I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11930

Son Preference and Human Capital Investment among China's Rural-Urban Migrant Households

Carl Lin
Yan Sun
Chunbing Xing

OCTOBER 2018

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11930

Son Preference and Human Capital Investment among China's Rural-Urban Migrant Households

Carl Lin
Bucknell University, China Institute for Income Distribution and IZA

Yan Sun
Beijing Normal University

Chunbing Xing
Beijing Normal University, China Institute for Income Distribution and IZA

OCTOBER 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Son Preference and Human Capital Investment among China's Rural-Urban Migrant Households

We use several datasets to study whether son preference prevails in the human capital investment among Chinese rural-urban migrant households. We find that son preference exists among the rural migrants' households and that it caused lower probabilities relative to that of their boy counterparts that school age girls will migrate with their parents - a difference that is absent for children of preschool age. We also find that (1) boys are more likely to migrate following the reduction in the number of rural primary schools, (2) migrant households with multiple children tend to take their sons to migrate more than they take their daughters, and (3) the fact that parents of boy students spend more on their children's education can be largely explained by the extra costs of schooling for migrant households. Finally, we show that the parents of rural children have higher expectations for boys than they do for girls. Our results suggest that son preference is detrimental to the human capital investment in girls in contemporary China when institutional arrangements result in high costs of schooling for migrants.

JEL Classification: J13, J17, J61, J24

Keywords: rural-urban migration, China, children, son preference, human capital

Corresponding author:
Chunbing Xing
Professor of Economics at Beijing Normal University.
19 Xinjiekouwai Street
Haidian District
Beijing
P.R. China.

E-mail: xingchb@bnu.edu.cn

1. Introduction

Households in many developing countries, such as China and India, prefer having sons rather than having daughters (i.e., son preference), leading to unfavorable consequences such as selective abortion, the ill-treatment of infant girls, an unbalanced sex ratio, and low human capital levels among females. In recent years, however, the socioeconomic status of females has improved substantially. In China, for example, the population census of 2010 indicates that the relative number of women with tertiary degrees reached the level of men for those aged 20-30. Existing research also argues that son preference gradually weakens as economic development and urbanization continue (Chung and Das Gupta, 2007). After decades of unprecedented economic growth, does son preference still prevail in contemporary China? If so, what are the consequences in terms of human capital investment? We empirically investigate these questions by studying the educational opportunities of boys and girls among China's rural-to-urban migrant households.

China has a large and ever growing rural to urban migrant population, reaching 172 million in 2017 (National Bureau of Statistics of China, 2018; NBS hereafter). Although the relaxation of the household registration (or hukou) system since the 1980s has made a large migration flow possible, this system still presents extra burdens for migrants who usually do not have the official hukou of their destination and are thus susceptible of being discriminated against in access to public services, such as enrolling in public schools. A considerable number of children, estimated at approximately 60 million in 2013 (All China Women's Federation, 2013), are left behind in rural areas by their migrant parent(s). At the same time, an increasing number of rural residents migrate with their children despite being discriminated against for their lack of urban hukou. In 2014, there were approximately 13 million migrant children enrolled in primary or middle schools in urban areas (NBS, 2015). An important feature of this group is that the male-to-female ratio reached 1.38 to 1—a ratio much more unbalanced than the national average (1.15 to 1) and that of children left behind (1.20 to 1), suggesting a prevalence of son preference among migrant households.

Migrant households can enhance human capital investment by taking their children with them. First, there are more educational resources in urban China than in rural China, as urban areas are more prosperous and have a better infrastructure. Second, parental supervision and companionship are vital for child development. With resource constraints and the high costs of schooling in urban areas, households having a strong son preference would be more willing to take their sons rather than their daughters along when they migrate, leading to a gender disparity of educational opportunity among rural children.

We use two datasets, the 2005 1% Census and the China Education Panel Survey (CEPS hereafter), to investigate whether the unbalanced sex ratio among migrant students is due to the son preference in human capital investment and the hukou restriction. Using the 1% Census data, we first exploit the school enrollment restriction in cities to show that migrant households prefer taking sons rather than daughters to cities only when schooling concerns set in. To reach this conclusion, we use a sample of rural children of school age (aged 6-12) and below school age (aged 0-5) from migrant households (households with migrant parents) and run regressions to see how gender and age influence whether or not a child migrates. As the gender differential in migration rates may be caused by considerations other than education, we introduce a dummy variable for school age: only at school age do concerns about schooling become valid. The results show that the gender differential in migration rates is significantly larger for school-age children than for children below school age, leading to a greater prevalence of boys in the migrant children of school age than in the migrant children of preschool age.

Next, we use the sharp reduction in the number of rural primary schools in Guizhou and Sichuan as a natural experiment to explore whether there is son preference when rural households respond to the increased schooling costs by taking their children to other regions. Using a Difference-in-Difference (DID) strategy, we establish the fact that boys of primary school age are more likely to respond to the policy than girls are.

Last, using the China Family Panel Studies (CFPS hereafter) data that provides information on the parental expectations of their children's education achievements and the CEPS data which surveyed approximately 20,000 middle school students in 2014, we provide more direct evidence of son preference in human capital investment and exclude some alternative explanations. Using CFPS data, we show that rural parents have higher expectations for their sons than for their daughters. By examining migrant students of different genders and related decisions and behaviors of their parents, we find that (1) classes with more migrant children have more boys; (2) migrant children with school-age siblings left behind in rural hometowns are more likely to be boys, suggesting that parents with multiple children and who are more likely to have girls[1] tend to migrate with their sons; and (3) relative to households with girls, those with boys are more likely to use social connections, pay higher sponsorship fees and are more willing to tackle difficult situations in gaining access to local schools. This difference is dominated by the fact that migrant households are more likely to engage in the above activities to get their children enrolled in school. The results indicate that eliminating discrimination in public services against rural migrants can potentially reduce the sex imbalance among migrant children.

This paper, to our knowledge, is the first to document the unbalanced sex composition among China's migrant children. We exploit the sharp difference between this group and left-behind children, to identify Chinese rural households' son preference in human capital investment. Focusing on the children in migrant households, we also examine how son preference in human capital investment combined with discrimination against migrant workers in local public service provision (or the hukou system) adversely affects the welfare of the female population. As son preference prevails in developing countries and many households face constraints in financing migration and education, examining son preference as a cause of the unbalanced sex ratio and as a factor in human capital investment decisions in the Chinese context offers powerful lessons for other developing countries that are experiencing rapid urbanization.

2. Literature review

Son preference prevails in developing countries and has profound impacts on household or individual decisions, such as fertility, human capital investment, marriage, and savings. As many studies emphasize, through selective abortion or the early death of infant girls from malnutrition and polluted water, son preference has led to tens of millions of missing women (Sen, 1989; 1992; Qian, 2008; Qiao, 2004; Wei et al., 2005; Jayachandran and Kuziemko, 2011). The resultant high male-to-female ratio has increased competition among males in the marriage market, leading to criminal behavior among anxious males, high saving rates of

[1] Evidence from the 2005 mini-census shows that the probability of being a girl increases as the number of children increases.

households with sons, and the tendency for wealthy households to be more able to afford to have sons (Cameron et al., 2017; Wei and Zhang, 2011; Edlund, 1999).

