Fertility, Child Gender, and Parental Migration Decision: Evidence from One Child Policy in China*

Zibin Huang† Lin Lin‡ Junsen Zhang§

April 10, 2019

Abstract

We investigate the effect of the number and gender of children on parents' rural-to-urban migration in China and try to evaluate the effect of One Child Policy (OCP) on China's urbanization. We propose a new semi-parametric method to solve an identification difficulty in previous studies and estimate the two effects separately. Results show that the addition of one girl in the family will result in a 13.7% increase in the probability that both parents in rural households migrate to urban areas; whereas the addition of one boy will result in a 24.3% increase in this probability. It implies that the reduction in the number of children will hinder the urbanization. Comparing to the traditional instrumental variable method, we find that without considering the effect of child gender, the estimate of the effect of the number of children will be heavily downward biased which leads to an opposite policy implication.

Keywords: Fertility, Child gender, Parental migration, One Child Policy, Invalid instrumental variable, Semi-parametric model

JEL Codes: D19, J13, O15

---

*We wish to thank Ronni Pavan, Joshua Angrist, Greg Caetano, Bin Chen, Lisa Kahn, Zhentao Shi, John Singleton, Kegan Tan, Hao Teng, Feng Yao and Nese Yildiz for their inspiring suggestions on this paper. We also thank all the participants and discussants attending China Meeting of Econometrics Society 2018, Chinese Economist Society Annual China Conference 2018 and a presentation at University of Rochester for their comments.

†Department of Economics, University of Rochester. Email: zhuang18@UR.Rochester.edu.

‡Department of Economics, The Chinese University of Hong Kong. Email: michellelinlin@link.cuhk.edu.hk.

§Department of Economics, The Chinese University of Hong Kong. Email: jszhang@cuhk.edu.hk.

1 Introduction

In this study, we aim to estimate the effect of the number of children and the effect of the gender composition of children on parents' migration decisions using One Child Policy (OCP) as a natural experiment. OCP, the most rigorous fertility control policy in human history, has affected millions of people for over 30 years in China. Previous studies on the effect of OCP on China's development mostly focus on human capital accumulation (Li and Zhang, 2016; Qian, 2009; Rosenzweig and Zhang, 2009). We seek to explore the effect of OCP from a new angle. A crucial feature of China's development has been a massive amount of agricultural population migrating to cities.[1] The vast wave of migration has continued to boom, which links rural to urban areas and spurs the development of cities. Therefore, we are interested in studying how OCP affects rural-to-urban migration, which can shed some light on the impact of OCP on China's urbanization process.

There are two possible channels for OCP to affect migration. One is through changing the number of children in families, the other one is through changing the gender composition of children in families due to the gender-based abortion.[2]

First, the number of children in the family can affect parental migration decision.[3] Due to the Hukou system in China,[4] it is tough for migrant workers with rural Hukou to send their children to good schools in cities. Thus, many rural migrant workers have to leave their children at hometown, which incurs the separation cost.[5] On the other hand, modern sectors in cities, such as manufacturing firms, offer better wages. The economic pressure resulted from raising more children may drive parents to migrate to cities for work (Du, Park, and Wang, 2005). Consequently, the direction of the fertility effect on parental migration depends on the relative magnitudes of the separation cost and the demand for income.

[1]According to the National Bureau of Statistics of China, the number of migrant workers had grown to about 281 million in 2016, which is undoubtedly the largest migrant worker group in human history.

[2]The problem of boy preference in China has been a long-lasting issue that is exacerbated by the implementation of OCP. Based on the government statistics, the proportion of boys to girls at birth rocketed up to 121:100 in 2004. Ebenstein (2010) finds that places with more restrictive fertility policies also have more distorted sex ratios towards boys in China. Chen, Li, and Meng (2013) find that approximately 40% to 50% of the increase in sex ratio can be attributed to the improved local access to ultrasound technology, which is the main gender detection method during pregnancy. These two studies imply the existence of gender selection (gender-based abortion) at high order births but not the first birth and indicate a positive relationship between sex ratio distortion and the intensity of fertility control policy.

[3]The fertility effects on various family outcomes other than migration, such as parental marital status (Cáceres-Delpiano, 2006; Jena, 2006) and parental labor market activity (Angrist and Evans, 1998; Adda, Dustmann, and Stevens, 2017; Jacobsen, Pearce III, and Rosenbloom, 2001), have been well studied.

[4]Hukou system is a household registration system that categorizes individuals by their hometowns in China. We will explain it in more detail in Section 2.1.

[5]The separation leads to lower human capital accumulation for the children (Zhang et al., 2014; Ye and Lu, 2011; Jia and Tian, 2010; Chang, Dong, and MacPhail, 2011).

Second, the gender of children can also affect parental migration decision.[6] Due to the traditional Chinese culture, rural families often have a strong boy preference. Li and Yi (2015) found that parents migrate to increase income, in order to improve their son's relative standing in response to the ever-rising pressure for males in China's marriage market. On the contrary, the son preference can lead to a higher separation cost for boys than for girls, because parents may feel worse for the human capital loss suffered by their son than by their daughter. Thus, the effect of the gender of children is determined by the relative magnitudes of these two opposing forces.

In order to have a comprehensive understanding of the effect of OCP on migration, we need to estimate the fertility effect and the child gender effect separately. However, a challenge arises in the estimation that both the number and gender of children are endogenous. While, we only have one exogenous variation from OCP which affects fertility and child gender at the same time, as OCP leads to not only the decline of the birth rate from the 1980s (García, 2017) but also a distorted sex ratio due to the strong preference for boys (See Figures 1 and 2).[7] This is the main identification difficulty in all the OCP-related studies of China. For current literature, all of them only consider the fertility effect and ignore the gender effect when using OCP as a natural experiment.[8]

In this study, we propose a new semi-parametric method in a heterogeneous treatment effect framework to address the challenge and estimate the fertility effect and child gender effect. Our estimation results show that the addition of one girl will lead to a 13.7% (1.33 percentage points) increase in the probability that both parents migrate, while the addition of one boy will result in a 24.3% (2.36 percentage points) increase in this probability. The results imply that having more children can motivate parents to migrate (demand for income dominates). Moreover, there exist heterogeneous treatment effects concerning child gender. Having a boy brings a larger effect than having a girl on promoting parental rural-to-urban migration.

The fertility effect estimated using our method is of opposite sign to the one

[6] Extensive research has demonstrated that the gender of children could directly influence various family outcomes other than through changing the number of children, such as marriage stability (Dahl and Moretti, 2008), investment in nutrition and childcare (Behrman, 1988; Barcellos, Carvalho, and Lleras-Muney, 2014; Jayachandran and Kuziemko, 2011), parental labor market performance (Lundberg and Rose, 2002; Ichino, Lindström, and Viviano, 2014), and specifically, parental migration (Li and Yi, 2015).

[7] The gender of children is endogenous since parents can implement gender-based abortion in China.

[8] For instance, Zhang (2017); Li and Zhang (2016); Huang (2016). These studies either use OCP as the instrument for the number of children in households, or take the reduced form effect of the OCP as the effect only through the channel of changing the number of children.

found in Zhang (2017), the only study examining the effect of OCP on rural-to-urban migration in China, as far as we know. Zhang (2017) suggests a negative fertility effect on migration, though the size of the effect is moderate. Using the local OCP implementation intensity as an instrument variable (IV) for fertility, however, Zhang (2017)'s method suffers the problem we have just discussed that he does not take child gender into consideration. The effect estimated in Zhang (2017) is a combination of the fertility effect and child gender effect. Specifically, OCP leads to fewer children at home, which decreases parents' migration probability, but at the same time increases the relative number of boys, which increases parents' migration probability. The two effects cancel out with each other, resulting in a moderately positive or even zero overall effect of OCP on migration. Hence, the moderate accelerating effect of OCP on migration in China found in Zhang (2017) actually comes from the distortion of child gender but not from the reduction of the number of children. If mistakenly interpreting the overall effect of OCP as a pure fertility effect, policy-makers may reach to an erroneous conclusion that reducing fertility can promote the process of urbanization, which should actually work the other way around.

The comparison above indicates that researchers should be cautious about using an OCP related variable as a naive IV for fertility and interpreting the estimated effect as a pure fertility effect. However, the literature mainly focuses on the endogeneity of fertility but show less concern about the possible confounder caused by child gender. The two frequently-used IVs for fertility, the randomness of the gender of children and the implementation of a fertility control policy, both violate the excludability assumption if child gender matters concerning the outcome variable.[9] Some literature entails the use of the occurrence of twins as an instrument for fertility (Oliveira, 2016; Rosenzweig and Wolpin, 1980; Black, Devereux, and Salvanes, 2005). However, this method is problematic in the Chinese context, as the twin record may be unreliable in Chinese data after the launch of OCP.[10] Consequently,

[9] Angrist and Evans (1998), Conley (2004), and Lee (2003) used the sex sameness of children as an instrument for family size to estimate the effect of childbearing on the parents' labor supply. However, as criticized by Rosenzweig and Wolpin (2000), the gender of children may have a direct effect on the outcome variable other than through changing fertility. Using local enforcement intensity of OCP as an IV for fertility, researchers examine the fertility effect on various family outcomes in China, such as children's educational attainment (Li and Zhang, 2016; Qian, 2009), children's economic outcomes in their later lives (Huang, 2016), and parents' labor participation (Zhang, 2017).

[10] As discussed by Huang, Lei, and Zhao (2016), the problem of man-made twins is severe due to OCP. Man-made twins are not real twins. Parents report regularly spaced children as twins to avoid policy violation punishment. Estimated by Huang, Lei, and Zhao (2016), OCP accounts for more than one-third of the increase in the reported births of twins since the 1970s.

there is no off-the-shelf method can be utilized in our context.

To address the endogeneity of fertility and child gender and separately identify the two effects, we propose a semi-parametric model, incorporating the fertility control policy as a natural experiment and a specific matching strategy. To our knowledge, our paper is the first one addressing this problem explicitly. Under OCP in China, families in rural areas whose first child is a girl are usually allowed to have a second child, whereas families whose first child is a boy are not. We call this kind of relaxation as the 1.5 Child Policy. Therefore, the group of rural households who have a first-born girl is randomly assigned to a group allowing them to have a second birth. Based on Angrist, Imbens, and Rubin (1996), we can divide the households in our sample into three behavioral types, namely, always-takers, never-takers, and compliers. Always-takers (Never-takers) are those people who will always (never) choose to give birth to a second child regardless of the gender of the first child. Compliers will choose to give birth to a second child if the first child is a girl and will stop with one child if the first child is a boy. To deal with the problem of two endogenous variables, we impose two exogeneity-related assumptions. First, we assume that the first child gender is random.[11] Second, we assume that conditional on some observables and the behavioral type, the gender selection process is exogenous. This is a weakened version of the standard "selection on observables" assumption.[12]

Under these two assumptions, we identify the local average treatment effect (LATE) of an additional girl on parents' migration decision as the difference between the expectation of parents' potential migration decisions of the compliers who have only a first-born boy and the compliers who have a first-born girl and a second-born boy.[13] Moreover, we can identify the treatment effects of an additional boy on parents' migration decisions for always-takers, compliers, and never-takers by comparing households analogously, with different numbers of boys and children.[14] Consequently, the main task is to identify these expectations of potential migration decisions for different groups with certain numbers of children and boys. We

[11] As the previous literature has claimed, the gender selection of babies happens only at higher order births but not at the first birth (Ebenstein, 2010; Li and Yi, 2015; Chen, Li, and Meng, 2013). We will discuss the reliability of the assumption in more detail later.

[12] In this paper, the "selection on observables" assumption is imposed within fertility types (always-takers, never-takers, and compliers) for gender selection, which means that we not only control for observed variables but also implicitly control for fertility behavioral types.

[13] We need to assume that the birth order of the boy and the girl does not matter.

[14] For instance, the effect of an additional boy for compliers can be identified as the difference between the expectation of parents' potential migration decisions of the compliers who have a first-born girl and a second-born boy, and those compliers who have a first-born girl and a second-born girl.

boil down the problem to a system of linear equations with these expectations as unknowns and solve this system to derive the corresponding treatment effects.[15]

Our study makes three major contributions to the literature. First, we provide the first attempt to examine how fertility control policy affects urbanization process in China and separately estimate the fertility effect and child gender effect.[16] We find that reducing fertility hinders the urbanization process, and the usage of the off-the-shelf IV method may result in a wrong policy implication. Second, we propose a new method which can solve the problem of two endogenous variables—the number and gender of children—with only one exogenous variation from OCP. As far as we know, there is no study in the literature discussing this problem before. Our model can be generalized to the studies on the effect of the number and gender of children on various outcomes under OCP, such as parental labor participation and family stability. Third, our model extends the prevailing constant effects framework in OCP studies to a semi-parametric model which allows for heterogeneous treatment effects.

The remainder of the paper is organized as follows. Section 2 describes China's unique "Hukou" registration system, OCP, and data. Section 3 presents a simple off-the-shelf IV strategy. Section 4 presents our new identification strategy. Section 5 describes our estimation method. Section 6 reports the empirical results. Section 7 presents the robustness checks. Section 8 concludes.

2 Background and Data

2.1 "Hukou" System and China's Population Policy

The definition of migration for Chinese families is different from its counterparts in western countries owing to a special household registration system called "Hukou". In China, all households must register in this national-wide household registration system, according to their place of origin. There are two categories of Hukou, rural and urban. The social welfare systems, such as the education system and the public medical insurance system, are separated for people with different Hukou types. In this respect, sending children to normal schools in cities is very difficult for a migrant

[15] Details of how to construct the system of linear equations will be shown in Section 4.

[16] To the best of our knowledge, the only two papers that examine the effect of the number of children on the migration decision are Oliveira (2016) and Zhang (2017). The only paper investigating the effect of the gender of children on the migration decision is Li and Yi (2015). Nonetheless, both Zhang (2017) and Li and Yi (2015) suffer from the identification problem that we mentioned, that is, both the number and gender of children are endogenous.

worker with a rural Hukou. Thus, many rural migrant workers are forced to leave their children at home in the care of their grandparents if the parents opt to work in cities. In China, people call it "Left-behind children problem". In this situation, children can be an important factor affecting rural parents' migration decisions. We only consider rural-to-urban migration in this paper.

China initiated the OCP in 1979. At the initial stage, almost all families in China were allowed to have only one child, and the second or higher-parity births were penalized.[17] During the mid-1980s, a few provinces in China slightly relaxed the OCP. Several special situations were exempted from punishment for the second birth. One of the critical and prevailing exemptions was that, families in rural areas with a first-born girl were allowed to give birth to a second child (if the first child is a boy, then the second birth will still incur punishment from the government). The exemption was a response and a compromise to the protest of peasant families who needed male labor forces to farm the land. This policy was named "1.5 Child Policy". Afterward, the fertility policy remained relatively stable and was rigorously implemented across the country until recent years.

The "1.5 Child Policy" is implemented in the majority of the rural areas in China. Of the 31 provinces in mainland China, 25 offer this exemption to at least some of the rural population in their provinces, and 19 offer this exemption to all the rural population in their provinces. Table 1 lists the provinces and corresponding exemptions. The 19 provinces that offer the exemption to all the rural people account for 71.3% of the total GDP and 74.9% of the total population in mainland China. Among the remaining six provinces, five directly allow a second birth to all rural residents. They are considered to actually implement a Two Child Policy. Figure 4 depicts the geographic distribution of provinces with different policies. The figure clearly shows that this exemption covers the majority of the populated and economically active parts of the country.[18]

2.2 Data

The data we use in this paper are taken from the 0.4% sample of the 2010 Chinese Population Census. Based on a household-level survey, the census provides rich

[17] To guarantee the efficiency of the implementation, fertility control achievement was put into the list of local officers' evaluation sheet, which would be an important reference for the promotion. The punishment for the violation includes a pecuniary fine, a reduction in social security, and in some extreme cases, forced abortion and sterilization. For more details, see Zhang (2017).

