. Thus, the measure of wage in the data may

Table 13: Average treatment effect on adult-level labour supply

Outcome	gender	sample size	mean of outcome	estimated ATT	standard error
number of days worked	men	332505	197.9680	5.5196	3.3957
number of days worked	women	314004	148.3935	-1.0893	3.8001
number of days worked outside local village/town	men	332505	91.4244	2.1470	2.9844
number of days worked outside local village/town	women	314004	53.3403	2.7515	2.4142
number of days worked locally outside agriculature	men	332505	48.6960	4.7967	3.3388
number of days worked locally outside agriculature	women	314004	32.8171	3.8218	2.8716
number of days worked locally in agriculature	men	332505	62.5597	-0.5274	3.0185
number of days worked locally in agriculature	women	314004	65.2630	-5.7492*	3.4463
whether worked at least 30 days	men	332505	0.7845	0.0287**	0.0122
whether worked at least 30 days	women	314004	0.6771	0.0170	0.0146
whether worked non-locally at least 180 days	men	332505	0.2952	0.0046	0.0097
whether worked non-locally at least 180 days	women	314004	0.1710	0.0096	0.0079
whether worked non-locally at least 30 days	men	332505	0.3633	0.0157	0.0124
whether worked non-locally at least 30 days	women	314004	0.2014	0.0110	0.0091
whether worked in non-agriculture for at least 30	men	332505	0.6760	0.0210	0.0138
whether worked in non-agriculture for at least 30	women	314004	0.5776	0.0165	0.0149

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 adult-year level. 'Mean of outcome' is the mean of the outcome variable across adults of the labelled gender.
*** p<0.01, ** p<0.05, * p<0.10.

increase less than the concept of wage in the model presented earlier. Table 12 shows the ATT on men's wage is 5.2% for all men, and 7.4% for lower-educated men. Only the latter is statistically significant at 5% level. The ATT on women's wage is smaller and insignificant. These differential results are in line with the hypothesis that male-specific wage gains cause an income effect while the absence of large wage effects for women implies no countervailing substitution effect, and therefore the overall impact on fertility is positive.

Finally, Table 13 reports the impact on the number of days worked in different categories. The ATT on total number of days worked is 5.5 for men and -1.1 for women, neither of which is significant. The estimated effects on the number of days of non-local work are positive for both gender, but small and insignificant. The standard errors are small enough to rule out effects larger than 15 days a year. Table 13 also shows a near-zero effect on the binary outcome of working non-locally for at least 180 days. In other words, the reform has not led to a large increase in individuals' temporary migration for work.[25] For labor supply outcomes, the only statistically significant result is a small increase (2.9%) in the proportion of men working at 30 days a year. The absence of large labour supply responses is consistent with multiple interpretations: labour supply may be inelastic due to labour market frictions; there is limited substitution from formal work to childcare because the primary carers (e.g. grandparents) did not do much work any way; the positive impact on fertility could stem from expectations of better future opportunities for children rather than an immediate significant increase in earnings.

7 Conclusion

This paper provides new causal evidence that reducing internal migration barriers can raise fertility in rural areas. Exploiting the staggered rollout of China's Hukou reform across 283 cities between 2002 and 2015, 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 per year - a substantial effect representing one-third of the baseline mean. For women under age 40, the policy effect is an increase of 0.016 newborns per woman per year.

The fertility response was more pronounced in households with certain characteristics: where women of childbearing age were present, multiple generations of adults co-residing, presence of healthy older adults, and where prospective parents were already working outside their local village/town at the start of the sample. These patterns point to the importance of extended family structures and pre-existing migration experience in shaping the household response.

[25] This finding is distinct from the positive impact we found on household-level attrition (Figure 8), which could be permanent migration of the whole family. By contrast, here we are looking at individuals' temporary migration while their family remain.

The observed increase in fertility likely reflects a combination of higher household incomes, access to better schools, and informal childcare support from rural grandparents. Evidence of increased household consumption and male wage following the reform supports the presence of a positive income effect. While other mechanisms may be at play and hard to quantify in existing data, the robustness of the main finding–a sizeable fertility increase following the Hukou reform–remains clear across a range of specifications and robustness checks.