The differential treatment of boys and girls in human capital investment is observed in many studies. Households show apparent son preference in their investment in children's health, which not only reduces the survival rate of infant girls but has an adverse effect on the girls' nutrient intake, height, and other health conditions later in life (Bandyopadhyay, 2003; Song and Burgard, 2008; Song and Tan, 2008; Barcellos et al., 2014). Jayachandran and Kuziemko (2011) point out that son preference leads to shorter breastfeeding of girl infants because the mother is more likely to get pregnant again. Households also show an evident son preference in educational investment. Parents tend to pay more attention to the education of boys, and girls are more likely to suffer the risks of late enrollment and early dropout (Zhou and Yuan, 2014). Gender difference in educational opportunities caused by son preference has led to the lower educational attainment and labor participation rates among females (Wang, 2005).

The impact of son preference on the decision of households is affected by technological, economic, and institutional changes. Some studies point out that the reduction in the cost of gender detection (the prevalence of ultrasound) leads to the decline of newborn girls (see Chen et al., 2013). Qian (2008) and Almond et al. (2003) emphasize the impact of changing economic situations. Qian (2008) shows that the increase in the tea price after the agricultural reform led to higher incomes of women, which in turn resulted in lower mortality and the higher education of girls. Almond et al. (2013) argue that the increased income following the household responsibility system in rural China led to a higher boy to girl ratio because rural households could better afford sex selection and raising boys. Both studies explore the policy change in the early reform period.

Son preference and its impact on human capital investment may gradually weaken with economic and social development (Chung and Das Gupta, 2007). For one thing, economic growth and urbanization have undermined the role of son preference, which is mainly rooted in the traditional agricultural society. For another, rising household income and public investment in education have reduced the costs of education significantly. Both could make son preference obsolete. Contrary to these predictions, Burgess and Zhuang (2002) find that the unfavorable treatment of girls in educational resource allocation exists in both poor and rich regions.[2]

Due to the household registration system in China, migrant households without local hukou face high education costs, as they are discriminated against in public services. This increases the cost of migration and impedes many rural households from migrating. To explore economic opportunities in cities, many rural households have labor age family members who are working in cities and have left their children behind in their hometowns, resulting in a vast number of left behind children. However, rural households are increasingly able and willing to pay extra costs to enroll their children in schools in cities.

A growing number of studies show that their children's education is an important factor that parents consider in migration. For example, Xing and Wei (2017) show that when migrating with children, parents are willing to give up better employment opportunities and higher incomes for access to schools, resulting in significant opportunity costs associated with taking their children to cities for education. Research shows that children moving with their parents are better off in health and academic performance than nonmigrant children are (Resosudarmo and Suryadarma, 2011; Xu and Xie, 2015). Reasons for the better performance of migrant children include better quality education in cities and more parental inputs in the children's human capital investment. Chen and Feng (2013) point out that migrant children in public schools can receive better quality education and that 82.2% of migrant children at primary school age are enrolled in public schools (National Bureau of Statistics of China, 2018).

Our study closely relates to the literature on China's hukou system. Many studies have shown that the hukou system not only hinders regional mobility but also reduces mobility across social hierarchies (Xing and Nie, 2010). Our paper finds that the hukou system affects rural households' investment in human capital unfavorably and that when they are discriminated against in public services, a son preference will lead girls to be treated more unfavorably.

3. Theoretical predictions

Gender preference may result from personal tastes. For example, research shows that

[2] Dahl and Moretti (2008) indicate that son preference even exists in U.S., but they do not consider education opportunities and migration decisions.

households in many societies prefer to have children of different sexes when the number of children exceeds one. Son preference in China and other developing countries may also be a practical matter. The social norm in rural China is that daughters will leave the family when they get married; parents are more likely to live with and obtain livelihood support from sons in their later years. Sons contribute more to their households, as the resource allocation in rural villages (such as land) depends on the number of family members and, in many cases, on the number of sons. Thus, increasing the earning potential of sons through education is more beneficial for parents in the long run. We illustrate this point with a simple human capital investment model similar to Laing (2011).[3]

We assume that the capital market is complete. The exogenously determined interest rate, r, is the marginal cost for investment in education, s. $\mathrm{W}(s)$ is the total revenue from the investment, and $\mathrm{MW}(s)$ is the marginal revenue, which is assumed to be diminishing as s increases. The optimal amount of investment in education is determined by the condition that marginal revenue equals marginal cost ($\mathrm{MW}(s) = r$). Although empirical evidence suggests that the private return to education for females is not lower than that of males, we assume that for parents, marginal revenue of human capital investment for girls ($MW_f = (1 - d)MW$ is less than that of boys ($MW_m = MW$), considering the specific circumstances of Chinese society. d indicates discrimination against girls (i.e., son preference). When $\mathrm{MW}(s) = r$, the level of human capital investment is optimal. As shown in Figure 1, the optimal investment in the human capital of girls is lower than that of boys.

[Figure 1 about here]

Because of the implementation of compulsory schooling and the increase in public investment in rural areas, the required education expenditure is below the optimal investment level. The differential treatment of boys and girls in human capital investment is minimal if they are enrolled in local rural schools. However, when a school age child migrates, extra

[3] We can also analyze the influence of son preference on human capital investment in the framework of consumer choice theory (Ben-Porath and Welch, 1976). Assume that families allocate resources between children and other consumer goods to maximize utility. As sons contribute more to the household income and take more responsibilities in supporting their elderly parents than daughters do, son preference can be interpreted as boys are less expensive than girls. For two identical households, the utility of a household with a girl will be lower than that of a household with a boy. Ben-Porath and Welch (1976) point out that they can lessen the adverse effects of having girls on their utility by reducing their expenditures on girls. Corresponding to our study, this means that migrant families are more reluctant to pay for educational costs when girls migrate.

expenditures (explicit or implicit) related to migration are incurred, which can also be regarded as an investment in education. The extra expenditure causes the households with boys to approach their optimal investment (where $MW(s) = r$); however, for girls, the migration cost is too high; therefore, it is best for the households to stay. Given the case depicted in Figure 1, it is obvious that there is an incentive for rural boys to move to cities, as this is a type of education investment. However, in this paper, we do not distinguish whether the migration is education induced or employment induced.

We can also use this framework to analyze the impact of the reduction in the number of primary schools in rural areas. To achieve the same academic performance, many rural children will have to make more effort (e.g., to commute) and their households will also need to allocate more resources. Thus, we model this impact as the reduction in the marginal returns to human capital investment in rural areas. As the marginal returns to investment decrease in rural areas, it creates an extra incentive for rural residents to move. When the migration costs are large, it is the boys that are more likely to respond to this change.[4]

4. Data and descriptive analysis

We use the 2005 1% Chinese Population Census data and the CEPS data to show son preference in human capital investment among migrant households. The census data covers 2,585,481 individuals in 31 provinces, autonomous regions, and municipalities. We limit the sample to children aged 0-12 with agricultural hukou, of whom children of preschool age (aged 0 to 5) account for 38% (125,513/331,748). Table 1 shows the sample summary statistics in the following aspects: migration status, gender, age, schooling status, and siblings' status. In addition to these variables, the data contains information on the prefectural city where her/his hukou is registered.