[18] The details of provincial-level policy differences are derived from Gu et al. (2007).

information on the surveyed households and each individual in the household. At the household level, the census provides basic demographic information, such as address, number of people registered in the household, and living conditions. At the individual level, the census contains information on each individual's age, sex, region of residence, education, relationship to household head, and working status.

The census data fits our analysis well.[19] First, we can identify rural-to-urban migration behavior in a household. All people who have a Hukou registered in the community will be surveyed at that community, even if he/she is currently living elsewhere (migrating to other places). Thus, all family members, including migrants and non-migrants, will be covered in the survey. Second, we can identify child information and mother's fertility history in a household using the data. As no question in the census directly links parents to children, we infer the parent-child relationship through a relationship identifier. Specifically, the parents are labeled as "household head" and "spouse of the household head", and the children are labeled as "child". Thus, we can obtain the number of children currently living in the household and detailed information for each child currently living in the household. However, we lack personal information for children who are currently not living in the parents' household. Fortunately, women aged between 15 and 64 are required to report their number of births and number of surviving children in the survey. Together with the child information in a household, we can obtain a full track of the mothers' fertility history and child information for families in which the number of children currently living in the household equals the number of children to whom the mother has given birth. For families who have incomplete child information, the grown-up children may have likely left their original families. However, our target is to assess the effect of juvenile children on the parents' decision, not the effect of those children who can make money by themselves. Therefore, the current sample fits our analysis. Third, the census data contain an extensive range of socioeconomic information at the household and individual levels, which enable us to utilize the machine learning method in our empirical strategy.[20]

To identify the fertility effect and gender effect of children on the rural-to-urban migration behavior of parents, we first restrict our sample to households that consist of a household head and his/her spouse who are both holding rural Hukous.

[19] For more detail of the survey, please refer to http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/html/fu07.htm

[20] The use of the machine learning method with regularization may introduce some difficulties in inference or even inconsistency. As a result, we also implement some bound estimations to make sure our result is robust.

Second, we confine the sample to households in provinces that offer 1.5 Child Policy to all rural citizens to utilize the exogenous variation of fertility induced by the policy. Third, we retain only the households with the household head and his/her spouse aged above 35—to select those who have completed their fertility, and below 64—to guarantee that both parents remain in the labor force. Fourth, we confine our sample to households in which all children are less than or equal to 18 years old. The existence of adult children is not imperative to our research because a grown-up child can independently generate income and support their parents. This condition will have a largely different implication in relation to the effect on parents' migration decision compared with that of juvenile children. Fifth, we further restrict our sample to households in which the number of children currently living in the household is equal to the number of born children reported by the mother to obtain an accurate account of children number and birth order.[21] Thus, we can avoid unnecessary confusion caused by children who died or left the household. Sixth, we exclude households in which at least one of the parents has been divorced, and those with twins. Seventh, as the population policy is only applicable to Han Chinese but not to minorities, we further restrict our sample to Han Chinese.[22] After all these data filtration steps, we obtain a sample of 64,095 households.

We use three dummy variables to measure parents' rural-to-urban migration behavior. The first dummy equals 1 if both parents migrate to urban areas, otherwise zero. The second dummy equals 1 if the male parent migrates. The third dummy equals 1 if the female parent migrates. Table 2 shows the summary statistics of the main variables, where the first child's gender equals 1 if it is a girl.

3 Naive Identification of the Effect of Number of Children

Our study mainly aims to identify the effects of the number and the gender of children on the migration decisions of parents. First of all, in this section, we consider a traditional IV estimation method to examine the effect of the number of children, similar to the method utilized in previous papers. We will show that the

[21] In order to examine if the fourth and fifth restrictions lead to a biased sample, we conduct a robustness check with a sample in which both parents are aged between 35 and 40 years old. In this sample, households in which all children are less than or equal to 18 years old account for more than 95% of all the households; households in which the number of children living in the household is equal to the number of children even born account for 97.5% of all the households in the sample. Therefore, these two restrictions should have a minor effect on the estimates of using this robustness-check sample. In this setting our main conclusions still hold. The results of the robustness check are available upon request.

[22] According to the census data in 2010, Han people consists of 91.5% of the whole population.

method does not work in the situation when the effect of the child gender cannot be ignored.

3.1 Simple IV Identification

We first assume that the gender of children does not affect parents' migration decision. Consider an additive constant-effects model for migration:

$$Mig_{il} = \beta_0 + \beta_1 c_{il} + \boldsymbol{X_{il}}\beta_2 + u_{il}, \tag{1}$$

where Mig_{il} is a dummy variable of whether the parents in household i with a home location l choose to migrate, which can be both parents migrating, father migrating, or mother migrating. c_{il} is the number of children in household i with home location l. $\boldsymbol{X_{il}}$ is a vector of observed controls including first child's age, number of preschool children, marriage time, parents' education, parents' age, whether parents are literate and hukou location dummies. u_{il} is the unobserved term.

To identify β_1, we need to address the endogeneity of fertility, because fertility choices can be correlated with u_{il}. For example, migration decision may be based on the attitude towards life and family, which may simultaneously affect fertility decision negatively or positively. For instance, in a traditional family, people will be more willing to have more children, and meanwhile, they may be less willing to move outside to cities due to their conservative attitudes. As a result, if we omit any of these unobserved factors in our analysis, then the OLS regression will be biased. Thus, we should employ an IV identification method.

The 1.5 Child Policy in rural areas enables us to use the first-born child's gender as an IV for fertility in the household. Families with a first-born girl will receive a policy shock to fertility, which will not be received by families with a first-born boy. We denote z_{il} as a dummy that equals to 0 if the gender of the first child in household i is male and equals to 1 if it is female. We use z_{il} as the instrumental variable for c_{il}.

An important assumption for the validity of this IV is no gender selection at the first birth. We claim that the assumption holds when considering rural families in provinces with 1.5 Child Policy. That is, the incentive to give up the first-born girl is almost null because parents are free to have a second chance. Moreover, previous research failed to find evidence for gender-selection at the first birth in

China. The main gender selection problem occurs at the second or even subsequent births (Ebenstein, 2010; Li and Yi, 2015; Chen, Li, and Meng, 2013). We check the gender ratio for the first child in our sample, and find 32,386 first-born boys and 31,727 first-born girls. The sex ratio, as calculated based on these numbers, is 102.02 (the number of boys for every 100 girls), which lies in the normal range of the child sex ratio for Caucasian and Asian populations aged 0-14 (Banister, 2004).[23]

3.2 Simple IV Results

Table 3 shows the first stage relationship for the IV estimation. In all specifications, the correlation between having a first-born girl and the number of children in a household is both statistically and economically significant. Having a first-born girl is related to a 0.418 increase in the number of children, compared with having a first-born boy. It means that approximately 42% of the families who were allowed to give birth to a second child, actually took advantage of the exemption. This evidence supports our statement that many rural people will utilize the extra quota provided by the 1.5 Child Policy and give birth to a second child. Consequently, we obtain a strong first stage for the IV estimation.

Table 4 shows the results of the IV estimation for a linear probability model. The dependent variables are both parents migrating, father migrating, and mother migrating. The IV estimates show that the effect of the number of children on parents' migration decision is small and statistically insignificant. The signs of the effect differ when we use various measures. For father migration, the sign is positive; whereas the sign for mother migration is negative. The causal link between fertility and migration decision seems weak according to the IV estimates.

3.3 Problem with Simple IV

Nevertheless, we should be cautious in using the IV identification strategy. As discussed, the gender of children may influence parents' migration decision. Notably, the IV we select, the gender of the first child, will not only change the number of children in the household because of the 1.5 Child Policy but also mechanically change the gender of children in the household. Therefore, if the gender of children directly affects parents' migration decision, then the excludability assumption is

[23] Naturally, there will be slightly more male infants, and the sex ratio will converge to 100 as time goes by since females generally have a higher life expectancy.

invalidated when the first child's gender is used as an IV for the number of children. As a result, the reduced-form estimate of the fertility effect is likely to be biased by the child gender effect.

As discussed in the literature review, this is not a specific misusing of the first child gender as the IV for the number of children, but a general trap for any policy-induced or gender-related IV method. When estimating the reduced-form fertility effect instrumented by the fertility policy, previous studies also use policy variables such as fine rates across provinces. However, these policy variables have two implications: a more restrictive fertility control policy will result in a lower fertility rate and a higher sex ratio imbalance (Ebenstein, 2010). A combination of these two effects cannot be interpreted as a pure fertility effect if the effect of the child gender is not ignorable. Attempting to fix this problem, we can try a simple revision of the IV regression by linearly controlling for the number of boys. Nevertheless, as the number of boys itself is endogenous, the estimation remains contaminated. The results are shown in Appendix A.

4 Identification Strategy

In this section, we discuss how to tackle the problem in the simple IV method and derive a semi-parametric identification strategy with the effect of the child gender taken into consideration. We extend the constant effects IV model in two ways. First, we allow for the violations of the exclusion restriction in the first-born child gender instrument. Second, we allow for heterogeneous causal effects. Our model is an extension of Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996)'s idea of treatment effects by behavioral types, which is inspired by the potential outcome framework of Rubin (Rubin, 1974). The treatment we employ is China's 1.5 Child Policy. Under this fertility policy, the gender of the first child can be considered random, and the treatment on the second child enables us to classify households into three behavior types, namely, always-takers, compliers, and never-takers. We aim to explicitly identify the LATEs for households in certain behavior types. Specifically, we aim to identify the LATE of an additional child on parents' potential migration decision, which we define as the effect of the number of children. The LATEs of an additional boy on parent's migration decision for always-takers, compliers, and never-takers, are defined as the effects of the gender of children. Our method proceeds as follows. First, we construct the birth tree of children, denoting each node in the tree as the households with a specific number of children and number of boys. Second, because different nodes are composed of different behavioral types of households and the conditional expectations of the potential migration decisions for the households at each node are known from the data, we expand the conditional expectations of the potential migration decisions at each node by behavioral types, resulting in a system of equations. Third, using a combination of the randomness of gender of first child, a weakened gender "selection on observables" (Assumption 3), and some other mild assumptions, we solve this system of equations and back up the conditional expectations of the potential migration decisions for different behavioral types at each node. Then, we can obtain the LATEs of interest. To simplify the model and render it tractable, we only consider households with less than three children, which accounts for approximately 89% of the whole sample. Initially, we present our model setup and apply it to a simple case without gender selection, and then extend it to a general situation when gender selection at the second birth is permitted.

4.1 Model Setup

We consider a general migration decision equation as follows:[24]

$$Mig_i = f(c_i, b_i, X_i, u_i); \tag{2}$$

and we have a first-stage fertility equation as follows:

$$c_i = g(z_i, X_i, e_i). \tag{3}$$

e_i is the unobserved term in the fertility decision equation, such as unobserved fertility preference and ability to deal with government punishment, and so on. z_i is the gender of the first child which equals 1 if it is a boy. b_i is the number of boys in household i. The definition of other variables remains the same as in the simple IV estimation.

Based on this model setup, we can explicitly explain the failure of the simple IV identification. First, the unobserved term in migration decision u_i and the unob-

[24] In this model, we implicitly assume that the birth order of the children does not affect migration and only the number of children and number of boys matter. For example, the combination of a first boy and a second girl, it has the same effect on the parents' migration compared with the combination of a first girl and a second boy.

served term in fertility decision e_i may be correlated with each other, which is the source of endogeneity of the number of children c_i. As there is no gender selection at the first birth, u_i and e_i are independent of the first child gender z_i. Hence, previous studies use z_i as an IV for c_i. However, b_i is correlated with z_i and Mig_i, so the traditional IV estimate is biased if b_i is omitted from the regression. Moreover, this problem cannot be addressed by simply controlling for b_i, as b_i itself is endogenous and z_i is correlated with u_i conditional on b_i. The endogeneity of b_i comes from two sources. On the one hand, there may be gender selection on children. On the other hand, b_i is a post-determined variable.[25] We should come up with a new strategy to solve this problem.

Following Angrist, Imbens, and Rubin (1996), we define three behavioral types in our model: always-takers, those who always choose to give birth to the second child regardless of the gender of their first child, i.e., $c_i(z_i = 1) = c_i(z_i = 0) = 2$; never-takers, those who always choose not to give birth to the second child regardless of the gender of their first child, i.e., $c_i(z_i = 1) = c_i(z_i = 0) = 1$; compliers, those who choose to give birth to the second child if the first is a girl but do not if the first is a boy, i.e., $c_i(z_i = 1) = 2, c_i(z_i = 0) = 1$.[26] The household's fertility preference (observed factors in X_i and unobserved factors e_i) and social capital (e.g., ability to bear government punishment) will determine its type.

As Pinto (2015) points out, the number of possible response types will grow exponentially if we expand the birth tree, which will prevent us from a meaningful identification of treatment effects for various types. Therefore, we decide to drop all households with more than two children. Another reason for us to do this is that families giving birth to more than two children are highly likely to be qualified for other local specific exemptions or even have illegal channels to avoid punishment. They may behave differently from the always-takers and compliers as defined in our setting.

For convenience, we denote Mig_{icb} as the potential migration decision for household i when $c_i = c$ and $b_i = b$. That is, $Mig_{icb} = f(c_i = c, b_i = b, X_i, u_i)$. It refers to the migration decision that the household would make if they had c children and b boys, regardless of how many children/boys they have in reality. For example, Mig_{i10} represents the potential migration decision for household i, if they

[25] The detailed explanation for why b_i is post-determined in the model we proposed is shown in Appendix B.

[26] We assume that there is no defier in our model.

had one child and zero boy. Moreover, we denote A, N, and C as always-takers, never-takers, and compliers, respectively, and denote n_A, n_C, and n_N as the number of always-takers, compliers, and never-takers in the population, respectively.

Our model aims to identify the LATEs of the number of children and the gender of children. First, we attempt to identify the treatment effect of an additional child on the migration decision of parents. Specifically, we take an addtional girl as the baseline of the effect of number of children. Thus, the LATE of the number of children, can be defined as the difference between the potential outcome of for households with two children and one boy, and households with one child and one boy:

$$FLATE = E[Mig_{i21}|C] - E[Mig_{i11}|C]. \tag{4}$$

As only compliers will change their fertility choice when being treated, the effect of the number of children can be identified only for compliers. Second, we attempt to identify the treatment effect of an additional boy on the migration decision of parents, which is the LATE of child gender denoted as $BLATE$. We define the child gender effect for always-takers, compliers, and never-takers, as follows:

$$BLATE_{A1} = E[Mig_{i22}|A] - E[Mig_{i21}|A], \tag{5}$$

$$BLATE_{A2} = E[Mig_{i21}|A] - E[Mig_{i20}|A], \tag{6}$$

$$BLATE_{C} = E[Mig_{i21}|C] - E[Mig_{i20}|C], \tag{7}$$

$$BLATE_{N} = E[Mig_{i11}|N] - E[Mig_{i10}|N]. \tag{8}$$

We can also derive an average $BLATE$ by integrating all the $BLATE$ over types using the number of households in each type as weights.

4.2 Identification without Gender Selection

In this subsection, we discuss how to separately identify the two effects when there is no gender selection at any birth order. Then we extend the model to a case with gender selection in Section 4.3.

Figure 3 shows that, a set of values for vector (z, b, c) will determine a node on the birth tree. In other words, a specific node on the tree refers to a group of households with the same number of children, number of boys, and gender of the first child. When there is no gender selection at any birth order, and households are randomly assigned with a boy or a girl with the probability of $\frac{1}{2}$ at each birth, the distributions of different types of people are shown in Figure 3 and Table 5. Taking households with a first-born boy as an example, never-takers and compliers will stop giving birth to the second child, but always-takers will continue to give birth to the second child. Always-takers with a first-born boy are randomly assigned a second-born boy or a second-born girl. The same logic can be applied to households with a first-born girl.