These results speak to a broader literature on the demographic consequences of public policy. Migration policy, typically viewed through the lens of labor mobility, can also generate large and unintended effects on fertility, particularly in rural settings where structural constraints on mobility have long shaped family decisions. The fertility response observed here runs counter to concerns that modernisation necessarily depresses fertility and suggests that loosening institutional constraints can create space for more optimistic demographic behaviour, especially where family networks remain strong.

Finally, while much prior research on Hukou reform has focused on migrants in urban destinations, our findings highlight that households who remain in rural areas are also deeply affected. Migration reforms alter expectations, economic conditions, and intergenerational dynamics even among current non-migrants, in ways that ultimately shape fertility and family formation. Looking ahead, future research could usefully explore how these migration opportunities affect human capital investments and family dynamics in rural areas.

References

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.

An, L., Qin, Y., Wu, J. and You, W. (2024). The local labor market effect of relaxing internal migration restrictions: evidence from china. *Journal of Labor Economics*, **42** (1), 161–200.

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.

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.

Bergsvik, J., Fauske, A. and Hart, R. K. (2021). Can policies stall the fertility fall? a systematic review of the (quasi-) experimental literature. *Population and Development Review*, **47** (4), 913–964.

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.

Chen, X. 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.

—, Shi, S. and Tang, Y. (2019). Valuing the urban hukou in china: Evidence from a regression discontinuity design for housing prices. *Journal of Development Economics*, **141**, 102381.

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

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

Galor, O. (2011). *Unified growth theory*. Princeton University Press.

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

— and Weil, D. N. (1996). The gender gap, fertility, and growth. *American Economic Review*, **86** (3).

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

Hart, R. K., Bergsvik, J., Fauske, A. and Kim, W. (2024). Causal analysis of policy effects on fertility. *Handbook of Labor, Human Resources and Population Economics*, pp. 1–25.

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. 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.

Johnston, L. A. (2019). The economic demography transition: is china's 'not rich, first old'circumstance a barrier to growth? *Australian Economic Review*, **52** (4), 406–426.

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.

Liu, X., Xu, Y. and Zou, J. (2024). Migration barrier relaxation and entrepreneurship: Evidence from the hukou reform in china. *Labour Economics*, **90**, 102605.

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

National Bureau of Statistics of China (2018). 1Census Survey (2015).

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.

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

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.

Zhang, W., Zou, X., Luo, C. and Yuan, L. (2024). Hukou reform and labor market outcomes of urban natives in china. *Journal of Population Economics*, **37** (2), 1–41.

8 Appendix

8.1 Derivation of model predictions

This section derives the solution to the household's optimisation problem described in section 3. The household maximises utility subject to a few constraints:

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

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

$$\frac{1-\gamma}{c} = \lambda_c \tag{13}$$

$$\gamma\beta = \lambda_q \tag{14}$$

$$\lambda_q \delta_e / E = \lambda_c \tag{15}$$

$$\lambda_w = \lambda_c w_w \tag{16}$$

$$\lambda_\tau = \lambda_w \tag{17}$$

$$\frac{\gamma}{n} - \frac{\lambda_q \delta_e}{n} - \lambda_\tau \tau(1-G) - \lambda_c p^n = 0 \tag{18}$$

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. $\Rightarrow$:

$$\gamma\beta\delta_e/E = \frac{1-\gamma}{c}$$
$$\Rightarrow E = \frac{\gamma\beta\delta_e c}{(1-\gamma)} \tag{19}$$

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:

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

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

$$\frac{\gamma-\gamma\beta\delta_e}{n} = \frac{(1-\gamma)p^*}{c} \tag{20}$$
$$\Rightarrow \frac{c}{n} = \frac{(1-\gamma)p^*}{\gamma-\gamma\beta\delta_e} \tag{21}$$

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:

$$c + E + p^n n + w_m(1-h_m) + w_w(1-h_w) = w_m + w_w + y_0 = \bar{y} \tag{22}$$
$$c + E + p^n n + w_m\tau_m + w_w\tau_w = \bar{y} \tag{23}$$
$$\tag{24}$$

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

$$c + E + p^n n + 0 + w_m n\tau(1-G) = \bar{y} \tag{25}$$
$$c + E + (p^n + w_m \tau(1-G))n = \bar{y} \tag{26}$$
$$c + E + p^* n = \bar{y} \tag{27}$$