[Table 1 about here]

Migrant children are defined as those who have left the place of their hukou registration;

[4] There are other explanations for the differential migration rate among rural children. Wei and Zhang (2011) point out that households will respond to the unbalanced sex ratio caused by son preference in fertility and saving behaviors. The marriage pressure in rural areas may also force rural households to migrate to cities where the sex ratio is less unbalanced. Alternatively, the education in urban areas can be regarded as a way of saving to increase the competitiveness of boys in the future marriage markets. It is also possible that the boys at school age need more supervision from parents or need a role model of the father or that boys are less costly to raise in urban areas. These alternative explanations are excluded in the following analysis.

Left-behind children are those who stay but with both or either parent migrating. To determine his/her left-behind status, we use a variable that indicates the relationship between the child and the household members to match the parental information for each child. Children with either or both parents missing (unmatched) are regarded as left-behind children because migration is a major reason for the difficulty of matching children with their parents. Children with one nonmigrant parent and the other unsuccessfully matched are considered as left-behind children only when the accompanying parent is currently married. Table 1 shows that migrant and left-behind children account for 8% and 48% of the rural children aged 0 to 12. The share of left-behind children may be overestimated because all missing parents are regarded as migrants. We experiment alternative ways of identifying left-behind children in the robustness check and the results do not change much. Table 1 also shows that 40% of the migrant children migrate out of their hukou province; One third of the rural children are the only child, whereas 60% of them are in school and 95% of the primary school age children are in school. The last two columns of Table 1 report the summary statistics for the sample of migrant and left-behind children.

Figure 2 shows the fraction of migrants in rural children of different genders by age. There is no significant difference in the proportion of boys and girls in the preschool period, ranging between 8% and 10%. However, the proportion of migrant children decreases markedly as age increases. The relative number of migrant children falls from 5 to 6 years old as a result of restrictions on school enrollment of migrant children in urban areas. In addition, the relative number of migrant girls in primary schools is significantly lower than that of boys. Correspondingly, there is no significant difference in the sex ratio between migrants and nonmigrants at the preschool age; at the primary school age, the proportion of boys among migrant children becomes significantly higher than that among nonmigrant children.

As Figure 2 is produced using cross sectional data, it is possible that the reduction in migration probability and its gender gap entering primary school age is dominated by cohort effect. If so, no gender gap in migration probability will show up when preschool age children grow into school age. No evidence supports this conjecture. In Figure 3, we use statistics released by the NBS in 2014 and 2015 to calculate the gender ratio for both migrant and left-behind children in primary and middle schools. For both levels of education, the boy to girl ratios for migrant students surpassed those of the left-behind by a sizable margin (approximately 1.4 vs. 1.2), suggesting a differential migration probability between boys and girls of compulsory schooling age.

[Figure 2 about here]

[Figure 3 about here]

Finally, to provide more direct evidence, we group migrant children aged 6-12 by the reasons for migration and calculate the proportion of boys in different groups (see Figure 4). Compared with other groups, the fraction of boys is the largest (approximately 56%) for the group whose primary reason for migration is for study (12.3% of the total sample of migrant children).

[Figure 4 about here]

The above descriptive analysis suggests that parents prefer to send primary school age daughters home and have their sons with them, paying higher costs of schooling for sons in urban areas. Since migration is affected by selection effects, family structure, age effects, and the socioeconomic characteristics of the hukou location, we provide our empirical analysis of son preference in human capital investment for rural household migration in the next section.

5. Empirical analysis

5.1. Model

We use a DID strategy to identify son preference in human capital investment in rural migrant households by estimating a Linear Probability Model (LPM):

$$migrant_i = \beta_0 + \beta_1 male_i + \beta_2 male_i \times schage_i + \beta_3 schage_i + \gamma X_i + \mu_i \quad (1)$$

where *migrant* is a dummy indicating whether a child had left his/her hukou location. *male* denotes gender and *schage* indicates whether a child is of school age; *male*×*schage* is the interaction of the two dummy variables. X is a vector of controls including age, siblings (whether from a one-child family), and city dummies of hukou. μ is the error term. β_1, β_2, and β_3, are the coefficients of the variables *male*, *interaction*, and *schage*, respectively. γ is the coefficient vector of the control variables. i denotes individual observations.

Next, we briefly discuss the dependent variable *migrant*. As most 0-12-year-old children migrate with their parents if the parents migrate, *migrant*=1 can be understood as parents migrating with their child. In the case of *migrant*= 0, the exact meaning depends on how we define our sample. When no restrictions are made on the sample, it includes both left-behind and nonleft-behind children. When we limit the sample to children in migrant households, observations with *migrant*= 0 refer to left-behind children. To better capture son preference in the migration by the gender dummy (*male*), we limit the sample to children in migrant families. Such a limited sample helps solve the selection problem of family migration to a certain extent. If families with boys are more likely to migrate or if migrant families are more likely to come from areas with higher boy-to-girl ratios, it will also lead to a higher boy-to-girl ratio among the left-behind children.

Migration also has different meanings for children of different ages. For preschool children, parents migrate with them mainly to live with and take care of them. However, for children of school age, parents taking children with them would have to arrange for school enrollment in their places of residence, which is not only mandatory by law but in line with the interests of the households.[5] Therefore, a son preference of the parents of school-age children includes a general son preference and a son preference in schooling. For this reason, we add the school-age dummy variable (*schage*) and its interaction with the gender variable (*male*×*schage*) to the model. The former coefficient reflects the influence of enrollment restrictions on the probability of migration, and the coefficient of the interaction reflects son preference in schooling under such constraints.

To better understand the identification strategy, we let *schage*=0 in (1), and the model becomes $migrant = \beta_0 + \beta_1 male + \gamma X + \mu$. This is equivalent to considering only the preschool children, and β_1 reflects son preference that does not include the schooling concern. When we only consider school-age children (*schage*=1), the model becomes $migrant = \beta_0 + \beta_3 + (\beta_1 + \beta_2) male + \gamma X + \mu$. The coefficient of *male* includes both general son preference and son preference in education. It is by considering the difference in son preference between these two groups ($(\beta_1 + \beta_2) - \beta_1) = \beta_2$, the coefficient of the interaction term, that we

[5] There are few observations of children aged 6-12 who are not in school, but whether or not we consider these observations, our results are unaffected.

identify son preference among migrant households in human capital investment.