Thus, we can obtain Equation (9) according to the number of households in different cells in Table 5.

$$\begin{cases} n(z=1, b=0, c=1) = \frac{1}{2}n_N \\ n(z=0, b=1, c=1) = \frac{1}{2}n_N + \frac{1}{2}n_C \\ n(z=0, b=1, c=2) = \frac{1}{4}n_A \end{cases} \tag{9}$$

For Equation 9, the left-hand side pertains to the number of households at each node, which is a summation of the number of households in different types at each node. In this case, the number of households in each type is overidentified. Thus, we are able to use the information of some nodes, although not all of them, to identify n_N, n_A, and n_C.[27] Then, we can further estimate the probability distribution of any specific type of households at each node, i.e., $P(A|z_i, b_i, c_i)$, $P(N|z_i, b_i, c_i)$ and $P(C|z_i, b_i, c_i)$.

Owing to the randomness of child gender at each birth, each specific type of households is identically distributed at different nodes on the birth tree. That is, the distributions of X_i and u_i are identical across all nodes on the birth tree conditional on types. Therefore, we propose that, conditional on any specific type of household, the expectation of the potential migration decision remains the same across the gender of the first child z_i, the number of boys b_i, and the number of children c_i, i.e., $E[Mig_{icb}|z_i, b_i, c_i, Type_i] = E[Mig_{icb}|Type_i]$. For instance, $E[Mig_{i21}|z_i = 1, b_i = 1, c_i = 2, A] = E[Mig_{i21}|z_i = 1, b_i = 0, c_i = 2, A] = E[Mig_{i21}|A]$.

Then, we can derive the expectation of potential migration decisions at each node of the tree as follows:

[27] In our data set, the differences of the estimated n_N, n_A and n_C across different choices of nodes are negligible.

$$
\begin{aligned}
E[Mig_i|z_i = 0, b_i = 1, c_i = 2] =& E[Mig_{i21}|A] && (10)\\
E[Mig_i|z_i = 0, b_i = 2, c_i = 2] =& E[Mig_{i22}|A] && (11)\\
E[Mig_i|z_i = 0, b_i = 1, c_i = 1] =& P(N|z_i = 0, b_i = 1, c_i = 1)E[Mig_{i11}|N] && (12)\\
&+P(C|z_i = 0, b_i = 1, c_i = 1)E[Mig_{i11}|C] &&\\
E[Mig_i|z_i = 1, b_i = 0, c_i = 1] =& E[Mig_{i10}|N] && (13)\\
E[Mig_i|z_i = 1, b_i = 0, c_i = 2] =& P(A|z_i = 1, b_i = 0, c_i = 2)E[Mig_{i20}|A] && (14)\\
&+P(C|z_i = 1, b_i = 0, c_i = 2)E[Mig_{i20}|C] &&\\
E[Mig_i|z_i = 1, b_i = 1, c_i = 2] =& P(A|z_i = 1, b_i = 1, c_i = 2)E[Mig_{i21}|A] && (15)\\
&+P(C|z_i = 1, b_i = 1, c_i = 2)E[Mig_{i21}|C] &&
\end{aligned}
$$

The conditional expectation at each node is a weighted sum of the conditional expectations of potential outcomes for various types of households at that node. All expectations on the left-hand side can be derived from the data, and all conditional probabilities on the right-hand side can be identified using the number of households in each type.

Therefore, we can identify our targeted $FLATE$ and $BLATE$ by solving the system of Equations (10)-(15). We should make two assumptions to solve the system.

Assumption 1 *For the always-taker group, the expectation of the boy's partial effect on migration decision remains the same for from no boy to one boy, and from one boy to two boys. That is,* $E[Mig_{i22}|A] - E[Mig_{i21}|A] = E[Mig_{i21}|A] - E[Mig_{i20}|A]$.

This is a mild assumption, because it only postulates that b_i enters into the migration decision function linearly for always-takers. Hence, $BLATE_{A_1} = BLATE_{A_2}$.

Assumption 2 *The expectation of the boy's partial effect on migration decision remains the same across the never-taker and always-taker groups. That is,* $E[Mig_{11}|N] - E[Mig_{10}|N] = E[Mig_{22}|A] - E[Mig_{21}|A]$.

The economic intuition of this assumption is that always-takers and never-takers have the same attitude toward boys in relation to migration decision. This assumption is reasonable because the difference between never-takers and always-takers likely lies on fertility preference than on boy preference, which means that they have a similar preference to boys. Therefore, boys are likely to have similar effects on the migration decision of the parents.[28]

To simplify the notation, we denote the conditional expectation of migration decision at each node (z,b,c), $E[Mig_i|z_i = z, b_i = b, c_i = c]$ as y_{zbc}. After solving the system of Equations (10)-(15), we can identify all the aforementioned treatment effects with the following propositions.[29]

Proposition 1 *If Assumption 1 and Assumption 2 hold, we can identify the Local Average Treatment Effect of an additional child as:*

$$FLATE = \frac{(n_A + n_C)y_{112} - (n_A + n_N)y_{012} - (n_N + n_C)y_{011} + n_N(y_{022} + y_{101})}{n_C} \tag{16}$$

Proposition 2 *We can identify the Local Average Treatment Effect of an additional boy, holding number of children constant, for different types of people as follows:*
(1) $BLATE_A = y_{022} - y_{012}$ *for always-takers*
(2) If Assumption 1 holds,
$BLATE_C = \frac{(n_A+n_C)(y_{112}-y_{102})-n_A(y_{022}-y_{012})}{n_C}$ *for compliers*
(3) If both Assumption 1 and Assumption 2 hold, $BLATE_N = BLATE_A = y_{022} - y_{012}$ *for never-takers.*

Propositions 1 and 2 illustrate that the treatment effects are identified as the average values of the conditional expectations of the potential migration decisions at the nodes weighted by the distribution (number of households) of types across the population.

4.3 Identification with Gender Selection

4.3.1 Setup with Gender Selection

The previous subsection shows the identification of treatment effects without considering gender selection at both the first and the second births. However, this is not the case in China, where according to empirical evidence, gender-based abortion is a severe problem at the second or higher-order births. In this section, we construct a

[28]One of the special cases satisfying Assumptions 1 and 2 is that the migration decision is a linear function regarding the number of boys. However, it is not necessary to impose such a restrictive condition, although it is a prevailing underlying assumption in most parametric research. We can allow for a heterogeneous effect of the gender of children for the complier group and the always-taker/never-taker group.
[29]The proof is shown in Appendix C.

general identification strategy based on the previous one but taking gender selection at the second birth into account.

At the first birth, all types of household do not implement gender selection, and the probability of having a first-born boy is equal to P_0, a natural rate of birth gender ratio which can be derived from the data. At the second birth, we assume that compliers and always-takers choose to carry out gender selection because of gender preference. However, the selection process is imperfect due to the limited access to detection/selection technology or government supervision. The gender of the second birth is determined by the following function:

$$\mathbf{1}(Boy)_i = h(X_i, \eta_i), \tag{17}$$

where the indicator function on the left-hand side equals 1 if the second birth of household i is a boy. On the right-hand side, η is a vector of unobserved factors, such as the luck of not being spotted by the government or the willingness to bear the abortion cost.

We call the realized distorted gender ratio at birth as the "success rate", which can vary across types. We define the average probability of getting a second-born boy for always-takers as P_{A_0} when $z = 0$ and P_{A_1} when $z = 1$, and the average probability of getting a second-born boy for compliers as P_C. The three probabilities remains unknown because we cannot observe the illegal gender selection process.

The distribution of different types of households is shown in Figure 5 and Table 6. At the first birth, the assignment of birth gender remains random with natural rate P_0. Therefore, a proportion of $1-P_0$ always-takers, never-takers, and compliers have a first-born girl. Never-takers comply with one girl, while always-takers and compliers choose to give birth to a second child. At the second birth, the existence of gender selection distorts the gender ratio. Hence, $(1-P_0)P_{A_1}$ of always-takers have a first girl and a second boy, and the remaining $(1-P_0)(1-P_{A_1})$ of always-takers have a first-born girl and a second-born girl. Similarly, $(1-P_0)P_C$ of compliers have a first-born girl and a second-born boy, and the remaining $(1-P_0)(1-P_C)$ of compliers have a first-born girl and a second-born girl. The same logic can be applied to the parts with a first-born boy.

Then, we can obtain Equation (18) according to the number of households in different cells in Table 6 with gender selection at the second birth.

$$
\begin{cases}
n(z=1,b=0,c=1) = (1-P_0)n_N \\
n(z=0,b=1,c=1) = P_0 n_N + P_0 n_C \\
n(z=0,b=1,c=2) = P_0(1-P_{A_0})n_A \\
n(z=0,b=2,c=2) = P_0 P_{A_0} n_A \\
n(z=1,b=0,c=2) = (1-P_0)(1-P_{A_1})n_A + (1-P_0)(1-P_C)n_C \\
n(z=1,b=1,c=2) = (1-P_0)P_{A_1}n_A + (1-P_0)P_C n_C
\end{cases}
\tag{18}
$$

4.3.2 Assumption for Gender Selection

Owing to the endogeneity of the gender selection process, the distributions of X_i and u_i are no longer identical across all nodes on the birth tree conditional on a specific type. Therefore, we do not have $E[Mig_{icb}|z_i, b_i, c_i, Type_i] = E[Mig_{icb}|Type_i]$ at nodes with gender selection. For example, $E[Mig_{i21}|z_i = 1, b_i = 1, c_i = 2, A] \neq E[Mig_{i21}|A]$, as always-takers with a higher boy preference are more likely to sort into the node with a first-born girl and a second boy ($z_i = 1, b_i = 1, c_i = 2$). Consequently, the conditional expectations of the potential migration decisions for certain type of households at certain node is $E[Mig_{icb}|z_i, b_i, c_i, Type_i]$, rather than $E[Mig_{icb}|Type_i]$.

To solve this problem, we assume that there is no further selection for success rate within types after conditioning on observed controls, X. After controlling for X, compliers/always-takers who successfully have a boy and compliers/always-takers who do not are not systematically different in the unobserved term in the migration function (u_i). Specifically, the unobserved factors in the gender selection equation η_i are independent of those in the migration equation u_i within types. Then we have $E[Mig_{icb}|z_i = z, b_i = b, c_i = c, A, x_i] = E[Mig_{icb}|A, x_i]$ and $E[Mig_{icb}|z_i = z, b_i = b, c_i = c, C, x_i] = E[Mig_{icb}|C, x_i]$. This is a crucial assumption, which is similar to the "selection on observables" assumption in the common matching method but in a weakened version because we also control for fertility behavioral type. Section 7 features a discussion of the possible violation of this assumption and its implication to our conclusion. Moreover, the data clearly show that the sex ratio of the second child for households with a first-born boy is very close to 1:1 (95.3:100). To simplify the model, we can safely assume that gender selection does not occur for always-takers at the second birth when the first child is a boy.

Assumption 3 *always-takers do not implement gender selection when they have a first-born boy, that is, $P_{A_0} = P_0$. Compliers and always-takers may implement gender-based abortion at the second birth if their first child is a girl. However after controlling for observables, conditional on types, their sucess of gender selection depends on exogenous factors (to migration decision) that are independent of u_i, i.e., $\eta_i \perp u_i|A$ and $\eta_i \perp u_i|C$.*

This assumption is actually not as strong as one may think. The "selection on observables" is assumed within the fertility type group, which means that, to a certain extent, we simultaneously and implicitly control for the unobserved characteristics of households in relation to fertility choices. Assumption 3 implies that the unobserved term of gender selection η_i and the unobserved term of migration u_i are independent after controlling for both observed X and some unobserved characteristics related to fertility decision. We will discuss the advantage of using this assumption in our method, compared with the assumptions we have to make in the traditional IV method in section 7.1.

We then obtain a similar system of equations about conditional expectations as in the case without gender selection for equations (10)-(15). We expand the conditional expectations of the potential migration decisions at each node (left-hand side) in terms of the conditional expectations of the potential migration decisions for different types at that node (right-hand side) weighted by the probabilities of the types at that node.

$$
\begin{aligned}
E[Mig_i|z_i=0,b_i=1,c_i=2,X] =& E[Mig_{i21}|A,X] && (19)\\
E[Mig_i|z_i=0,b_i=2,c_i=2,X] =& E[Mig_{i22}|A,X] && (20)\\
E[Mig_i|z_i=0,b_i=1,c_i=1,X] =& \frac{n_{Nx}}{n_{Nx}+n_{Cx}}E[Mig_{i11}|N,X] && (21)\\
&+\frac{n_{Cx}}{n_{Nx}+n_{Cx}}E[Mig_{i11}|C,X]\\
E[Mig_i|z_i=1,b_i=0,c_i=1,X] =& E[Mig_{i10}|N,X] && (22)\\
E[Mig_i|z_i=1,b_i=0,c_i=2,X] =& \frac{(1-P_{A_1x})n_A}{(1-P_{A_1x})n_{Ax}+(1-P_{Cx})n_{Cx}}E[Mig_{i20}|A,X] && (23)\\
&+\frac{(1-P_{Cx})n_{Cx}}{(1-P_{A_1x})n_{Ax}+(1-P_{Cx})n_C}E[Mig_{i20}|C,X]\\
E[Mig_i|z_i=1,b_i=1,c_i=2,X] =& \frac{P_{A_1}n_{Ax}}{P_{A_{1x}}n_{Ax}+P_{Cx}n_{Cx}}E[Mig_{i21}|A,X] && (24)\\
&+\frac{P_{Cx}n_{Cx}}{P_{A_1x}n_{Ax}+P_{Cx}n_{Cx}}E[Mig_{i21}|C,X]
\end{aligned}
$$

where all probabilities and numbers of households are conditional on X. Subscript x stands for the variable conditional on x.[30]

4.3.3 Assumption for Identifying Gender Selection Probability

Another difficulty in identification when we incorporate gender selection is how to identify P_{A_1x} and P_{Cx}, the distorted gender ratio at birth for always-takers and compliers, respectively. To solve for the two unknown probabilities, we build a prediction model for the probability of each observation to be an always-taker. We know whether it is an always-taker or not for all the households in the upper part of the birth tree (z=0), and we observe some features of these households.[31] Using the data of these households as the training data set, we can train a prediction model for whether a household is an always-taker conditional on observed predetermined features X_A (e.g., parents' nationality, age, education, etc.) with a machine learning algorithm.[32] Then we use this model to predict the probability of being an always-taker for each household at nodes (z=1,b=1,c=2) and (z=1,b=0,c=2).

[30]For example, $n_{Nx}=n(N,x)$, $n_{Cx}=n(C,x)$ which leads to $\frac{n_{Nx}}{n_{Nx}+n_{Cx}}=P(N|z_i=0,b_i=1,c_i=1,X=x)$.

[31]In machine learning field, people usually use the term "features" or "predictors" to describe the independent variables in the model. We will follow this tradition here.

[32]If we need to predict a certain outcome Y with a set of inputs X_A, we call the data set used to estimate the prediction model as the training dataset. Moreover, X_A is a superset of X which includes all possible predetermined characteristics.

To guarantee that we can obtain a good prediction model and achieve the identification using observed X_A, we need to further make an assumption on X_A.

Assumption 4 *We can observe a set of predetermined features X_A such that b will not provide with more information about being an always-taker after conditioning on X_A for households with a first-born girl and two children, (z=1,c=2). That is, $P(A|X_A, b, c = 2, z = 1) = P(A|X_A, z = 0)$.*

Subsequently, we obtain P_{A_1x} by dividing the sum of the predicted probability as follows:[33]

$$P_{A_1x} = \frac{P(A, b = 1|c = 2, z = 1, X)}{P(A, b = 1|c = 2, z = 1, X) + P(A, b = 0|c = 2, z = 1, X)}. \tag{25}$$

4.3.4 Point Identification under Assumptions 3 and 4

Another additional identification work we need to do in this context is that now we are identifying everything conditional on X, and if we want to get the averages of the treatment effects within types, we still need to identify the distribution of X conditional on types, that is, $P(X|Type)$. $P(X|A)$ and $P(X|N)$ are easy to identify because we have the gender of the first child z_i to be independent of X and types, which leads to that $P(X|A) = P(X|A, z = 1) = P(X|z = 1, c = 2)$ and $P(X|N) = P(X|N, z = 0) = P(X|z = 0, c = 1)$. The intuition of these two equations is that always-takers and never-takers are randomly separated into families with a first-born girl or a first-born boy. All the families at node (z=1,b,c=2) are always-takers; all the families at node (z=0,b,c=1) are never-takers. After $P(X|A)$ and $P(X|N)$ are derived, $P(X|C)$ can be identified as $P(x|C) = \frac{P(X)-P(X|A)P(A)-P(X|N)P(N)}{P(C)}$, where the unconditional type probabilities are $P(A) = \frac{n_A}{n}$, $P(N) = \frac{n_N}{n}$ and $P(C) = \frac{n_C}{n}$.