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:

$$c + c\frac{\gamma\beta\delta_e}{(1-\gamma)} + p^* n = \bar{y} \tag{28}$$
$$c[1 + \frac{\gamma\beta\delta_e}{(1-\gamma)}] + p^* n = \bar{y} \tag{29}$$
$$c[\frac{1-\gamma+\gamma\beta\delta_e}{(1-\gamma)}] + p^* n = \bar{y} \tag{30}$$

Use (21) 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 (21) 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 (19) to get E:

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

8.2 Additional empirical results

Figure 11 examines the impact of the reform on the presence of a woman under age 40. Figure 12 examines the impact of the reform on grandparents' availability in rural households, using two proxies.

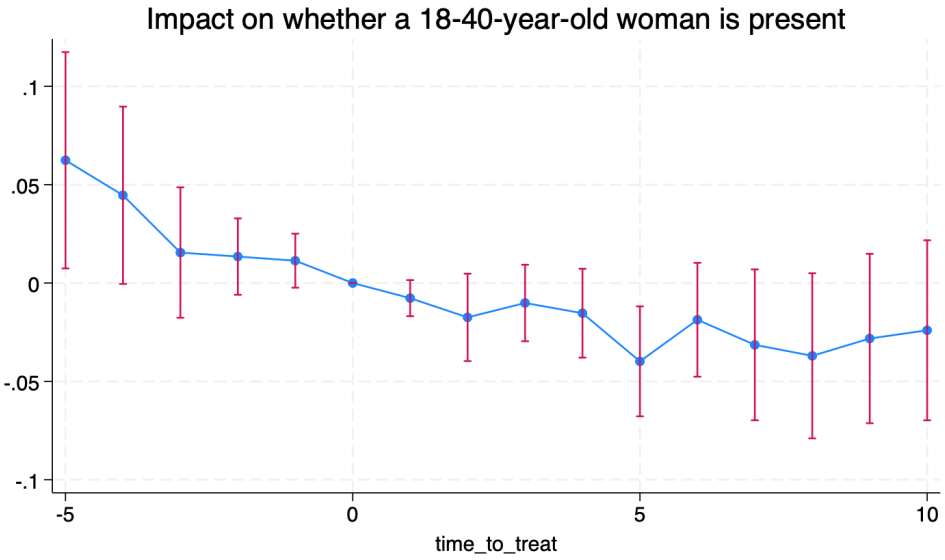


Figure 11: Effect of reform on the presence of women under 40

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

Table 16 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

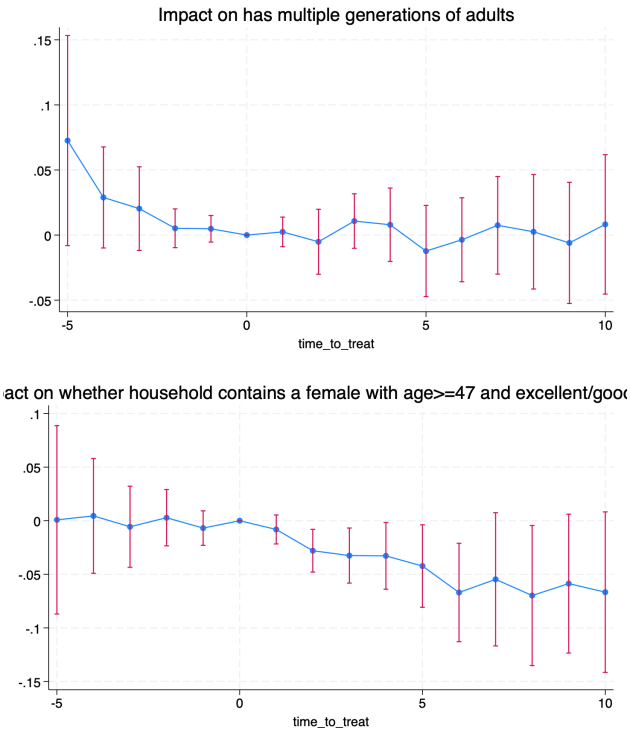


Figure 12: Effect of reform on grandparents availability

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+.