Our empirical strategy also faces some challenges. First, school age and preschool age children are of different ages, and their tendencies to migrate with their parents may be different, even without human capital considerations. Moreover, since we are using cross-sectional data, children of different ages belong to different birth groups, and their sex ratios and migration probability may vary. To address such concerns, we control for the age of the children.[6] Due to the strong correlation between age and school-age dummy variables, we do not control the interaction between *male* and age. Second, there may be differences in the number of siblings among children of different age groups and regions, and having siblings is likely to affect the decision of parents to migrate with their children. For this reason, we control for whether the observation comes from a one-child family. Third, we control for the dummies of the location of the children's household registration because many regional factors (such as the level of economic development and marital market conditions) may affect migration decisions and may also relate to the probability of families having boys.

5.2. Results

Table 2 reports the regression results from using the sample of migrant and left-behind children with rural hukou. Column 1 shows the results without control variables. The coefficient of *male* is not statistically significant, indicating that migrant households do not show son preference when an education concern is absent (*schage*=0). Compared to that of preschool children, the migration probability of primary school children declines. For girls, the probability (relative to preschool) decreases significantly by 3.9 percentage points. The coefficient of the interaction term suggests that the migration probability of boys is significantly higher than that of girls by 1.2 percentage points. This is one of the key results of this paper. It shows that the enrollment restrictions on migrant children significantly reduce the probability of rural households moving with their children. A larger impact on girls than on boys reflects a son preference in educational investment among rural households.

[Table 2 about here]

In column 2, we include the children's ages to control for the influence of sex ratios in

[6] We have shown by Figure 3 that cohort effect is unlikely to dominate.

different birth groups. Due to the high correlation between age and the school-age group, the magnitude and the precision of the coefficient of *schage* declines, while the coefficient of the interaction term stays unchanged. In column 3, we control for whether a child has siblings. The interaction coefficient becomes 0.005, and it is insignificant. In column 4, we include dummies of the prefecture city of hukou, and the coefficient of the interaction term becomes significant at 0.007. Wei and Zhang (2011) show that the unbalanced sex ratio in rural China leads men to face more competition in the marriage market. To respond to this pressure, rural households with boys save more money to increase their sons' competitiveness in the marriage market. Similarly, rural households may respond to this competition by taking sons to urban areas for education. If the competition mainly comes from the region of the migrants' hukou origins (assuming that marriage partners are from the hukou region), the results in column 4 indicate that characteristics of the household registration region cannot explain son preference in rural migrant families. This means that son preference in household migration is not a response to the unbalanced sex ratio.

5.3. Heterogeneity analysis

As fiscal transfers for education expenditures from the central government are based on the size of the population with local hukou, local governments lack incentives to provide education for migrant children. Households face stronger institutional constraints and higher costs in interprovince than in intraprovince migration. Therefore, households are more likely to show son preference in human capital investment in interprovincial migration. In columns 1 to 4 of Table 3, we classify migrant children into two groups, intraprovincial and interprovincial migrants, and compare them with left-behind children.

In the interprovincial case, the coefficient of school age is -0.034 when other controls are not added, indicating that the probability of interprovincial migration of school-age children (relative to preschool) reduces significantly. The coefficient of *schage*×*male* is positive, showing the existence of son preference in the educational investments parents make. After controlling for age, siblings, and the city of hukou, the effect of school age decreases (which is mainly caused by controlling age), but the coefficient of the interaction term varies little. In comparison, the coefficient of the interaction term in the intraprovincial migration estimation is not significant after controlling for these variables. The absolute values of the coefficients on school age in the first two columns are greater than those in columns 3 and 4, suggesting that interprovincial migrant families face stronger constraints to their children's enrollment and son preference is more noticeable in such cases. The results here is also consistent with the pattern presented in Figure 3 where interprovincial migrant students show the higher boy to girl ratio than intraprovincial ones do.

[Table 3 about here]

To study the differences of son preference in migrant households with different numbers of children, we estimate the effects for children from different types of households and report the results in columns 5 to 8 in Table 3. When considering children from one-child households, the coefficient of the interaction item is positive but insignificant regardless of whether control variables are added. The migration probability of boys in primary school is about one percentage higher than that of girls. In the last columns, we identify those from households with at least one boy and one girl. While we know whether a child has brothers or sisters, we may not have information on each child within a household. To ensure that all children within the considered households are below 18 years of age, we limit the sample to children whose mothers' ages are between 20 and 37. The last two columns show that both the school age variable and the interaction item have a greater impact on migration than those in the first two columns do. In particular, the migration probability of school-age boys is 3.2 percentage points higher than that of girls. This suggests that son preference in multichild households is more apparent. We provide more evidence on this in section 8, using the CEPS data.

5.4. Alternative explanations and robustness check

In this section, we provide further evidence that boys of school age are more likely to migrate, but not because of gender differences in children's personalities and other motivations. For the relatively high probability of migration for boys, a possible explanation is that boys are naughtier than girls and parents' absence is more likely to have a negative effect on their academic achievement. Therefore, parents are more likely to bring them along when they reach school age. To investigate this possibility, we delete observations with both parents migrating. In this case, children staying in the countryside will have parental supervision and companionship. Columns 1 to 3 in Table 4 reports the corresponding results. The coefficient of the interaction term between the variables *male* and *schage* is lower than that in Table 2 and becomes insignificant. However, when we consider the children from households with multiple children and compare interprovincial migrant with left-behind children, the coefficients on the interaction term become significant. It shows that when the difference in personality between boys and girls is considered, the result that boys (relative to girls) of school age are more likely to migrate still holds.

[Table 4 about here]

The higher probability of migration for rural boys might be part of the preparation of rural households' children for the urban labor market. This alternative explanation can be ruled out based on the following evidence. First, many studies demonstrate that the returns to education are higher for females than for males in China's urban labor market (Liu, 2008; Chen and Hamori, 2009) and that the higher the level of female schooling is, the higher the returns to female schooling compared to that of male schooling (Rosenzweig and Zhang, 2013). Second, recent studies show that younger females are in higher demand among employers than their male counterparts are (Helleseter *et al.*, 2016). Third, we find that the rural females aged 16-25 are more likely to migrate to cities than their male counterparts are (see Figure 5).

[Figure 5 about here]

To see the robustness of our results, we first extend the school-age children in school from primary to middle school in columns 4 to 6 in Table 4. Compared with the results in Table 2, including middle school age children only slightly reduces the coefficient of the interaction item, and again, when we consider the children from households with multiple children and when we compare interprovincial migrant children with left-behind children, the coefficients on the interaction term become significant.

In the previous analysis, nonmigrant children with both or either parent's migration status unknown (because of unsuccessful matching) are treated as left behind children, which might cause left-behind children to be overrepresented. In this section, we use alternative definitions of left-behind children to examine the robustness of our findings.

In our first alternative, nonmigrant children with both parents missing in the matching process are still regarded as left-behind children, but those with one nonmigrant parent and the other unmatched are not regarded as left-behind children and are deleted in the regressions. The estimated results of model (1) are reported in Part A of Table 5 and suggest a larger impact of son preference than that in Table 2. While there is no significant gender difference in migration probability among preschool children, the migration probability of boys at primary school age is approximately 1 to 2 percentage points higher than that of their girl counterparts. In Part B of Table 5, nonmigrant children with both parents missing in the matching process are not regarded as left-behind children, while those with one nonmigrant parent and the other unmatched are regarded as left-behind children (the nonmigrant parent needs to be married). The regression results are similar to those in panel A when other variables are not controlled for, but the coefficient of the interaction term turns insignificant when other variables are controlled for. Finally, in panel C, when both parents are missing in the matching procedure or when one parent is missing and the other is at home, nonmigrant children are not regarded as left-behind children. Although the coefficient of the interaction term is insignificant, it is positive and similar in magnitude to the results previously reported. Therefore, our results do not change qualitatively when we use different definitions to identify left-behind children.