Thus, we can derive a point identification for each treatment effect presented above under Assumptions 1-4 when gender selection is possible at the second birth.[34]

Proposition 3 *If Assumptions 1, 2, 3 and 4 hold simultaneously, we can identify the LATE of an additional child as:*

[33] The detailed analysis for the identification of P_{A_1} is shown in Appendix D.

[34] The proof is the same with what we did when there was no gender selection on second birth.

$$FLATE = \int_X \frac{(P_{A_1x}n_{Ax} + P_{Cx}n_{Cx})y_{112x} - (P_{A_1x}n_{Ax} + P_{Cx}n_{Nx})y_{012x} - P_{Cx}(n_{Nx} + n_{Cx})y_{011x} + P_{Cx}n_{Nx}(y_{022x} + y_{101x})}{P_{Cx}n_{Cx}} P(X|C)dx \tag{26}$$

Proposition 4 *We can identify the LATE of an additional boy, holding the number of children constant, for different types of people as follows:*
(1) $BLATE_A = \int_X (y_{022x} - y_{012x})P(X|A)dx$ *for always-takers*
(2) If Assumption 1, 3 and 4 hold,

$$BLATE_C = \int_X [\frac{y_{112x}(P_{A_1x}n_{Ax}+P_{Cx}n_{Cx})-P_{A_1x}n_{Ax}y_{012x}}{P_{Cx}n_{Cx}} - \frac{[(1-P_{A_1x})n_{Ax}+(1-P_{Cx})n_{Cx}]y_{102x}-(1-P_{A_1x})n_{Ax}(2y_{012x}-y_{022x})}{(1-P_{Cx})n_{Cx}}]P(X|C)dx$$

for Compliers
(3) If both Assumption 1 and Assumption 2 hold, $BLATE_N = \int_X (y_{022x} - y_{012x})P(X|N)dx$ *for never-takers.*

All treatment effects, probabilities and y_{zbcx} above are conditional on controls X. To derive the average effect, we need to integrate over X with respect to $P(X|type)$.

4.3.5 Set Identification

One problem of Assumption 4 is that we need a very good feature set X_A, which may not be available. Therefore, we can take a step back and turn to set identification. If P_{A_1x} and P_{Cx} do not depend on X, that is, the distorted gender ratio does not vary across X within always-takers or compliers, we can use an upper bound and a lower bound for P_{A_1} and P_C, respectively, to investigate the behavior of our estimation within the bounds. For the effect of the number of children, we rearrange the terms inside Equation (26) and get the $FLATE$ as follows:

$$FLATE = \frac{P_{A_1}}{P_C} \int_X \frac{n_{Ax}(y_{112x} - y_{012x})}{n_{Cx}} P(X|C)dx + k \tag{27}$$

where P_{A_1} and P_C are the distorted gender ratios (success rates), and k is a constant not related to P_{A_1} and P_C. Given that P_{A_1} and P_C are negatively correlated, this equation is a monotonic function of P_{A_1}. As a consequence, the bound for P_{A_1} is just the bound for $FLATE$. We define $P^l_{A_1}$ as the lower bound and $P^u_{A_1}$ as the upper bound. We can use Equation (18) to derive the corresponding bounds for P^u_C and P^l_C.

Proposition 5 *If Assumptions 1, 2 and 3 hold simultaneously and in addition $P_{A_1x} = P_{A_1}$ and $P_{Cx} = P_C$, which do not depend on X, then we can bound $FLATE$ as follows:*

$$FLATE^1 = \int_X \frac{(P^l_{A_1} n_{Ax} + P^u_C n_{Cx}) y_{112x} - (P^l_{A_1} n_{Ax} + P^u_C n_{Nx}) y_{012x} - P^u_C (n_{Nx} + n_{Cx}) y_{011x} + P^u_C n_{Nx} (y_{022x} + y_{101x})}{P^u_C n_{Cx}} P(X|C) dx \tag{28}$$

$$FLATE^2 = \int_X \frac{(P^u_{A_1} n_{Ax} + P^l_C n_{Cx}) y_{112x} - (P^u_{A_1} n_{Ax} + P^l_C n_{Nx}) y_{012x} - P^l_C (n_{Nx} + n_{Cx}) y_{011x} + P^l_C n_{Nx} (y_{022x} + y_{101x})}{P^l_C n_{Cx}} P(X|C) dx \tag{29}$$

$FLATE$ lies between $FLATE^1$ and $FLATE^2$.

However, for $BLATE$, we do not have a good monotonic property for P_{A_1}. That is, the bound of P_{A_1} is not necessarily the bound of $BLATE$. Meanwhile, "P_{A_1x} and P_{Cx} do not depend on X" is a strong assumption. Therefore, to make sure our main conclusion does not vary because of the assumptions we make in the probability here, we check for several different grids to see how $BLATE$ changes with different choices of P_{A_1}. The results for the estimations for different grids are shown in Appendix F. Our main conclusion holds in all situations. We discuss how to choose the bounds reasonably in the estimation in Section 5.

5 Estimation

In this section, we put the identification strategy of the model with gender selection into the application using our data from China. We use the model with gender selection because gender selection at second birth is prevalent in rural China. We first discuss the point estimation and then move on to the bound estimation.

5.1 Point Estimation

As discussed in Section 4, to conduct point estimation of the targeted $LATEs$, we need to estimate the number of households in each type, the conditional expectation of migration decision at each node, and the selection probability of each type conditional on X.

First, given that we know which node on the birth tree each household belongs to, we can estimate the number of households in each type by substituting the number of households in different nodes calculated from the data into Equation (18). Then we obtain the estimation results as follows:

$$\hat{P}_0 = \frac{\sum_i \mathbf{1}(z_i = 0, X = x_i)}{\sum_i \mathbf{1}(z_i = 0, X = x_i) + \sum_i \mathbf{1}(z_i = 1, X = x_i)} \tag{30}$$

$$\hat{n}_N = \frac{1}{1-\hat{P}_0}\hat{n}_{101} = \frac{1}{1-\hat{P}_0}\sum_i \mathbf{1}(z_i = 1, b_i = 0, c_i = 1, X = x_i) \tag{31}$$

$$\begin{aligned}\hat{n}_A =& \frac{1}{\hat{P}_0}(\hat{n}_{012} + \hat{n}_{022}) = \frac{1}{\hat{P}_0}[\sum_i \mathbf{1}(z_i = 0, b_i = 1, c_i = 2, X = x_i) \\ &+ \sum_i \mathbf{1}(z_i = 0, b_i = 2, c_i = 2, X = x_i)]\end{aligned} \tag{32}$$

$$\hat{n}_C = \frac{\hat{n}_{011} - \hat{P}_0\hat{n}_N}{\hat{P}_o} \tag{33}$$

However, the dimension of X is too high and there are too few observations available at each grid $X = x_i$. As a result, we approximate the conditional probabilities directly by unconditional counts.[35]

Second, we can further estimate the conditional expectation of migration decision y_{zbcx} at each node from the data with a partial linear model.[36] A fully nonparametric estimation is not feasible because the dimension of vector X is too high.

$$\begin{aligned}\hat{y}_{zbcx} =& \hat{E}[Mig_i | x_i = x, (z_i = z, b_i = b, c_i = c)] \\ =& x'\hat{\beta} + \hat{g}(z, b, c)\end{aligned} \tag{34}$$

Another advantage for the first two simplifications in our estimation is that the linearly separable form of the conditional expectation function for migration at nodes, results in an elimination of the $x'\hat{\beta}$ term in the estimations of treatment effects. For example, for $BLATE_A$, $y_{022x} - y_{012x} = x'\hat{\beta} + \hat{g}(0,2,2) - [x'\hat{\beta} + \hat{g}(0,1,2)] = \hat{g}(0,2,2) - \hat{g}(0,1,2)$. The same thing happens for all the average treatment effects. Furthermore, all probabilities are approximated using unconditional counterparts, which means that all the conditional average treatment effects will not depend on X in the estimation. Specifically, there is no need to take the average over the distribution of X within types. Removing the need to estimate $P(X|type)$ reduces the computational burden.

[35]To check for the robustness, we also run all the estimations for a totally unconditional model, that is, when both the probabilities and the expectations are unconditional (getting rid of controls X). The estimated treatment effects are qualitatively the same and even more salient compared with our main results in the paper. The results are available upon request.

[36]The details of estimation for this partial linear model are shown in Appendix E.

Third, we need to estimate the selection probability P_{A_1x} given a set of features X_A. Instead of using P_{A_1x}, that is, the selection ratio conditional on X, we also approximate it by the unconditional counterpart. As the set of characteristics X_A can be large, a machine learning method can be utilized. The algorithm we choose in this paper is the L-1 penalized logit model, which is a combination of the logistic model and the LASSO model. It can be presented as follows.

We assume that the conditional probability function is in a logit form. However, we do not know what predictors to put in the function, so we need to conduct a variable selection process. For a given tuning parameter λ,[37] we try to find the best β to maximize the L-1 regularized sum of log-likelihood:

$$\max_{\beta} \sum ln[\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}] - \frac{1}{2}\lambda\|\beta\|_1 \tag{35}$$

The difference between a common logit model and this penalized logit model lies in the additional regularization term. The appearance of this term will penalize those model specifications with high absolute values of β, that is, with high L-1 norms of β. In a situation with a large number of X_A at hand, it is tempting for researchers to include all the regressors into the common logit model to achieve the highest in-sample prediction accuracy. However, this will lead to a typical over-fitting problem with poor out-of-sample behavior. To avoid this problem, the penalized term in the L-1 penalized logit model can serve as a filter to drop those characteristics in X_A with little explanatory power. The tuning parameter λ determines how hard we penalize the over-fitting such that a higher value of λ results in a harder penalization. We choose the best λ by 10-folds cross-validation.[38] After obtaining the selected predictors and the corresponding coefficients β by maximizing (35), we can calculate the predicted probabilities using the logit model.

We can then derive an estimate of probability of being always-takers $\hat{P}_i(A|X_A)$ for each household. By substituting the estimated $\hat{P}_i(A)$ in Equation (25),[39] we can

[37]Tunning parameter refers to a parameter controlling the strength of regularization or penalty in the model.

[38]In consideration of computational burden, we only train a model for the original data set and use this model for each draw during the process of bootstrap, rather than training a new model once we have a new draw. This may introduce some biases in the estimation of standard error, but we claim that it will not be very severe because the predicted P_A is pretty stable as we change drawn samples.

[39]See Appendix D for details.

get the estimate for the selection probability P_{A_1} as follows:

$$\hat{P}_{A_1} = \frac{\sum\limits_{i\in(z=1,b=1,c=2)} \hat{P}_i(A|X_A)}{\sum\limits_{i\in(z=1,b=1,c=2)} \hat{P}_i(A|X_A) + \sum\limits_{i\in(z=1,b=0,c=2)} \hat{P}_i(A|X_A)} \tag{36}$$

5.2 Bound Estimation

As we have claimed, Assumption 4 is strong and the use of the machine learning method may result in some difficulties in statistical inference and even leads to inconsistency and biasness. In this subsection, we retreat from the attempt to obtain the point estimate of the treatment effect and turn to bound estimation. One challenge is that we need to choose the bounds for P_{A_1}.[40]

The most straightforward bound is that $P_0 \leq P_{A_1} \leq 1$, and $P_0 \leq P_C \leq 1$. It means that the selection rate should be larger than the natural sex ratio and smaller than 1, which leads to a very simple linear programming problem.

$$\begin{cases} n(z=1,b=1,c=2) = (1-P_0)P_{A_1}n_A + (1-P_0)P_C n_C \\ P_0 \leq P_{A_1} \leq 1 \\ P_0 \leq P_C \leq 1 \end{cases} \tag{37}$$

We can substitute $\hat{P}_0$, $\hat{n}_A$, and $\hat{n}_C$ which are estimated using Equations (30)-(33) and solve for a bound for P_{A_1} and P_C. The resulting bounds are as follows:

$$\begin{cases} 0.532 \leq P_{A_1} \leq 0.871 \\ 0.532 \leq P_C \leq 0.980 \end{cases} \tag{38}$$

As we can see from Proposition 4, the estimate of $BLATE_C$ contains a term involving $(1-P_C)$ in the denominator. Therefore, if P_C is too close to 1, this term will be magnified and exploded. To find a meaningful bound for our targeted treatment effect, we need additional criteria. One of the conditions we can add is $P_{A_1} \geq P_C$, which claims that always-takers are those who care less about government rules. Therefore, always-takers are more likely to use the forbidden gender detection tech-

[40]In this subsection, we still approximate the conditional probabilities directly by unconditional counts.

nology and do gender selection. As a consequence, the bounds can be shrunk to the following:

$$\begin{cases} 0.725 \le P_{A_1} \le 0.871 \\ 0.532 \le P_C \le 0.725 \end{cases} \tag{39}$$

This restriction seems to be too strong and shrinks the bounds too much. When doing estimations, we relax the upper bound of P_C to 0.9. In the meantime, as discussed in Section 4.3.5, the effect of the child gender is not a monotonic function of P_{A_1} and the assumption of "P_{A_1x} and P_{Cx} do not depend on X" is very strong, so we will do more estimations with different P_{A_1} in Appendix F to check the robustness.

6 Results

In this section, we report our estimation results for the model with gender selection. We first report the point estimation results under Assumptions 1 to 4. Then, we relax Assumption 4 and report the bound estimation results.

6.1 Results for Point Estimation

In this subsection, we report the estimation results for the point estimation under Assumption 4. First, we display the estimates of the number of households in different types according to Equations (31)-(33) as follows:

$$\begin{cases} \hat{n}_N = 11233 \\ \hat{n}_C = 19624 \\ \hat{n}_A = 25938 \end{cases} \tag{40}$$

Equation system (18) is actually over-identified. Thus, we have two ways to get $n_C + n_A$. If the first child's gender is random for all the households, that is, the gender ratio is P_0 for all the three types, then these two ways should give us similar estimates of $n_C + n_A$. The first way is to solve with the first four equations in (18), which leads to an estimate of 45562 for $n_C + n_A$. The second way is to add up the last two equations in (18), which leads to an estimate of 45559 for $n_C + n_A$. The difference between these two estimates is slim, which can be regarded as an indirect evidence of the randomness of the first child's gender.

Second, we show the estimation results of selection probability using the L-1 penalized logit model. The feature set X_A consists of 8573 variables, including household information, parents' personal information, and detailed location information. Figure 6 shows a 10-fold cross-validation prediction accuracy for different choices of tunning parameter λ.[41] The best λ approximately equals 2 with an in-sample cross-validation accuracy of 0.77. The resulting estimated selection probability is $\hat{P}_{A_1} = 0.787$, $\hat{P}_C = 0.673$.

To investigate whether the prediction model works well, we implement several out-of-sample checks. The first out-of-sample check is to predict the probability of being an always-taker for the households at the node (z=1,b=0,c=1), who are never-takers for certain. If our prediction model is reliable, then most of these households should be classified as non-always-takers. With 5257 observations at the node, on average, the predicted probability of being an always-taker is 0.29; the accuracy of classifying the observations to be a non-always-taker is 0.746, slightly lower than the in-sample prediction above.