ATT is much bigger in household with less-educated adults.

8.3 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 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, migration stock, 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.

Table 14: Average treatment effect by current presence of migrants

Outcome	sample size	mean of outcome	estimated ATT	standard error
households without migrant workers	110413	0.0197	0.0106***	0.0036
households with migrant workers	92393	0.0366	0.0115*	0.0065
households without migrants	102324	0.0194	0.0098***	0.0036
households with migrants	100482	0.0356	0.0116	0.0071

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.

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.

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 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 17 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 sector. According to Table 17, 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 0.22 (Table 17).

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 share of state-owned enterprises (SOEs) in employment** based on data from the Chinese Annual Survey of Industrial Firms (CASIF). Table 17 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 17 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.

In summary, Table 17 shows that most city-level initial conditions are only weakly correlated with the year of reform. There are four conditions where the correlation coefficient exceed 0.15 in absolute terms. As a robustness check, we added interactions between those four conditions and year into the event-study regression and the results were shown in Figure 6. It did not much noticeable differences to the estimated effects.

Table 15: Average treatment effect on other household outcomes in balanced sample

Outcome	sample size	mean of outcome	estimated ATT	standard error
number of newborns in household this year	98735	0.0273	0.0202***	0.0052
has a newborn that's 1st in that generation	98735	0.0160	0.0136***	0.0044
has additional newborn in an existing generation	98735	0.0108	0.0060**	0.0027
number of persons age<18 in household	98735	0.6532	-0.0208	0.0278
number of persons age>=18 in household	98735	3.1222	0.0106	0.0367
number of people aged 0-6	98735	0.2231	0.0408**	0.0197
number of people aged 7-17	98735	0.4301	-0.0616**	0.0293
number of people aged 18-40	98735	1.3013	0.0207	0.0294
number of people aged 41-60	98735	1.2498	-0.0254	0.0301
number of people aged 61+	98735	0.5812	0.0212	0.0186
number of adults working away for 180+days	98735	0.7199	0.0029	0.0394
whether anyone works away for 180+days	98735	0.4416	-0.0152	0.0200
total earnings from non-local work	98735	1.22e+04	-126.7179	915.3257
total log earnings from non-local work	51692	9.5718	0.0697	0.0728
mean number of days worked non-locally	98730	64.2898	-0.8254	3.1010
mean number of days worked in local agriculture	98733	62.6463	-2.6005	3.2821
mean number of days worked in local nonagriculture	98735	39.4464	3.2269	3.3542
log household total income	98469	10.1650	0.0035	0.0378
log household net annual income	98074	9.9152	0.0059	0.0387
log household annual expenditure	98153	9.8157	0.0270	0.0417
health expenditure	55811	1355.6783	16.9993	113.6065
educational expenditure	33937	2584.2275	-330.7797	226.7905
log educational expenditure	31641	7.0699	-0.1557*	0.0940
remittance from non-permanent family members	7969	2315.7764	-691.7272	605.4288

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.
*** p<0.01, ** p<0.05, * p<0.10.

Table 16: 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	0.0272	0.0123**	0.0052
if head has 7+ years of education	103182	0.0276	0.0083*	0.0049
if the most educated person has <12 years of education	159207	0.0259	0.0121***	0.0041
if the most educated person has 12+ years of education	43599	0.0328	0.0031	0.0104

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 17: Summary Statistics of initial city-level conditions

City-level condition in 2001	mean	standard deviation	correlation coefficient with year of Hukou reform
Urban unemployment rate	0.0888	0.0450	0.2158
Share of fiscal expenditure on education	0.3004	0.2499	0.1849
Fertility rate	1.2311	0.2905	0.1689
Employment share of state sector	0.3707	0.2199	0.1570
Migration ratio (migrant/population)	0.1054	0.0842	0.0998
Ratio of fiscal expenditures to fiscal revenue	1.9109	1.4195	0.0680
Total urban population (10,000)	101.5347	92.1365	0.0044
Labor productivity gap (urban vs rural)	0.3651	0.1976	-0.0745
Logarithim of Per capita GDP (RMB)	8.8508	0.5917	-0.0966
NTR gap	0.3089	0.0842	-0.1365

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.