[Table 5 about here]

6. Closures of rural primary schools and the migration of children: Evidence from a natural experiment

In this section, we use the large-scale school closures in rural China in the early 2000s as a natural experiment to examine son preference in rural households' migration and human capital investment decisions. To improve the quality (in particular hardware) of rural education under limited fiscal capacity, in response to the encouragement of the central government in 2001, many village primary schools or teaching spots were merged into larger ones that were usually located in towns (Liu and Xing, 2016).[7] Local governments often made arbitrary decisions without consulting local residents and implemented them hastily. For many rural households, the closing of nearby schools increased the monetary and nonmonetary schooling cost as the students either had to commute long distances without school buses or were forced to board in school, which led to the absence of parental caring and supervision. Meanwhile, no

[7] See Decisions on the Reform and Development of Primary Education, the State Council, 2001.

evidence suggests that school mergers improved the students' academic performance. Therefore, the school closures reduced the marginal return to educational expenditures as we discussed in the theoretical analysis. School mergers in rural China provided incentives for rural residents to migrate for their children's education. Liu and Xing (2016) show that the reduction in rural primary schools increased the probability of migration for rural residents, particularly those with primary school age children. We further this study by examining the policy's differential impact on the migration of boys versus girls, which reflects the parents' gender preference in human capital investment.

It is also possible that the school closure policy at the local level is a response to the reduction in the number of students due to the large-scale out-migration. To avoid the endogeneity problem, we exploit the sharp difference in the implementation of this policy in specific provinces. As the implementation of the policy is at the discretion of local governments, it shows considerable variations in time and extent. One noticeable province is Guizhou. Documents released by Guizhou government indicate that it implemented the school closure program hastily approximately 2001. Panel A in Figure 6 shows that the number of rural primary schools in Guizhou declined sharply by 15% between 2000 and 2001, and the reduced number accounted for 44% of the total reduction in the number between 1995 and 2005. The sharp reduction of primary schools in Guizhou offers a natural experiment to examine how the increased schooling costs drove the residents to move. However, to establish the causal linkage is still difficult as (1) there may be other changes that happened at the same time, such as China's entry into the World Trade Organization, which increased the demand for migrants in coastal regions, and (2) the closure of primary schools might be a response (either continuous or discontinuous) to the loss of the pupils due to the massive rural to urban migration and the rural residents' actively seeking a better education in urban areas. To alleviate these concerns, we use Yunnan as a control group and use a DID strategy for identification. Yunnan is a bordering province with an economic development level similar to that of Guizhou, but it implemented large-scale school closures only approximately 2010. As is clear, the number of primary schools decreased continuously in Yunnan, Guizhou's neighboring province. Importantly, there is a similar decreasing trend in the number of primary schools for both provinces before 2000, a trend that has continued for Yunnan after 2001 but stopped for

Guizhou.

[Figure 6 about here]

We examine the migration probability of primary school age children directly, and we link the rural households' migration decision and the school closing policy by examining the number of primary school migrants, that is, those who were at primary school age (5-12) when they first migrated, before and after the policy. For example, the primary school migrants for 2000 are those aged 10 to 17 in 2005 who have left their hukou registration place for 5 to 6 years: their households' decision to migrate was unaffected by the policy. In contrast, the primary school migrants for 2003, who are of age 7 to 14 in 2005 and have migrated for 2 to 3 years, were under the influence of the policy if they were from Guizhou. To allow for a year of adjustment due to the hasty manner of the policy, primary school migrants for 2002 are regarded as mostly unaffected by the policy.[8] See Table 6 for a detailed illustration.

[Table 6 about here]

Our strategy compares the relative amount of primary school migrants between Guizhou and Yunnan, before and after the policy, and the model is as follows:

$$\begin{aligned} pscmigrant_{ict} = \beta_0 + \beta_1 post_{ict} + \beta_1 GZ_{ic} + \delta post_{ict} \times GZ_{ic} \\ + \gamma_1 age_{ic} + \gamma_2 age2_{ic} + \lambda year_{it} + c_i + \mu_{ict} \end{aligned} \quad (2)$$

We estimate the model by using a sample of migrants from Guizhou and Yunnan aged below 60, and as before, migrants are defined as those who have left their hukou registration place for over six months. The dependent variable *pscmigrant* is a dummy for being a migrant who left the hukou registration place at the primary school age; *post* is dummy variable indicating that an observation left his/her hukou place after 2002; *GZ* is a dummy indicating that one is from Guizhou, and *post*×*GZ* is an interaction of these two dummies. We control for a quadratic term of age to account for the fact that migration probability changes with age, and we assume that the change is continuous. We also control for the time when an individual left his/her hukou registration (*year*) to account for the fact that the migration probability changes over time. Finally, the dummies of the origin cities for each observation are also controlled for.

We report the results in Table 7. When the sample is not restricted (columns 1 to 3), the

[8] Treating this group as affected by the policy reduces the significance level but does not change the conclusion.

coefficients on the interaction term are positive, suggesting that the sharp decline in the number of primary schools in rural Guizhou has increased the probability of migration among those of primary school age. However, the coefficients are not statistically significant, although there is indication that the effect is slightly larger for male than for female students. In columns 4 to 6 of Table 7, we restrict the sample to migrants whose host regions were different from their city of hukou registration. When we control for the fixed effects of birth cities (column 4), boys of primary school age are relatively more likely to migrate to other cities by 5.4 percentage points after the sharp decline in primary schools and the effect is significant at the 5% level, while for girls, the effect is only approximately 0.5 percentage points and statistically insignificant. The hypothesis that these two coefficients are equal is rejected at the 10% level. Controlling for a common age effect and the time effect (column 5) does not change the results much. To allow the age and time effects to vary across cities actually produces a larger coefficient for the male sample and a smaller coefficient for females (column 6). Comparing the first and last three columns of Table 7 suggests that after nearby primary schools are closed, parents are more ready to bear the cost of migration to other cities for boys than for girls.

To alleviate the concern that the pattern presented above is just reflecting an existent trend of migration, we perform a placebo test, using the census data for 2000, a year when the school closure policy was not effective. Using Guizhou and Yunnan as the treated and control groups and classifying those migrated between 1998 and 2000 as the after group and those between 1996 and 1997 as the before group, we run similar regressions as (2) for boys and girls. All coefficients of the interaction term ($post \times GZ$) either for boys or girls are insignificant, suggesting an absence of a preexisting trend (see Table 14 in the appendix).