In the second out-of-sample check, we predict the probability of being an always-taker for the households in the lower part of the birth tree. We can get the predicted proportion of always-takers for the households in the lower part of the birth tree, which equals about 38%. We can also get this proportion using another method by estimating Equation (18), and the estimated proportion equals about 46%. The difference between these two estimates is not small, which indicates that our prediction of probability to be always-takers is lower than it should be. To alleviate this problem, we can turn to bound estimation as we will show in Section 6.2.

Table 7 shows the estimation results for $FLATE$ and $BLATE$ with the estimated selection probability.[42] For the effect of the number of children, one additional child in the household will increase the probability that both parents migrate by 1.33 percentage points, the father migrates by 2.31 percentage points, and the mother migrates by 0.248 percentage points. Relative to the average migration rate (see Table 2), one additional child will increase the probability that both parents migrate

[41] In a classification model, people usually classify observations with conditional probabilities higher than 0.5 to be Class 1 and observations with conditional probabilities less than 0.5 to be Class 0. By comparing the in-sample prediction of the model to the true class of the observation in the sample, we can calculate an overall accuracy of the model for in-sample prediction.

[42] We also show the 90 percentile confidence interval of the estimation in Figure 7.

by 13.71%, the father migrates by 13.59%, and the mother migrates by 2.07%. Compared with the results in the simple IV estimation, the effect of the number of children becomes considerably more positive. For father migration, the new estimate is four times larger than the corresponding simple IV estimate. For both parents migration and mother migration, the estimates change their signs.

For the effect of the child gender, almost all estimates are both economically and statistically significant. Holding the number of children in the household constant, compared with a girl, one additional boy will further increase the probability that both parents migrate by 1.03 percentage points, the father migrates by 0.91 percentage points, and the mother migrates by 0.86 percentage points. Relative to the average migration rate, one additional boy will further increase the probability that both parents migrate by 10.62%, the father migrates by 5.35%, and the mother migrates by 7.17%. Compared with the effect of the number of children, the child gender effect is comparable in terms of absolute magnitude and is considerably more significant. As a result, it is safe to conclude that the gender of children does affect parents' migration decision, which means that the exclusion restriction of the simple IV method fails.

All these estimates are positive, implying that more children will drive parents to migrate to cities for work. This is the evidence for the dominance of the need for income over separation cost. Furthermore, boys are more likely to motivate parents to migrate. One explanation is that parents need to prepare for the competitive marriage market their sons have to face in the future (Wei and Zhang, 2011; Li and Yi, 2015). The implementation of OCP leads to a highly distorted sex ratio at birth in China, which increases the competitiveness of the marriage market for boys. To enhance their children's attractiveness, parents have to save money for their sons, compelling them to go to cities for work.

As both the effect of the number of children and the effect of the gender of children are positive, the simple IV strategy fails. On the one hand, having a first-born girl increases the number of children in the household under the 1.5 Child Policy, on the other hand, it mechanically decreases the number of boys in the household. Given that both the effects are positive on migration decision, having a first-born girl will encourage migration through the number effect but discourage it through the gender effect, which results in an offset between the two. In the end, the total effect will be attenuated and even flip the sign to negative. Hence, we cannot interpret a policy effect to be purely the effect of the number of children.

The policy implication here is very clear that the controlling the population will not nudge the urbanization process in developing countries like China. Rather, it may hinder the migration of people from rural area to urban area. If some other developing country thinks they can accelerate the urbanization by encouraging the use of the contraception according to the results of OCP in China, then they are going on the absolutely wrong way. Previous studies use traditional IV method and find that OCP can reduce the number of children in the household and the decrease of the number of children will further increase the rural-urban migration. However these studies do not consider the effect of child gender and mistakenly attribute the positive effect of OCP on urbanization in China, to the reduction of the number of children. In fact, it is the distortion of the gender ratio that is playing the role to encourage parents to migrate. However, this comes with huge price in terms of both economic benefits[43] and human rights.

We can draw four conclusions from the point estimation. First, both the effect of the number of children and the effect of the gender of children are positive. Second, the effect of the gender of children is comparable with the effect of the number of children in terms of magnitude and even more statistically significant which means that exclusion restrction in simple IV method fails in our problem. Third, the effect of the gender of children and the effect of the number of children will offset each other. Therefore, a traditional IV method that only considers the number of children is biased, which may give a wrong policy implication. Fourth, both effects are more significant for fathers than for mothers.

6.2 Results for Bound Estimation

As discussed in Section 4, Assumption 4 is essential for point identification. If we are not willing to impose such an assumption, we can turn to set identification and estimate the bounds of the treatment effects. We discuss the results of the bound estimation in this subsection.

The bounds we choose for the two selection probabilities are the ones we mentioned in Section 5.2.[44]

[43] Wei and Zhang (2011) shows that the imbalance of the gender ratio distorts the savings rate and causes the roaring real estate price in China.

[44] Here we relax the upper bound of P_C to be 0.9 rather than 0.725 to make the bounds more flexible.

$$\begin{cases} 0.593 \le P_{A_1} \le 0.871 \\ 0.532 \le P_C \le 0.9 \end{cases} \tag{41}$$

Estimation results for P_{A_1} upper bound (correspondingly, P_C lower bound) and P_{A_1} lower bound (correspondingly, P_C upper bound) are shown in Tables 8 and 9, respectively.[45] Considering the effect of the number of children, we can see that the estimates are smaller when P_{A_1} achieves its upper bound. All the estimates are still larger than those in the simple IV estimation (Tables 4 and 13). The estimates of the effect of the gender of children are still positive and comparable with the effect of the number of children in terms of absolute magnitude for both the upper bound and lower bound estimation. In general, the conclusions we draw from the point estimation still hold in the bound estimation.

The monotonicity of the effect of the number of children guarantees that for any reasonable selection probability we may choose, we can always claim that the simple IV estimates are downward biased. However, concerning the effect of the gender of children, we need to check more choices of P_{A_1} to see the possible estimates of $BLATE$. We show the bound estimation estimates of $BLATE$ in Appendix F, which indicate that our main conclusions are unchanged for all of these different choices of P_{A_1}.

7 Robustness and Discussion

7.1 Exogeneous Gender Selection Assumption

In this subsection, we investigate the robustness of Assumption 3, which states that there is no endogenous gender selection after controlling for observed variables X and fertility behavioral type. We will compare our estimation method under Assumption 3 with other off-the-shelf methods. Moreover, we will discuss what the implication is if Assumption 3 fails.

[45] As we did in point estimation, we also provide the 90 percentile confidence interval of the estimation in Figures 8 and 9.

7.1.1 Comparison to Off-The-Shelf Methods

As discussed, both the number and the gender of children are endogenous in the migration decision equation. The concern is what the advantage of our semi-parametric method is compared to the naive revision method discussed in Section 3.4, which conducts an IV estimation with a linear control of the number of boys. Under this IV estimation method, if we want to get an unbiased estimate, we need to make two assumptions on the number of boys b_{il}. First, conditional on X_{il} and the gender of the first child z_{il}, the number of boys b_{il} is independent of the unobserved u_{il}, which is intuitively similar, but naturally stronger than the "selection on observables" assumption we make. Second, conditional on X_{il} and b_{il}, the gender of the first child z_{il} is independent of the unobserved u_{il}. However, as long as the endogeneity of fertility decision exists, that is, e_i is correlated with u_i, which is the original problem we want to solve, both of the two assumptions do not hold mechanically, no matter what further assumptions we make. The details are shown in Section 4.1 and Appendix B.[46] However, our method works as long as Assumption 3 holds. Assumption 3 in our identification method states that there is no endogenous gender selection after controlling for not only the observed variables but also fertility behavioral type, which is even weaker than the common exogeneity assumption in matching methods.

Generally speaking, the assumption we need for our method is weaker than the assumptions for the off-the-shelf method and the assumptions for the off-the-shelf method do not hold mechanically. The empirical results of this off-the-shelf method is displayed in Appendix A, which shows that the point estimates are slightly improved, but still biased.

In the IV regression framework, the parameter of interest possesses nonparametric interpretation as the LATE with only one endogenous variable and one binary instrument. Nevertheless, once we have two endogenous variables, the joint estimation nature of regression method requires two exclusion restrictions to hold simultaneously and jointly, which is infeasible. On the contrary, in the semi-parametric framework proposed in this study, we directly target on the treatment effects themselves rather than attempt to incorporate them into a parametric tool (linear regression). This more direct and flexible framework demands a considerably weaker "selection

[46] Mathematically, we need the assumption that for W_{il}=[X_{il} C_{il} b_{il} z_{il}]', we have $W_{il} \perp u_{il}$ jointly.

on observables" assumption than off-the-shelf methods.

7.1.2 Effect of Failure of Assumption 3

Considering the situation in which Assumption 3 fails for always-takers and compliers with a first-born girl, that is, even if we control for observed characteristics X, always-takers/compliers with a first-born girl and a second boy, are different from those with a first-born girl and a second girl in some unobserved factors in the migration decision equation. For Chinese households, we assume that $Cov(\eta_i, \epsilon_i|A, X) < 0$ and $Cov(\eta_i, \epsilon_i|C, X) < 0$, which means that households with a stronger boy preference are those with weaker migration preference. For instance, families with more traditional values will be less likely to move to cities and more likely to have boys than those with less traditional values.

In the context of a negative correlation between η_i and ϵ_i, households with lower migration preferences are more likely to conduct gender selection to have a second boy. As a result, we have:

$$E[Mig_{i21}|A, z = 1, b = 1, c = 2, X] < E[Mig_{i21}|A, X] \tag{42}$$

$$E[Mig_{i20}|A, X] < E[Mig_{i20}|A, z = 1, b = 0, c = 2, X] \tag{43}$$

$$E[Mig_{i21}|C, z = 1, b = 1, c = 2, X] < E[Mig_{i21}|C, X] \tag{44}$$

$$E[Mig_{i20}|C, X] < E[Mig_{i20}|C, z = 1, b = 0, c = 2, X] \tag{45}$$

With the above four inequalities, we can determine the sign of the biases in the estimated conditional expectations of the potential migration decisions for different types. Considering Equations (19)-(24), without Assumption 3, the first four equations are unchanged,[47] which means that we can correctly identify $E[Mig_{21}|A, X]$ and $E[Mig_{20}|A, X]$. The last two equations will be altered because $E[Mig_{icb}|z_i, b_i, c_i, Type_i] = E[Mig_{icb}|Type_i]$ does not hold for nodes (z=1,b=1,c=2) and (z=1,b=0,c=2) due to the selection on a second boy. Equations (23) and (24) are altered respectively as follows:

[47] For $E[Mig_{i21}|A, X]$, it can be identified by y_{012x}; for $E[Mig_{i22}|A, X]$, it can be identified by y_{022x}; and for $E[Mig_{i20}|A, X]$, it can be identified by the difference between y_{022x} and y_{012x}. The same logic applies to $E[Mig_{i11}|N, X], E[Mig_{i11}|Co, X]$ and $E[Mig_{i10}|N, X]$. Thus the bias only happens to $E[Mig_{i21}|C, X]$ and $E[Mig_{i20}|C, X]$.

$$
\begin{aligned}
y_{102x} =& \frac{(1-P_{A_1x})n_{Ax}}{(1-P_{A_1x})n_{Ax}+(1-P_{Cx})n_{Cx}} E[Mig_{i20}|A, z=1, b=0, c=2, X] \\
&+ \frac{(1-P_{Cx})n_{Cx}}{(1-P_{A_1x})n_{Ax}+(1-P_{Cx})n_{Cx}} E[Mig_{i20}|C, z=1, b=0, c=2, X], \quad (46) \\
y_{112x} =& \frac{P_{A_1x}n_{Ax}}{P_{A_1x}n_{Ax}+P_{Cx}n_{Cx}} E[Mig_{i21}|A, z=1, b=1, c=2, X] \\
&+ \frac{P_{Cx}n_{Cx}}{P_{A_1x}n_{Ax}+P_{Cx}n_{Cx}} E[Mig_{i21}|C, z=1, b=1, c=2, X]. \quad (47)
\end{aligned}
$$

Therefore, if we still solve for the conditional expectations of the potential migration decisions for compliers using Equation (23) and (24), we will mistakenly identify $E[Mig_{i21}|C, X]$ as follows:

$$
\frac{(P_{A_1x}n_{Ax}+P_{Cx}n_{Cx})y_{112x} - P_{A_{1x}}n_{Ax}E[Mig_{i21}|A, X]}{P_{Cx}n_{Cx}}, \tag{48}
$$

(48) is smaller than $E[Mig_{i21}|C, z=1, b=1, c=2, X]$, which can be identified as below:

$$
\frac{(P_{A_1x}n_{Ax}+P_{Cx}n_{Cx})y_{112x} - P_{A_1x}n_{Ax}E[Mig_{i21}|A, z=1, b=1, c=2, X]}{P_{Cx}n_{Cx}}. \tag{49}
$$

Given that Inequality (44) holds, we conclude that (48)<(49)<$E[Mig_{i21}|C, X]$. Therefore, $E[Mig_{i21}|C, X]$ will be downward biased. The same thing happens when identifying $E[Mig_{i20}|C, X]$, but in the opposite direction, which results in an upward bias. In total, we will have a downward bias when identifying $FLATE$ and $BLATE$, according to Equations (4) and (7).

As all the estimates are positive under Assumption 3, the true effect of the number of children and the effect of the gender of children will be more positive and will still offset each other. All main qualitative conclusions remain valid.

7.2 simple IV Estimation Results for Families with Less than Three Children

The simple IV estimation in Section 3 utilizes the whole sample including all families with more than two children, which is usually what other researchers do. However, to successfully identify treatment effects in our semi-parametric model, we only keep households with less than three children, which accounts for about 90% of the whole sample. To make the comparison between the simple IV model and our semi-parametric model more parallel, we run all the IV regressions again with the sample we used in our semi-parametric method.

The results of the first stage are shown in Table 10. The relation between the instrument z_{il} and the endogenous variable c_{il} is still very strong. The estimate shows that a first-born girl increases the number of children in the household by 0.34. This effect is a little lower than the one estimated using the whole sample because we drop all households with more than two children. Many people with three children are those who are eager to get at least one boy but fail at the first two tries. They definitely respond to the gender of the first child more in subsequent fertility choices.

The IV results in Table 11 illustrate the same pattern we discover in Section 3; that is, for most of the migration measures, the estimates of the effect of the number of children are small and insignificant. Compared to the results estimated using the whole sample, the results estimated using the restricted sample turn out to be more negative, which strengthens our conclusion that the simple IV estimate is downward biased.

8 Conclusion

In this paper, we examine the effect of children on parents' migration decision, specifically, focusing on both the number and the gender of children. We propose a new semi-parametric method to separately identify these two effects, incorporating matching strategy and China's OCP as a natural experiment. Although the model is applied to the parents' migration decision in this paper, it can be applied to any household decisions including parents' divorce/marriage decision, mother's labor participation decision, and household savings.

The results show that one additional child in the household will increase the probability of both parents migrating by 13.7%; one additional boy will increase the probability of both parents migrating by 10.6%, which confirms that the need for income motivation dominates the separation cost. For both boys and girls, the effects are more significant on fathers' migration decisions than for mothers' migration decisions. The gender of children plays an important role in determining parents' migration decision and cannot be neglected because the magnitudes of the number effect and gender effect are comparable with each other. Comparing the results with the ones estimated from the simple IV estimation, we find that the traditional IV method confounds the two effects and causes them to cancel out each other, resulting in a relatively small estimate with no statistical significance. This comparison has an important implication to all previous and future research on OCP. Specifically, any misuse of the policy instrument may give researchers a contaminated estimation and mask the real effect the researchers want to identify. The results show important policy implications. It is the distorted gender ratio that leads to the positive effect of OCP enforcement intensity on China's rural-to-urban migration, rather than the reduction of the number of children. Any effort to reduce the population growth in developing countries are likely to hinder the urbanization but not accelerate it.

The paper still has some limitations. First of all, our model does not allow birth order to affect the outcome. For instance, a first boy and a second girl must have the same effect as a first girl and a second boy.[48] Second, to derive an accurate point identification, we need to have an excellent predictor set, which requires high data quality. Third, the model only fits well in a society with a strong boy preference and a strict fertility control policy, so that we are able to eliminate the existence of the defier (those who will give birth to a second child if the first child is a boy but will not if the first child is a girl). Our work is the first attempt to decompose the effects of the number and gender of children, and solve the problem of the violation of the exclusion restriction in the literature that uses child gender or policy-related variables as instruments for fertility. Further research should focus on solving these problems in other settings.

[48]However, the model specification is much less restrictive than traditional constant effect linear model which is widely used in previous studies.

References

Adda, Jérôme, Christian Dustmann, and Katrien Stevens. 2017. "The Career Costs of Children." *Journal of Political Economy* 125 (2):293–337.

Angrist, Joshua D and William N Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *The American Economic Review* 88 (3):450–477.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434):444–455.

Angrist, Joshua D and Jörn-Steffen Pischke. 2009. "Mostly Harmless Econometrics: An Empiricist's Companion." .

Banister, Judith. 2004. "Shortage of Girls in China Today." *Journal of Population Research* 21 (1):19–45.

Barcellos, Silvia Helena, Leandro S Carvalho, and Adriana Lleras-Muney. 2014. "Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently?" *American Economic Journal: Applied Economics* 6 (1):157–89.

Behrman, Jere R. 1988. "Intrahousehold Allocation of Nutrients in Rural India: Are Boys Favored? Do Parents Exhibit Inequality Aversion?" *Oxford Economic Papers* 40 (1):32–54.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education." *The Quarterly Journal of Economics* 120 (2):669–700.

Cáceres-Delpiano, Julio. 2006. "The Impacts of Family Size on Investment in Child Quality." *Journal of Human Resources* 41 (4):738–754.

Chang, Hongqin, Xiao-yuan Dong, and Fiona MacPhail. 2011. "Labor Migration and Time Use Patterns of the Left-behind Children and Elderly in Rural China." *World Development* 39 (12):2199–2210.

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

Conley, Dalton. 2004. "What is the "True" Effect of Sibship Size and Birth Order on Education? Instrumental Variable Estimates from Exogenous Variation in Fertility."." *New York University, Working Paper* .

Dahl, Gordon B and Enrico Moretti. 2008. "The Demand for Sons." *The Review of Economic Studies* 75 (4):1085–1120.

Du, Yang, Albert Park, and Sangui Wang. 2005. "Migration and Rural Poverty in China." *Journal of Comparative Economics* 33 (4):688–709.

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

García, Jorge Luis. 2017. "Fertility After China's More-than-One-Child Policy." *Unpublished Ph.D. Thesis, University of Chicago, Department of Economics* .

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

Huang, Wei. 2016. "Fertility Restrictions and Life Cycle Outcomes: Evidence from the One Child Policy in China." *Unpublished Ph.D. Thesis, University of Harvard, Department of Economics* .

Huang, Wei, Xiaoyan Lei, and Yaohui Zhao. 2016. "One-Child Policy and the Rise of Man-made Twins." *Review of Economics and Statistics* 98 (3):467–476.

Ichino, Andrea, Elly-Ann Lindström, and Eliana Viviano. 2014. "Hidden Consequences of a First-born Boy for Mothers." *Economics Letters* 123 (3):274–278.

Imbens, Guido W and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2):467–475.

Jacobsen, Joyce P, James Wishart Pearce III, and Joshua L Rosenbloom. 2001. "The Effects of Child-Bearing on Women's Marital Status: Using Twin Births as a Natural Experiment." *Economics Letters* 70 (1):133–138.

Jayachandran, Seema and Ilyana Kuziemko. 2011. "Why Do Mothers Breastfeed Girls Less Than Boys? Evidence and Implications for Child Health in India." *The Quarterly Journal of Economics* 126 (3):1485–1538.

Jena, Anumpam. 2006. "Do Children Cause Divorce? Evidence from a Natural Experiment Using Twins." *University of Chicago, Working Paper* .

Jia, Zhaobao and Wenhua Tian. 2010. "Loneliness of Left-behind Children: A Cross-sectional Survey in a Sample of Rural China." *Child: Care, Health and Development* 36 (6):812–817.

Lee, Jungmin. 2003. "Children and Household Education Decisions: An Asian Instrument." *Working Paper* .

Li, Bingjing and Hongliang Zhang. 2016. "Does Population Control Lead to Better Child Quality? Evidence from China's One-Child Policy Enforcement." *Journal of Comparative Economics* 45 (2):246–260.

Li, Wenchao and Junjian Yi. 2015. "The Competitive Earning Incentive for Sons: Evidence from Migration in China." *IZA Discussion Paper* No. 9214.

Lundberg, Shelly and Elaina Rose. 2002. "The Effects of Sons and Daughters on Men's Labor Supply and Wages." *Review of Economics and Statistics* 84 (2):251–268.

Oliveira, Jaqueline. 2016. "Fertility, Migration, and Maternal Wages: Evidence from Brazil." *Journal of Human Capital* 10 (3):377–398.

Pinto, Rodrigo. 2015. "Selection Bias in a Controlled Experiment: The Case of Moving to Opportunity." *Unpublished Ph.D. Thesis, University of Chicago, Department of Economics* .

Qian, Nancy. 2009. "Quantity-Quality and the One Child Policy:The Only-Child Disadvantage in School Enrollment in Rural China." *NBER Working Paper* No. 14973.

Robinson, Peter M. 1988. "Root-N-Consistent Semiparametric Regression." *Econometrica* 54 (4):931–954.

Rosenzweig, Mark R and Kenneth I Wolpin. 1980. "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica: Journal of the Econometric Society* :227–240.

———. 2000. "Natural "Natural Experiments" in Economics." *Journal of Economic Literature* 38 (4):827–874.

Rosenzweig, Mark R and Junsen Zhang. 2009. "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's "One-Child" Policy." *The Review of Economic Studies* 76 (3):1149–1174.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5):688.

Wei, Shang-Jin and Xiaobo Zhang. 2011. "The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China." *Journal of Political Economy* 119 (3):511–564.

Ye, Jingzhong and Pan Lu. 2011. "Differentiated Childhoods: Impacts of Rural Labor Migration on Left-behind Children in China." *The Journal of Peasant Studies* 38 (2):355–377.

Zhang, Hongliang, Jere R Behrman, C Simon Fan, Xiangdong Wei, and Junsen Zhang. 2014. "Does Parental Absence Reduce Cognitive Achievements? Evidence from Rural China." *Journal of Development Economics* 111:181–195.

Zhang, Junsen. 2017. "The Evolution of China's One-Child Policy and Its Effects on Family Outcomes." *Journal of Economic Perspectives* 31 (1):141–160.

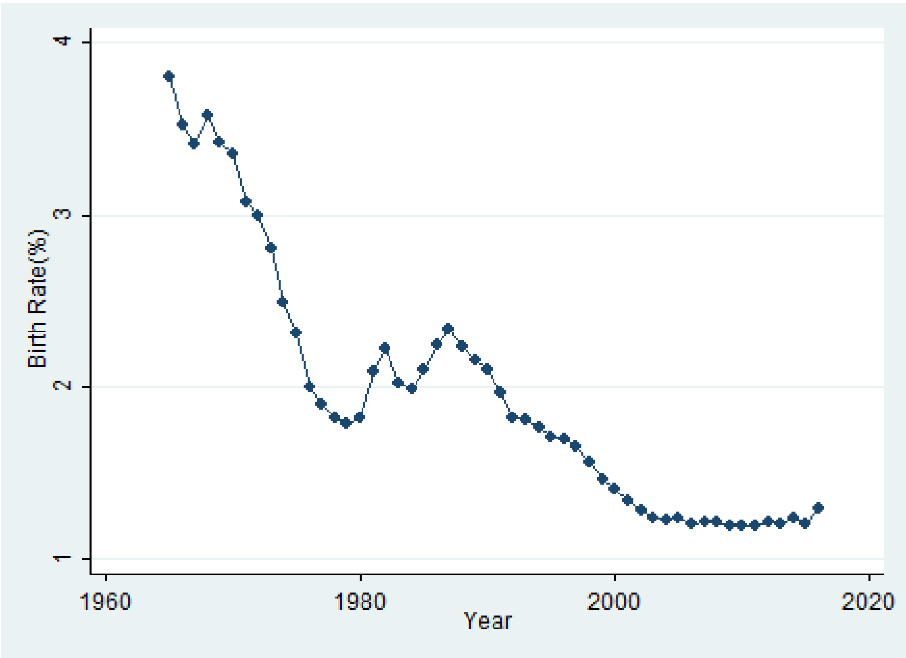


Figure 1: China's Birth Rate 1965-2016 (%)
Notes: Data from World Bank

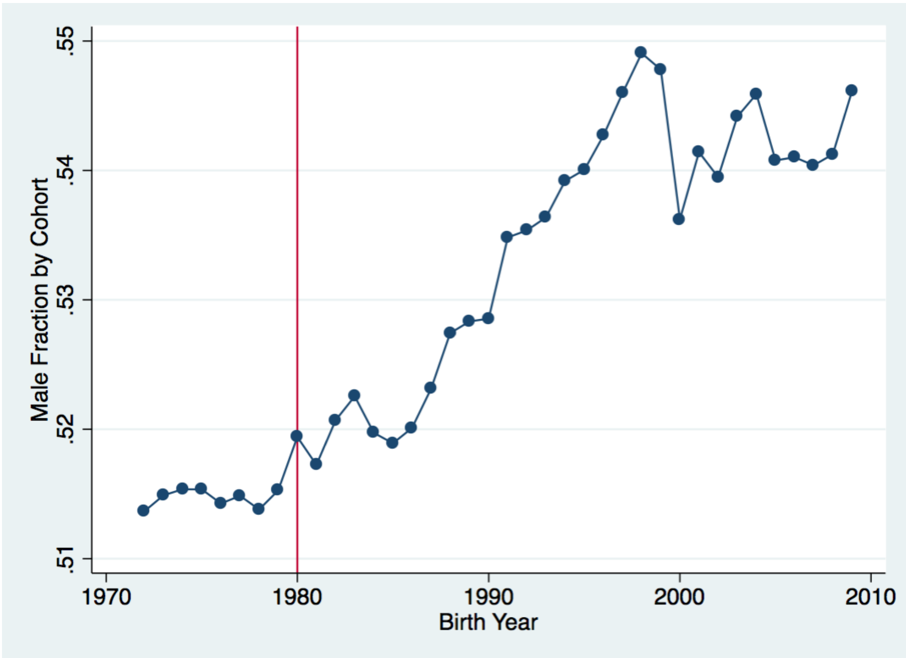


Figure 2: China's Gender Imbalance 1980-2010

Notes: Using data from China Census 1982, 1990, 2000, and 2010, we calculate the gender ratios for neonatus from 1972 to 2009. To avoid the confoundness of mortality, data from each census are only used to calculate the gender ratios for the 10 years before the survey year. The red line represents the starting time of the OCP.