[Table 7 about here]

Panel B of Figure 6 shows that Sichuan also witnessed a significant drop in the number of primary schools in 2001 and before that, the number of primary schools was roughly parallel to that of Yunnan. We therefore estimate a model similar to (2), using the migrant sample from Sichuan and Yunnan. The results are reported in Table 8, and the pattern emerges therein is similar to that in Table 7: boys at the primary school age are more likely to migrate to other cities than girls are after the number of rural primary schools declined sharply.

[Table 8 about here]

7. Parental expectations for children's educational achievements

In this section, we use the CFPS data in 2016 to examine son preference in the parental expectations for children's education. The CFPS data covers 25 provinces and collects information on the parental expectations for their children's highest education levels. We focus on children aged 0-15 with an agriculture hukou and examine how the parental expectations differ by their children's gender.

We run a linear probability model to examine whether a child's parents expect him/her to obtain a college degree (yes=1/otherwise=0). Columns 1 to 2 in Table 9 show that for preschool children (aged 0-5), parents have higher expectations for boys than for girls. The probability of the parents' expectation that their sons will go to college is 5.2 percentage points higher than that of their expectation that their daughters will. Controlling for children's age, parental education levels, net family income per capita, and county dummies does not change this association. For primary and middle school children, their parents' expectation is not significantly different between boys and girls (columns 3 to 4). However, when we control for children's academic performance, parents show a significantly higher expectation for boys than for girls. The probability of the parents' expectation that their sons will go to college is 3.7 percentage points higher than that of their expectation that their daughters will (column 5). This suggests that the girls' good academic performance can influence the parents' educational expectation.

[Table 9 about here]

8. Evidence from CEPS data

The CEPS data allows us to look into gender preferences in education among migrant households in more detail. The data covers 438 seventh or ninth grade classes in 112 schools in 28 counties/neighborhoods. All students in the sampled classes and their parents and teachers are surveyed, and information on students' personal characteristics, academic performance, family backgrounds, and parental attitudes is collected. Essential for this study, the survey collects information on the migration status of the students, which allows us to examine the human capital investment behavior of migrant households.

First, consistent with census data, we show that migrant students are more likely to be boys than nonmigrant students are. While classes in most local public schools have few migrant students, a few classes have a sizable share of migrants, and the higher the share of migrant students is, the higher the share of boy students in the class. A nonparametric analysis confirms this relationship (see Figure 7). In the linear estimations reported in Table 10, a 10% increase in the share of migrant students is associated with a 6% increase in the share of boys. Controlling for city dummies strengthens this association, but it disappears once school dummies are controlled for. This suggests that the migration-induced gender imbalance is mainly a school level phenomenon and that the gender composition varies considerably among schools within counties. Consistent with these results, columns 1 to 2 in Table 11 shows that compared with those who are born in the county of the school locality, those born outside of the school county are 4 percentage points more likely to be boys. Controlling for a rich set of variables, including school fixed effects and parental characteristics, does not change this coefficient considerably.

[Figure 7 about here]

[Table 10 about here]

[Table 11 about here]

As we mentioned earlier, migrant children are more likely to be boys because migrant households have a son preference in human capital investment not because of the selection effect. To provide further evidence, we use one more piece of information from CEPS. The survey asked whether a migrant student had siblings who were left behind to attend school at home. The regression results in columns 4 and 5 of Table 11 show that compared with other migrant counterparts, students with left-behind siblings are 7 percentage points more likely to be boys. This difference cannot be explained by factors such as academic performance, family structure, household income, parent education, region dummies, and so forth, suggesting that rural households prefer to take boys when they migrate. An alternative explanation is that boys of middle school age need more parental supervision than girls do. To assess this possibility, we drop migrant students who live with their parent(s). If migrant students of this subgroup are still more likely to be boys, it cannot be explained by the supervision consideration. Column 3 in Table 11 shows that the coefficient of the birthplace out of county is even larger at 0.072 and is significant at 10% level, suggesting that the larger fraction of boys in migrant students is not due to supervision concerns.

If households have a stronger preference for boys' education, will this preference be reflected at the intensive margins as well? Will it extend to pecuniary costs (like school choice fees and fees for after school programs), use of social connections, and time and effort to enroll the migrant student in school and other activities? Columns 1 and 4 in Table 12 show that households with boy students are significantly more likely to take more measures in sending their children to school. Controlling for migration status (columns 2 and 5) reduces the coefficient of *boy* slightly, and the migrant status has a much larger effect: parents of migrant students are more likely to make efforts and face difficulties than nonmigrant parents are by over 10 percentage points. The coefficients of the interaction term (*boy*×*migrant*) are mainly insignificant. For one thing, it indicates that there is no significant gender difference for a given migrant status. For another, the fact that the coefficient of *boy* turns insignificant (due to large standard errors) might be due to collinearity between the variables *boy* and *boy*×*migrant*. In general, these results indicate that the extra costs and effort involved in sending children to school are mainly the result of the hukou system. The results in Table 13 examine more details and regress a variety of human capital investment behaviors related to gender, migrant status, and their interactions. The results show that it is the migrant status that determines household behavior. Thus, the evidence from CEPS supports our previous findings that son preference manifests because the hukou system results in higher costs of schooling for migrant households.

[Table 12 about here]

[Table 13 about here]

9. Conclusion

China is experiencing unprecedented urbanization and the largest internal migration flow in history. Due to institutionalized restrictions, such as the household registration system and the financial constraints for education, China' rural-urban migrant households often face higher education costs in destination regions. These costs may be direct (such as school choice fees) or indirect (migrant households may give up good job opportunities for their children's education). Son preference in human capital investment among rural households leads to a higher probability of migration for boys than for girls. We use a DID strategy to estimate the effect of son preference on the migration probabilities of rural children by different gender. The results show that son preference reduces the migration probability of girls in the corresponding school age by 1 percentage point. If the number of migrant and left-behind children at the compulsory education stage is 40 million (which is likely a conservative number), our back-of-the-envelope calculations suggest that there will be 200,000 rural girls who do not move to urban areas with their parents because of son preference. When son preference is not considered, the results also imply that approximately 1.40 million rural children are not moving to urban areas with their parents due to restricted access to public education.

Although China has made remarkable achievements in gender equality and the educational level of females has improved, this study shows that disadvantageous institutional arrangements will hinder the improvement of girls' welfare. As households are more likely to show son preference when they face strong constraints, recent policies of population control carried out in some big cities may have a more adverse impact on girls (Xiong, 2016). Our results show that increasing the provision of education services and reducing barriers for migrants can improve the human capital of their children and reduce gender disparity in education in migrant households.

The high share of boys among migrant children as a result of household registration and son preference might be detrimental to the development of migrant children. Studies have shown that both males and females perform worse in math when there are more boys in classrooms (Hoxby, 2000) and that an increase in the number of girls improves the boys' and girls' cognitive outcomes (Black et al., 2013). These academic gains are due to lower levels of classroom disruption, improved interstudent and student-teacher relationships, and lessened teachers' fatigue (Lavy and Schlosser, 2011). Because migrant children are usually concentrated in schools established for migrant children, the higher boy-to-girl ratio among migrant children can potentially have a negative impact on the migrant children's academic performance.

References

All-China Women's Federation, 2013, "Report on the Status of Rural Migrant and Left-Behind Children in China".

Almond, D., H. Li, and S. Zhang, 2013, "Land Reform and Sex Selection in China", NBER Working Paper, No.19153.

Bandyopadhyay, M., 2003, "Missing Girls and Son Preference in Rural India: Looking beyond Popular Myth" , *Health Care for Women International*, 24(10):910-926.

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): 157-189.

Ben-Porath, Y., and F. Welch, 1976, "Do Sex Preference Really Matter?" *Quarterly Journal of Economics*, 90(2): 285-307.

Black, S. E., P. J. Devereux and K. G. Salvanes, 2013, "Under Pressure? The Effect of Peers on Outcomes of Young Adults", *Journal of Labor Economics*, 31(1):119-153.

Burgess, R., and J. Z. Zhuang, 2002, "Modernization and Son Preference in People's Republic of China", ERD Working Paper, NO. 20.

Cameron, L. A., X. Meng, and D. Zhang, 2017, "China's Sex Ratio and Crime: Behavioral Change or Financial Necessity?", *The Economic Journal*, 1-31.Chen, G. F., and S. Hamori, 2009, "Economic Returns to Schooling in Urban China: OLS and the Instrumental Variables Approach", *China Economic Review*, 20(2): 143-152.

Chen, Y. Y., and S. Z. Feng, 2013, "Access to Public Schools and the Education of Migrant Children in China", *China Economic Review*, 26(1): 75–88.

Chen, Y., H. Li, and L. 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.

Chung, W., and M. Das Gupta, 2007, "The Decline of Son Preference in South Korea_ The Roles of Development and Public Policy", *Population and Development Review*, 33(4): 757-783.

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

Edlund, L., 1999, "Son Preference, Sex ratios, and Marriage Patterns", *Journal of Political Economy*, 107(6): 1275-1304.

Helleseter Miguel Delgado, Peter Kuhn, Kailing Shen, 2016, "The Age Twist in Employers' Gender Requests: Evidence from Four Job Boards," NBER Working Paper, No. 22187.

Hoxby C., 2000, "Peer Effects in the Classroom: Learning from Gender and Race Variation", NBER Working Paper, No. w7867.

Jayachandran, S., and I. Kuziemko, 2011, "Why Do Mothers Breastfeed Girls Less than

Boys? Evidence and Implications for Child Health in India", *Quarterly Journal of Economics*, 126(3): 1485-1538.

Laing, D., 2011, *Labor Economics: Introduction to Classic and the New Labor Economics*, W. W. Norton & Company

Lavy, V., and A. Schlosse, 2011. "Mechanisms and Impacts of Gender Peer Effects at School", *American Economic Journal: Applied Economics*, 3(2):1-33.

Liu, J., and C. B. Xing, 2016, "Migrate for Education: An Unintended Effect of School District Combination in Rural China", *China Economic Review*, 40: 192-206.

Liu, Z. Y., 2008, "Why the Returns to Education are Higher for Females?" *Economic Science* (*Jijingji Kexue*), No. 2.

National Bureau of Statistics of China, 2015, "China Statistical Yearbook", http://www.stats.gov.cn/tjsj/ndsj/2015/indexch.htm.

National Bureau of Statistics of China, 2018, "Survey Report on Migrant Workers", http://www.stats.gov.cn/tjsj/zxfb/201804/t20180427_1596389.html.

Qiao, X. C., 2004, "Sex Preference, Sex Selection, and Sex Ratio at Birth", *Chinese Journal of Population Science*, (1): 14-22.

Qian, N., 2008, "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance", *Quarterly Journal of Economics*, 123(3): 1251-1285.

Resosudarmo, B., and D. Suryadarma, 2011, "The Effect of Childhood Migration on Human Capital Accumulation: Evidence from Rural-Urban Migrants in Indonesia", ANU Working Papers in Trade and Development, No. 2011/02.

Rosenzweig, M. R., and J. S. Zhang, 2013, "Economic Growth, Comparative Advantage, and Gender Differences in Schooling Outcomes: Evidence from the Birthweight Differences of Chinese Twins", *Journal of Development Economics*, 104(3): 245-260.

Sen, A., 1989, "Women's Survival as A Development Problem", *Bulletin of the American Academy of Arts and Sciences*, 43(2): 14-29.

Sen, A.,1992, "Missing Women", *British Medical Journal*, 304(6827): 587-588.

Song, Y. P., and L. Tan, 2008, "Son Preference and Sex Differentials in Children's Health: A Study Based on Rural Family Planning Policy Environment", *Population Research*, 32(3): 67-74.

Song, S., and Sarah A. Burgard, 2008, "Does Son Preference Influence Children's Growth in Height? A Comparative Study of Chinese and Filipino Children", *Population Studies*, 62(3): 305-320.

Wang, W., 2005, "Son Preference and Educational Opportunities of Children in China—I Wish You Were a Boy!", *Gender Issues*, 22(2): 3-30.

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

Wei, Y., S. Z. Li, and W. Feldman, 2005, "Boy Preference and Induced Abortion in Rural China: Findings from the 2001 Survey", *Chinese Journal of Population Science*, (2):12-21.

Xiong, B. Q., 2016, "'Education Controller' Casts A Shadow over Educational Equity", http://opinion.people.com.cn/n1/2016/1025/c1003-28804304.html.

Xing, C. B., and Y. H. Wei, 2017, "Does Migrating with Children Influence Migrants' Occupation Choice and Income?", *Journal of the Asia Pacific Economy*, 22(1): 156-172.

Xing, C.B., and H. F. Nie, 2010, "When an Urban Boy Falls in Love with a Country Girl: Baby's Hukou, Hukou Reform and Inter-Marriage", *World Economic Papers*, (4): 1-18.

Xu, H. W., and Y. Xie, 2015, "The Causal Effects of Rural-to-Urban Migration on Children's Well-being in China", *European Sociological Review*, 31(4):502-519.

Zhou, Q., and Y. Yuan, 2014, "Theoretical and Empirical Study of Gender Inequality in the Family's Compulsory Educational Investment", *Population Journal*, 36(4):14-24.

Table 1 Summary statistics (children aged 0-12)

variable	All children (including nonleft-behind) N=331,748		Migrant and left-behind Children N=183,324	
	Mean	Std. dev.	Mean	Std. dev.
Migrant child	0.075	0.264	0.136	0.343
Intraprovincial migrant	0.045	0.208	0.082	0.275
Interprovincial migrant	0.030	0.170	0.054	0.226
Left-behind child	0.477	0.499	0.864	0.343
Male	0.526	0.499	0.529	0.499
In school	0.591	0.492	0.592	0.491
In school (not include 0-5 children)	0.951	0.216	0.955	0.208
Age	6.736	3.706	6.723	3.679
Only child	0.329	0.470	0.353	0.478

Note: Migrant children are defined as children who leave the household registration (hukou) place. The left-behind children are those who have not left the hukou place and either or both of their parents have migrated. Children who have not migrated and for whom we cannot match both or either parent observations are also regarded as left-behind children.