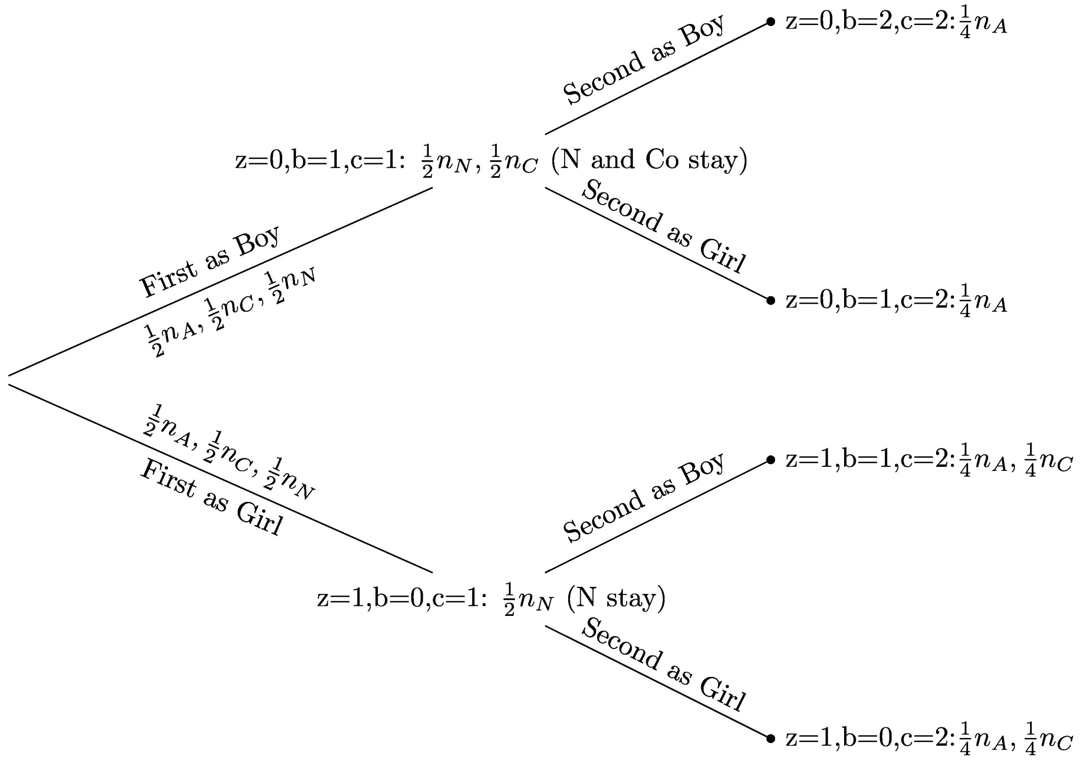


Figure 3: Birth Tree

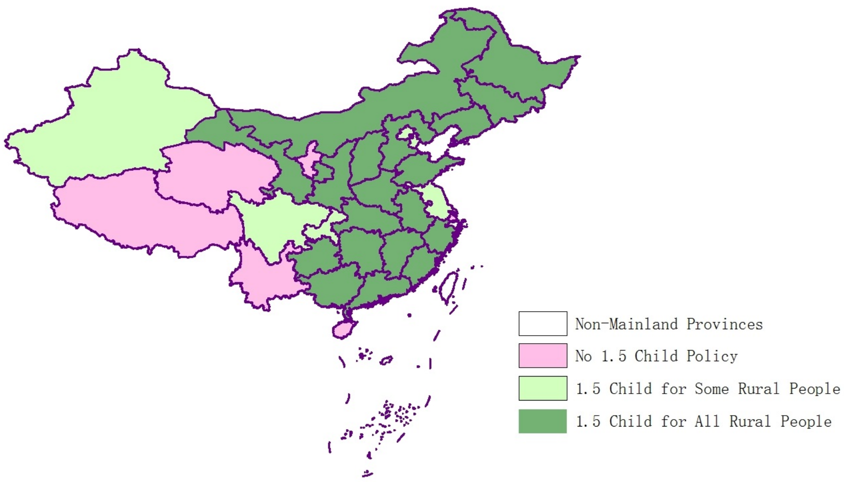


Figure 4: The Distribution of Provinces with 1.5 Child Policy

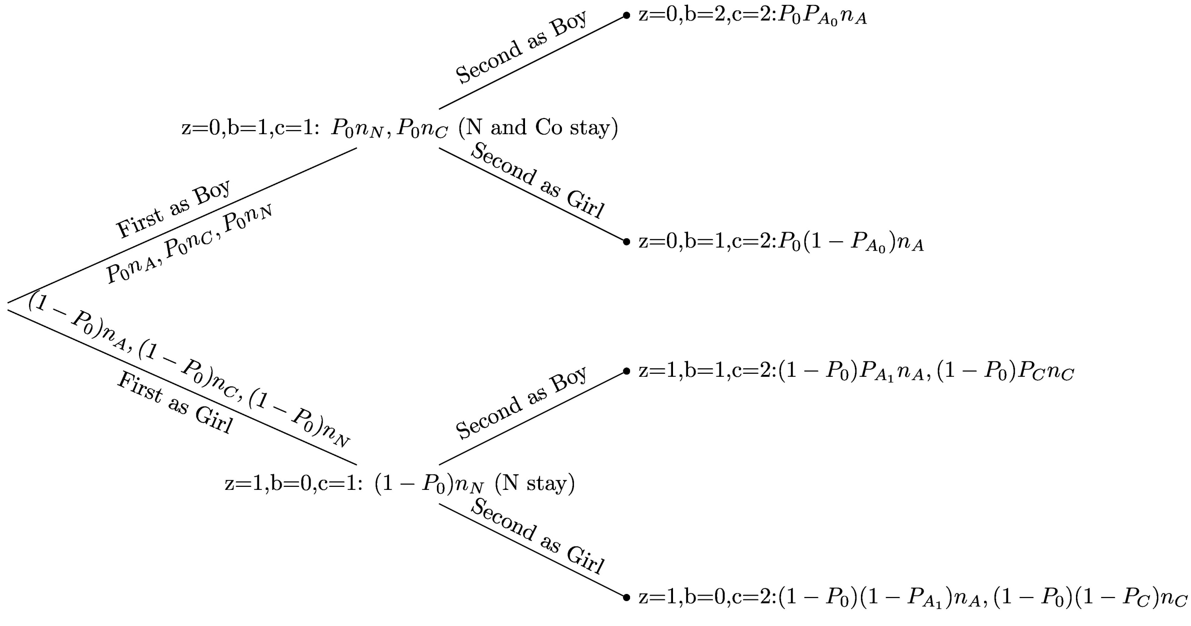


Figure 5: Birth Tree with Gender Selection

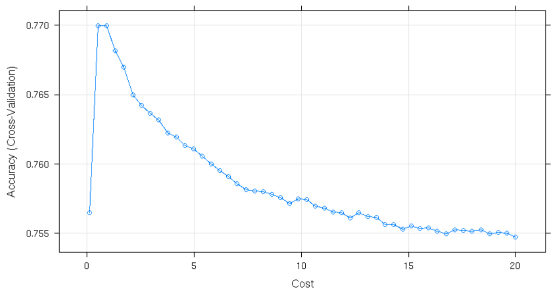
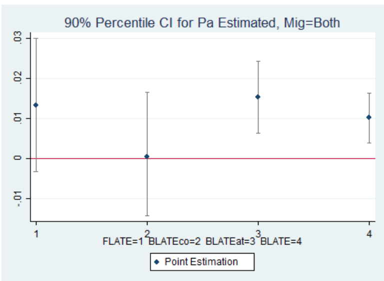
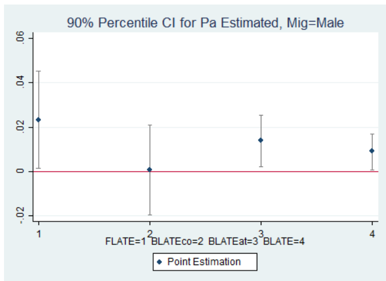


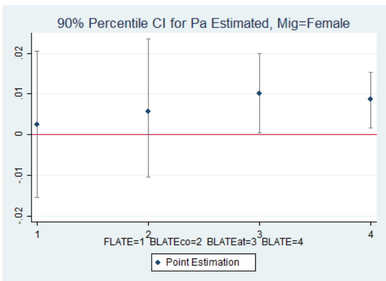
Figure 6: CV Accuracy for Different Choices of Cost Parameter λ



(a) P_{A_1} Estimated, Mig=Both

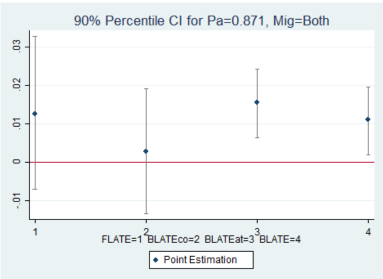


(b) P_{A_1} Estimated, Mig=M

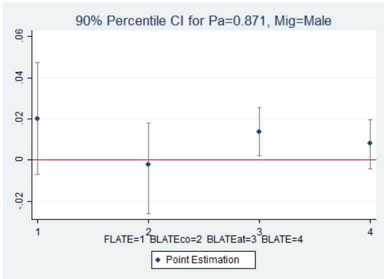


(c) P_{A_1} Estimated, Mig=F

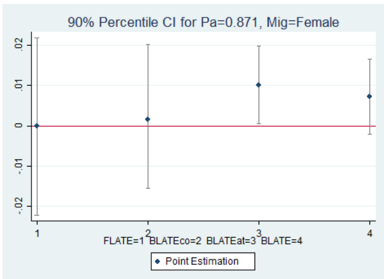
Figure 7: 90% Percentile CI of Estimations for Estimated P_{A_1}



(a) P_{A_1}=0.871, Mig=Both

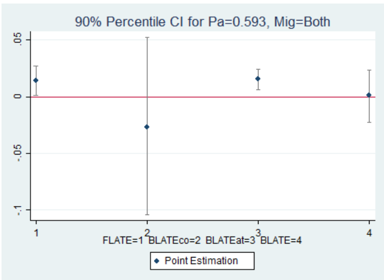


(b) P_{A_1}=0.871, Mig=M

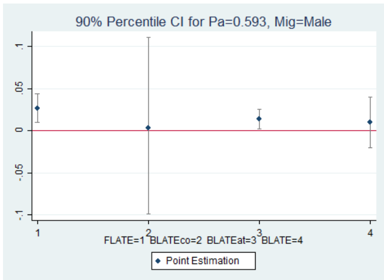


(c) P_{A_1}=0.871, Mig=F

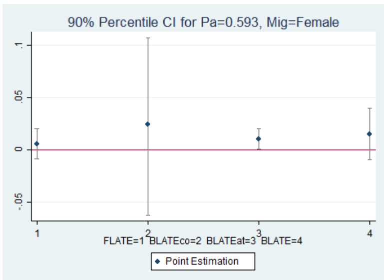
Figure 8: 90% Percentile CI of Estimations for Upper Bound of P_{A_1}



(a) P_{A_1}=0.593, Mig=Both



(b) P_{A_1}=0.593, Mig=M



(c) P_{A_1}=0.593, Mig=F

Figure 9: 90% Percentile CI of Estimations for Lower Bound of P_{A_1}

Table 1: List of Provinces offering 1.5 Child Policy

Province	All 1.5	Some 1.5	Province	All 1.5	Some 1.5
Beijing	NO	YES	Henan	YES	YES
Tianjin	NO	YES	Hubei	YES	YES
Hebei	YES	YES	Hunan	YES	YES
Shanxi	YES	YES	Guangdong	YES	YES
Inner Mongolia	YES	YES	Guangxi	YES	YES
Liaoning	YES	YES	Hainan	NO	NO
Jilin	YES	YES	Chongqing	NO	YES
Heilongjiang	YES	YES	Sichuan	NO	YES
Shanghai	NO	NO	Guizhou	YES	YES
Jiangsu	NO	YES	Yunnan	NO	NO
Zhejiang	YES	YES	Tibet	NO	NO
Anhui	YES	YES	Shaanxi	YES	YES
Fujian	YES	YES	Gansu	YES	YES
Jiangxi	YES	YES	Qinghai	NO	NO
Shandong	YES	YES	Ningxia	NO	NO
Xinjiang	NO	YES			

Table 2: Summary Statistics

Variable	Mean	Std.Dev	Min	Max
Dependent Variables				
If both parents migrate	0.097	0.30	0	1
If mother migrates	0.12	0.33	0	1
If father migrates	0.17	0.38	0	1
Interested Independent Variables				
Number of children in the household	1.80	0.71	1	8
Number of boys in the household	0.98	0.60	0	5
Family Characteristics				
Age of the first-born child	14.47	2.99	0	18
Father's age	40.38	3.60	35	65
Mother's age	39.18	3.04	35	64
If father is literate	0.99	0.072	0	1
If mother is literate	0.98	0.15	0	1
Father's education	8.62	1.68	0	19
Mother's education	7.96	2.01	0	19
First child gender	0.50	0.50	0	1

Notes: Data source: China 2010 Census Data.

Table 3: First Stage Results for the simple IV Estimation

	#children in the household		
	(1)	(2)	(3)
first child sex(=1 if girl)	0.465***	0.411***	0.417***
	(0.0155)	(0.0114)	(0.0112)
Household Controls	No	Yes	Yes
Parent Personal Controls	No	Yes	Yes
Home Place Fixed Effect	No	No	Yes
Prob>F	0.000	0.000	0.000
Number of Observations	64095	64095	64095

Notes: Household controls include first child's age, number of pre-school children, and marriage time. Parent personal controls include both of the parents' ages, both of the parents' schooling year, and two dummies indicating whether they are literate or not. Standard errors in parentheses are clustered at the prefecture level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 4: IV Estimation Results for Household's Parents' Migration

	Both	Male	Female
	(1)	(2)	(3)
#children	-0.00499	0.00607	-0.00861
	(0.00595)	(0.00658)	(0.00691)
Household Controls	Yes	Yes	Yes
Parent Personal Controls	Yes	Yes	Yes
Home Place Fixed Effect	Yes	Yes	Yes
Number of Observations	64095	64095	64095

Notes: Household controls include first child's age, number of pre-school children, and marriage time. Parent personal controls include both of the parents' age, both of the parents' schooling year, and two dummies indicating whether they are literate or not. Column (1) reports the IV estimation results when the dependent variable is *Both parents' Migration*. Column (2) reports the IV estimation results when the dependent variable is *father's Migration*. Column (3) reports the IV estimation results when the dependent variable is *mother's Migration*. Standard errors in parentheses are clustered at the prefecture level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 5: Distribution of Different Types of Households without Gender Selection

<table>
<tr><th></th><th>$z_i = 1$ (G)</th><th>$z_i = 0$ (B)</th></tr>
<tr><td>$c_i = 1$</td><td>$b_i = 0$: $\frac{1}{2}n_N$</td><td>$b_i = 1$: $\frac{1}{2}n_N$, $\frac{1}{2}n_C$</td></tr>
<tr><td rowspan="2">$c_i = 2$</td><td>$b_i = 0$: $\frac{1}{4}n_A$, $\frac{1}{4}n_C$</td><td>$b_i = 1$: $\frac{1}{4}n_A$</td></tr>
<tr><td>$b_i = 1$: $\frac{1}{4}n_A$, $\frac{1}{4}n_C$</td><td>$b_i = 2$: $\frac{1}{4}n_A$</td></tr>
</table>

Table 6: Distribution of Different Types of Individuals with Gender Selection

	$z_i = 1$ (G)	$z_i = 0$ (B)
$c_i = 1$	$b_i = 0$: $(1 - P_0)n_N$	$b_i = 1$: $P_0 n_N$, $P_0 n_C$
$c_i = 2$	$b_i = 0$: $(1 - P_0)(1 - P_{A_1})n_A$, $(1 - P_0)(1 - P_C)n_C$	$b_i = 1$: $P_0(1 - P_{A_0})n_A$
	$b_i = 1$: $(1 - P_0)P_{A_1}n_A$, $(1 - P_0)P_C n_C$	$b_i = 2$: $P_0 P_{A_0} n_A$

Table 7: semi-parametric Estimation Results with Estimated Selection Probability

	Both	Male	Female
	(1)	(2)	(3)
$FLATE$	0.0133	0.0231*	0.00248
	(0.0103)	(0.0134)	(0.0113)
$BLATE_C$	0.000471	0.000738	0.00577
	(0.00920)	(0.0125)	(0.00996)
$BLATE_A$	0.0154***	0.0138**	0.0101*
	(0.00526)	(0.00675)	(0.00591)
$BLATE$	0.0103***	0.00928*	0.00864**
	(0.00369)	(0.00504)	(0.00406)
Number of Observations	56794	56794	56794

Notes: All estimations are conducted with estimated selection probability. Column (1) reports the semi-parametric estimation results when the dependent variable is *Both Parents' Migration* for $FLATE$, $BLATE_C$, $BLATE_A$, and $BLATE$, respectively. Column (2) reports the semi-parametric estimation results when the dependent variable is *Father's Migration*. Column (3) reports the semi-parametric estimation results when the dependent variable is *Mother's Migration*. Standard errors in parentheses are estimated using a bootstrap method. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 8: semi-parametric Estimation Results with $P^u_{A_1} = 0.871$

	Both	Male	Female
$FLATE$	0.0124	0.0201	-0.000130
	(0.0125)	(0.0165)	(0.0138)
$BLATE_C$	0.00264	-0.00222	0.00146
	(0.00962)	(0.0131)	(0.0104)
$BLATE_A$	0.0154***	0.0138**	0.0101*
	(0.00526)	(0.00675)	(0.00591)
$BLATE$	0.0110**	0.00826	0.00715
	(0.00520)	(0.00696)	(0.00575)
Number of Observations	56794	56794	56794

Notes: All estimations are conducted with $P^u_{A_1} = 0.871$. Column (1) reports the semi-parametric estimation results when the dependent variable is *Both Parents' Migration* for $FLATE$, $BLATE_C$, $BLATE_A$, and $BLATE$, respectively. Column (2) reports the semi-parametric estimation results when the dependent variable is *Father's Migration*. Column (3) reports the semi-parametric estimation results when the dependent variable is *Mother's Migration*. Standard errors in parentheses are estimated using a bootstrap method. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 9: semi-parametric Results with $P^{l}_{A_1} = 0.593$

	Both	Male	Female
$FLATE$	0.0142*	0.0265**	0.00532
	(0.00798)	(0.0104)	(0.00890)
$BLATE_C$	-0.0267	0.00293	0.0240
	(0.0485)	(0.0639)	(0.0530)
$BLATE_A$	0.0154***	0.0138**	0.0101*
	(0.00526)	(0.00675)	(0.00591)
$BLATE$	0.000873	0.0100	0.0150
	(0.0140)	(0.0186)	(0.0151)
Number of Observations	56794	56794	56794

Notes: All estimations are conducted with $P^{u}_{A_1} = 0.593$. Column (1) reports the semi-parametric estimation results when the dependent variable is *Both Parents' Migration* for $FLATE$, $BLATE_C$, $BLATE_A$, and $BLATE$, respectively. Column (2) reports the semi-parametric estimation results when the dependent variable is *Father's Migration*. Column (3) reports the semi-parametric estimation results when the dependent variable is *Mother's Migration*. Standard errors in parentheses are estimated using a bootstrap method. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 10: First Stage Estimation Results for simple IV Estimation with Restricted Sample

	#children in the household		
	(1)	(2)	(3)
first child sex(=1 if girl)	0.346***	0.322***	0.340***
	(0.0200)	(0.0152)	(0.0146)
Household Controls	No	Yes	Yes
Parent Personal Controls	No	Yes	Yes
Home Place Fixed Effect	No	No	Yes
Prob>F	0.000	0.000	0.000
Number of Observations	56794	56794	56794

Notes: Here the restricted samples only includes families with less than three children. Household controls include first child's age, number of pre-school children, and marriage time. Parent personal controls include both parents' age, both parents' schooling year, and two dummies indicating whether they are literate or not. Standard errors in parentheses are clustered at the prefecture level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 11: IV Estimation Results for Household's Parents' Migration with Restricted Sample

	Both	Male	Female
	(1)	(2)	(3)
#children	-0.0105	0.00102	-0.0148*
	(0.00780)	(0.00884)	(0.00883)
Household Controls	Yes	Yes	Yes
Parent Personal Controls	Yes	Yes	Yes
Home Place Fixed Effect	Yes	Yes	Yes
Number of Observations	56794	56794	56794

Notes: Here the restricted samples only includes families with less than three children. Household controls include first child's age, number of pre-school children, and marriage time. Parent personal controls include both parents' age, both parents' schooling year, and two dummies indicating whether they are literate or not. Column (1) reports the IV estimation results when the dependent variable is *Both Parents' Migration*. Column (2) reports the IV estimation results when the dependent variable is *Father's Migration*. Column (3) reports the IV estimation results when the dependent variable is *Mother's Migration*. Standard errors in parentheses are clustered at the prefecture level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Appendix

A A Naive Revision

A naive way to fix the problem is by simply controlling for the child gender effect in the regression and instrument the number of children with an OCP-related IV. However, this method is also problematic. As parents can illegally implement gender selection for their children, the gender of children is also an endogenous decision that can be correlated with unobserved family characteristics. In addition, the number of boys is itself a post-determined endogenous variable even if there is no gender selection. The number of boys is still endogenous if we use the first child gender as the instrument. The reason is that they are determined after the realization of the first child gender, which are called "bad controls" by Angrist and Pischke (2009). Now we show a rigorous mathematical proof under the context of our model in this paper.

We run the following regression, instrumenting the number of children by the gender of the first child:

$$Mig_{il} = \beta_0 + \beta_1 c_{il} + \beta_2 b_{il} + \boldsymbol{X_{il}}\beta_3 + u_{il}, \tag{50}$$

where b_{il} refers to the number of boys in household i from location l. The results of the first-stage regression are displayed in Table 12, which clearly support a strong positive relationship between the gender of the first child and the number of children in the household. The results of the revised IV estimation are shown in Table 13. Only the effect of the number of children on father's migration is now statistically significant. For both parents' migration and mother's migration, the estimates are still not significantly different from zero.

After controlling for the number of boys in the household, point estimates in all specifications increase. However, the estimates are still biased, because the number of boys in the household is endogenous. The endogeneity comes from two kinds of sources. First, conditional on the gender of the first child z_i and X_{il}, the number of boys b_i is not independent of the unobserved term u_{il}. Second, conditional on the number of boys b_i and X_{il}, the gender of the first child z_i is no longer independent of the unobserved term u_{il}. Intuitively, the number of boys in the household

Table 12: First Stage Results for Revised IV Estimation with Linearly Controlling for Boy Number

	#children in the household		
	(1)	(2)	(3)
first child sex(=1 if girl)	0.948***	0.844***	0.723***
	(0.0214)	(0.0200)	(0.0110)
#boys	Yes	Yes	Yes
Household Controls	No	Yes	Yes
Parent Personal Controls	No	Yes	Yes
Home Place Fixed Effect	No	No	Yes
Prob>F	0.000	0.000	0.000
Number of Observations	64095	64095	64095

Notes: Household controls include first child's age, number of pre-school children, and marriage time. Parent personal controls include both parents' age, both parents' schooling year, and two dummies indicating whether they are literate or not. Standard errors in parentheses are clustered at the prefecture level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

is determined after the realization of the first-child gender lottery. Although the randomness of the first child gender guarantees a random division of all households at the time of the realization of first child gender, the post-determined choices (here, number of boys) happening after that will be affected by the realization. Therefore, after controlling for the number of boys b_{il} in the regression, the distributions for u_{il} may no longer be identical for households with a first-born boy and households with a first-born girl.

B An Explanation for the Endogeneity of b_i

We now explain in detail why the first child gender z_i is not independent of u_i conditional on the number of boys b_i. We can see from Table 5, conditional on $b_i = 1$, that when $z_i = 1$, the households are composed of a quarter of the always-takers and a quarter of the compliers $(\frac{1}{4}A, \frac{1}{4}C)$. Similarly, conditional on $b_i = 1$, when $z_i = 0$, the households are composed of a half of the never-takers, a half of the compliers, and a quarter of the always-takers $(\frac{1}{2}N, \frac{1}{2}C, \frac{1}{4}A)$. As the always-takers, never-takers, and compliers have different e_i, the distribution of e_i conditional on $b_i = 1$ is different for $z_i = 0, 1$. So conditional on b_i, z_i is correlated with e_i. As e_i and u_i may also be correlated,[49] z_i may be correlated with u_i conditional on b_i.

[49]Some common unobserved factors affect fertility decision and migration decision at the same time. For instance, an individual with higher acceptance of modern values may not only have less children but may also be more likely

Table 13: Revised IV Estimation Results with Linearly Controlling for Boy Number

	Both	Male	Female
	(1)	(2)	(3)
#children	0.00284	0.0152***	-0.00264
	(0.00394)	(0.00422)	(0.00503)
#boys	Yes	Yes	Yes
Household Controls	Yes	Yes	Yes
Parent Personal Controls	Yes	Yes	Yes
Home Place Fixed Effect	Yes	Yes	Yes
Number of Observations	64095	64095	64095

Notes: Household controls include first child's age, number of pre-school children, and marriage time. Parent personal controls include both parents' age, both parents' schooling year, and two dummies indicating whether they are literate or not. Column (1) reports the IV estimation results when the dependent variable is *Both Parents' Migration*. Column (2) reports the IV estimation results when the dependent variable is *Father's Migration*. Column (3) reports the IV estimation results when the dependent variable is *Mother's Migration*. Standard errors in parentheses are clustered at the prefecture level. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Similarly, we can prove that b_i maybe correlated with u_i conditional on z_i. Hence, we cannot simply control for b_i to net out the boy effect.

C Identification of LATEs without Gender Selection

First of all, we can directly derive from (10),(11) and (13) that $E[Mig_{i21}|A] = y_{012}$, $E[Mig_{i22}|A] = y_{022}$ and $E[Mig_{i10}|N] = y_{101}$. By substituting (10) into (15), we can derive the identification of $E[Mig_{i21}|C]$ as follows:

$$E[Mig_{i21}|C] = \frac{(n_A + n_C)y_{112} - n_A y_{012}}{n_C} \tag{51}$$

Under the first assumption, we have $E[Mig_{i22}|A] - E[Mig_{i21}|A] = E[Mig_{i21}|A] - E[Mig_{i20}|A]$. As a consequence, we can identify $E[Mig_{i20}|A] = 2E[Mig_{i21}|A] - E[Mig_{i22}|A] = 2 \cdot y_{012} - y_{022}$, where $E[Mig_{i22}|A] - E[Mig_{i21}|A]$ is the partial effect of the number of boys on migration for the always-taker group. By substituting the expression of $E[Mig_{i20}|A]$ into (14) we can identify $E[Mig_{i20}|C]$ as follows:

$$E[Mig_{i20}|C] = \frac{(n_A + n_C)y_{102} - n_A(2 \cdot y_{012} - y_{022})}{n_C} \tag{52}$$

to migrate to urban areas. Thus, e_i is correlated with u_i

Then we can identify the average partial effect of a boy compared with a girl on migration for the complier group by $E[Mig_{i21}|C] - E[Mig_{i20}|C]$.

However, without imposing additional assumptions, we cannot distinguish between $E[Mig_{i11}|N]$ and $E[Mig_{i11}|C]$ in this semi-parametric model because there is only one variation in the number of boys for never-takers, from 0 to 1, and we cannot replicate the trick we did for always-takers. If we want to identify the LATE, we need the second assumption.

If both Assumptions 1 and 2 hold, we can derive $E[Mig_{i11}|N]$ as follows:

$$y_{022} - y_{012} = E[Mig_{i11}|N] - y_{101} \tag{53}$$

$$\Rightarrow \quad E[Mig_{i11}|N] = y_{022} - y_{012} + y_{101} \tag{54}$$

Substituting it into (12) we get:

$$E[Mig_{i11}|C] = \frac{(n_N + n_C)y_{011} - n_N(y_{022} - y_{012} + y_{101})}{n_C} \tag{55}$$

Hence, we can identify FLATE as follows:

$$\begin{aligned} & E[Mig_{i21}|C] - E[Mig_{i11}|C] \\ = & \frac{(n_A + n_C)y_{112} - (n_A + n_N)y_{012} - (n_N + n_C)y_{011} + n_N(y_{022} + y_{101})}{n_C} \end{aligned} \tag{56}$$

D Identification of P_{A_1}

$P_{A_1 x}$ can be calculated as follows:

$$P_{A_1 x} = \frac{P(A, b=1|c=2, z=1, X)}{P(A, b=1|c=2, z=1, X) + P(A, b=0|c=2, z=1, X)}, \tag{57}$$

where

$$P(A, b=1|c=2, z=1, X) = \int_{X_A \setminus X} P(A|X_A, z=1, b=1, c=2) P(X_A \setminus X, b=1|c=2, z=1, X) dx_A \setminus x, \tag{58}$$

$$P(A, b=0|c=2, z=1, X) = \int_{X_A \setminus X} P(A|X_A, z=1, b=0, c=2) P(X_A \setminus X, b=0|c=2, z=1, X) dx_A \setminus x. \tag{59}$$

For a comprehensive set X_A with X as its subset, the integrals are with respect to the elements included in X_A but not in X. If Assumption 4 holds, we can write down (58) and (59) as:

$$P(A, b=1|c=2, z=1, X) = \int_{X_A \setminus X} P(A|X_A, z=0) P(X_A \setminus X, b=1|c=2, z=1, X) dx_A \setminus x \quad (60)$$

$$P(A, b=0|c=2, z=1, X) = \int_{X_A \setminus X} P(A|X_A, z=0) P(X_A \setminus X, b=0|c=2, z=1, X) dx_A \setminus x \quad (61)$$

where $P(A|X_A, z=0)$ can be derived from the upper part of the tree, and $P(X_A \setminus X, b=1|c=2, z=1, X) = \frac{P(X_A \setminus X, b=1, c=2, z=1, X)}{P(c=2, z=1, X)}$ can also be derived from the data. We can then identify (57) by plugging in (60) and (61).

A sufficient condition for Assumption 4 is that $P(A|X_A, b, c, z) = P(A|X_A)$ where X_A is a sufficient statistic in predicting always-takers.

E How to Incorporate Controls X in the Estimation

To incorporate controls X, for each y_{zbcx}, that is, the the conditional expectations of the potential migration decisions at nodes, we use a partial linear model developed by Robinson (1988). Let Z be a vector representing the node, Z=(z,b,c), which is a vector including three elements: the gender of the first child, the number of children and the number of boys. We do not impose any functional form restriction to the vector Z in migration decision function but assume that X enters into the function linearly.

$$Mig_i = x_i'\beta + g(Z_i) + u_i \quad (62)$$

where $E[u_i|X_i, Z_i] = 0$. We can express y_{zbc} as follows:

$$\begin{aligned} y_{zbcx} =& E[Mig_i|x_i = x, (z_i = z, b_i = b, c_i = c)] \\ =& x'\beta + g(Z) \end{aligned} \quad (63)$$

In principle, we can incorporate X into the conditional expectation function non-parametrically. However, as the location fixed effect dummies increase the dimensionality considerably, a fully non-parametric method becomes infeasible. Considering that the partial effect of X is not our target, we sacrifice some model functional form flexibility in the migration decision, to avoid the curse of dimensionality.

To estimate y_{zbcx}, the following steps are employed. Step 1: We derive estimates of $E(Mig_i|Z_i)$ and $E(x_i|Z_i)$ using the sample average at the corresponding node. Step 2: We regress $Mig_i - \hat{E}(Mig_i|Z_i)$ on $x_i - \hat{E}(x_i|Z_i)$ and get an estimate of $\hat{\beta}$. Step 3: We calculate $Mig_i - x_i'\hat{\beta}$ for each i and get the estimate of $\hat{g}(Z_i) = \frac{1}{n} \sum_{j \in Z_i} (Mig_j - x_j'\hat{\beta})$.

Then to get the estimates of all treatment effects, we can plug in the estimates above for y_{zbcx}. Standard errors and confidence intervals will be derived by bootstrapping with 500 replications.

F Results for Different Choices of P_{A_1} and P_C

Now we discuss the results when we set probability P_{A_1} to different values. In the main text, we show the estimates for both upper and lower bounds of P_{A_1}, and when $P_{A_1} = \hat{P}_{A_1} 0.787$, which is derived by predicting the number of always-takers. We also estimate our model, setting P_{A_1} equal to 0.8, 0.75, 0.7, and 0.65 (Table 14-16).

Even though we cannot analytically prove the monotonicity of the boy effect in terms of P_{A_1}, in our sample and bootstrap simulation, the monotonicity seems to hold not only for the number effect but also for all the boy effects (The boy effect for always-takers will not be changed when we vary P_{A_1} because it is not a function of it). According to the results, we can safely conclude that all our main statements are true no matter which probability we use.

Table 14: semi-parametric Estimation Results for Both Parents' Migration with Different P_{A_1}

P_{A_1}	0.871 (Upper Bound)	0.8	0.787 ($\hat{P}_A$)	0.75	0.7	0.65	0.593 (Lower Bound)
$FLATE$	0.0124	0.0131	0.0133	0.0135	0.0138	0.0140	0.0142*
	(0.0125)	(0.0108)	(0.0103)	(0.00987)	(0.00914)	(0.00854)	(0.00798)
$BLATE_C$	0.00264	0.00120	0.000471	-0.00069	-0.00389	-0.00979	-0.0267
	(0.00962)	(0.00877)	(0.00920)	(0.0103)	(0.0146)	(0.0233)	(0.0485)
$BLATE_A$	0.0154***	0.0154***	0.0154***	0.0154***	0.0154***	0.0154***	0.0154***
	(0.00526)	(0.00526)	(0.00526)	(0.00526)	(0.00526)	(0.00526)	(0.00526)
$BLATE$	0.0110**	0.0105***	0.0103***	0.00985***	0.00875**	0.00671	0.000873
	(0.00520)	(0.00403)	(0.00369)	(0.00341)	(0.00354)	(0.00567)	(0.0140)
Number of Observations	56794	56794	56794	56794	56794	56794	56794

Notes: Standard errors in parentheses are estimated using a bootstrap method. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 15: semi-parametric Estimation Results for father's Migration with Different P_{A_1}

P_{A_1}	0.871 (Upper Bound)	0.8	0.787 ($\hat{P}_A$))	0.75	0.7	0.65	0.593 (Lower Bound)
$FLATE$	0.0201	0.0224	0.0231*	0.0237*	0.0248**	0.0256**	0.0265**
	(0.0165)	(0.0141)	(0.0134)	(0.0129)	(0.0119)	(0.0111)	(0.0104)
$BLATE_C$	-0.00222	0.0000569	0.000738	0.00125	0.00216	0.00280	0.00293
	(0.0131)	(0.0120)	(0.0125)	(0.0140)	(0.0194)	(0.0308)	(0.0639)
$BLATE_A$	0.0138**	0.0138**	0.0138**	0.0138**	0.0138**	0.0138**	0.0138**
	(0.00675)	(0.00675)	(0.00675)	(0.00675)	(0.00675)	(0.00675)	(0.00675)
$BLATE$	0.00826	0.00904*	0.00928*	0.00945**	0.00977**	0.00999	0.0100
	(0.00696)	(0.00547)	(0.00504)	(0.00471)	(0.00493)	(0.00770)	(0.0186)
Number of Observations	56794	56794	56794	56794	56794	56794	56794

Notes: Standard errors in parentheses are estimated using a bootstrap method. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

Table 16: semi-parametric Estimation Results for mother's Migration with Different P_{A_1}

P_{A_1}	0.871 (Upper Bound)	0.8	0.787 ($\hat{P}_A$))	0.75	0.7	0.65	0.593 (Lower Bound)
$FLATE$	-0.000130	0.00187	0.00248	0.00296	0.00385	0.00460	0.00532
	(0.0138)	(0.0119)	(0.0113)	(0.0109)	(0.0101)	(0.00949)	(0.00890)
$BLATE_C$	0.00146	0.00463	0.00577	0.00696	0.00979	0.0139	0.0240
	(0.0104)	0.00946	(0.00996)	(0.0112)	(0.0159)	(0.0255)	(0.0530)
$BLATE_A$	0.0101*	0.0101*	0.0101*	0.0101*	0.0101*	0.0101*	0.0101*
	(0.00591)	(0.00591)	(0.00591)	(0.00591)	(0.00591)	(0.00591)	(0.00591)
$BLATE$	0.00715	0.00825*	0.00864**	0.00905**	0.0100***	0.0115*	0.0150
	(0.00575)	(0.00446)	(0.00406)	(0.00375)	(0.00384)	(0.00612)	(0.0151)
Number of Observations	56794	56794	56794	56794	56794	56794	56794

Notes: Standard errors in parentheses are estimated using a bootstrap method. *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.