Table 2 Son preference in rural migrant families (0-12-year-old children，LPM)

Dependent variable: migrant=1/ left-behind=0				
	(1)	(2)	(3)	(4)
male	-0.004	-0.004	-0.003	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)
schage	-0.039***	-0.011***	-0.005	-0.006
	(0.002)	(0.004)	(0.004)	(0.004)
male×schage	0.012***	0.012***	0.005	0.007**
	(0.003)	(0.003)	(0.003)	(0.003)
Age	no	yes	yes	yes
Only child	no	no	yes	yes
City of hukou	no	no	no	yes
Obs.	183,324	183,324	183,324	183,324
Adj_R^2	0.002	0.003	0.008	0.046

Note: Robust standard errors are in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table 3 Heterogeneous effects of son preference

Dependent variable: migrant=1/ left-behind=0								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
variable	Interprovincial vs. left-behind		Intraprovincial vs. left-behind		Only child		Multiple children with both gender	
male	0.000	-0.003	-0.004*	-0.003	-0.005	-0.006	-0.013	-0.022***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.008)	(0.008)
schage	-0.034***	-0.006**	-0.012***	-0.002	-0.023***	-0.006	-0.030***	-0.038***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.005)	(0.007)	(0.007)	(0.010)
male×*schage*	0.007***	0.006**	0.007**	0.002	0.010	0.008	0.021**	0.032***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.006)	(0.006)	(0.010)	(0.010)
Age	no	yes	no	yes	no	yes	no	yes
City of hukou	no	yes	no	yes	no	yes	no	yes
Only child	no	yes	no	yes				
Obs.	168,239	168,239	173,463	173,463	64,787	64,787	27,081	27,081
Adj_R^2	0.004	0.067	0.000	0.046	0.001	0.045	0.001	0.078

Note：In the last two columns, the mother's age is limited to 20-37 to ensure that all children are below 18 years old. Robust standard errors are in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table 4 Gender preference in migrant children in all stages of compulsory education and within households without both parents migrating

Dependent variable: migrant=1/ left-behind=0						
	Excluding children with both parents migrating			Aged 0-15 (children age 6 to 15 in school)		
	all	Multiple children w/ both gender	Interprovince vs. left-behind	all	Multiple children w/ both gender	Interprovince vs. left-behind
	(1)	(2)	(3)	(4)	(5)	(6)
male	-0.005*	-0.011**	-0.004**	-0.004	-0.022***	-0.002
	(0.003)	(0.006)	(0.002)	(0.003)	(0.008)	(0.002)
schage	-0.011***	-0.023***	-0.009***	-0.016***	-0.035***	-0.004*
	(0.004)	(0.007)	(0.003)	(0.003)	(0.010)	(0.002)
male×*schage*	0.004	0.014**	0.005**	0.005	0.032***	0.005**
	(0.003)	(0.007)	(0.002)	(0.003)	(0.010)	(0.002)
age	yes	yes	yes	yes	yes	yes
Only child	yes	—	yes	yes	—	yes
City of hukou	yes	yes	yes	yes	yes	yes
Obs.	111,710	23,174	106,195	236,902	31,620	216,952
Adj_R^2	0.033	0.035	0.047	0.044	0.072	0.064

Note: Based on Table 2, the samples in which both parents migrate are omitted. Robust standard errors are in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table 5 Robustness check with different definitions of left-behind children

Dependent variable: migrant=1/ left-behind=0						
	(1)	(2)	(3)	(4)	(5)	(6)
	A		B		C	
male	-0.005	-0.003	-0.005	-0.006*	0.002	-0.002
	(0.005)	(0.005)	(0.004)	(0.004)	(0.007)	(0.007)
schage	-0.054***	-0.006	-0.051***	-0.010**	0.025***	-0.018*
	(0.005)	(0.007)	(0.003)	(0.005)	(0.007)	(0.011)
male×*schage*	0.017***	0.014**	0.017***	0.006	0.013	0.011
	(0.006)	(0.006)	(0.005)	(0.004)	(0.010)	(0.009)
Age	no	yes	no	yes	no	yes
Only child	no	yes	no	yes	no	yes
City of hukou	no	yes	no	yes	no	yes
Obs.	89,570	89,570	128,669	128,669	34,915	34,915
Adj_R^2	0.002	0.068	0.003	0.074	0.001	0.086

Note: (1) Parts A to C correspond to different definitions of left-behind children. In part A, children with both parents missing in the matching process are regarded as left-behind children. Those nonmigrant children with one nonmigrant parent and the other unmatched are not regarded as left-behind children and are deleted in the regressions. In part B, we reverse the practice in part A: children with both parents missing in the matching process are not regarded as left-behind children and are deleted. Nonmigrant children with one nonmigrant parent and the other unmatched are regarded as left-behind children (the nonmigrant parent needs to be married). In part C, when both parents are missing in the matching procedure or when one parent is missing and the other is at home, children are not regarded as left-behind children. (2) Robust standard errors are in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table 6 Defining primary school student migrants, using age and migration duration

migration duration (r8=)	<1yr. (2/3)	1-2yr. (4)	2-3yr. (5)	3-4yr. (6)	4-5yr. (7)	5-6yr. (8)	>6yr. (9)
Yr. migration	2005	2004	2003	2002	2001	2000	1999-
Age in 2005	age at migration						
17	17	16	15	14	13	**12**	11
16	16	15	14	13	**12**	**11**	10
15	15	14	13	**12**	**11**	**10**	9
14	14	13	**12**	**11**	**10**	**9**	8
13	13	**12**	**11**	**10**	**9**	**8**	7
12	**12**	**11**	**10**	**9**	**8**	**7**	6
11	**11**	**10**	**9**	**8**	**7**	**6**	5
10	**10**	**9**	**8**	**7**	**6**	**5**	4
9	**9**	**8**	**7**	**6**	**5**	4	3
8	**8**	**7**	**6**	**5**	4	3	2
7	**7**	**6**	**5**	4	3	2	1
6	**6**	**5**	4	3	2	1	0
5	**5**	4	3	2	1	0	

Table 7 School closing and the migration of primary school age children in Guizhou and Yunnan

	Dependent variable: migrate at the primary school age=1					
	(1)	(2)	(3)	(4)	(5)	(6)
	unrestricted			migrants within hukou city dropped		
A: MALE sample						
Post×*GZ*	0.028	0.020	0.021	0.054**	0.048**	0.084*
	(0.019)	(0.017)	(0.038)	(0.024)	(0.021)	(0.041)
GZ	-0.148***	-0.141***	-1.955***	0.030***	0.047***	0.336***
	(0.015)	(0.013)	(0.058)	(0.010)	(0.009)	(0.059)
post	-0.027*	0.005	0.010	-0.037	-0.029	-0.051
	(0.016)	(0.021)	(0.031)	(0.023)	(0.026)	(0.039)
Obs.	4,573	4,573	4,573	3,255	3,255	3,255
Adj. R2	0.005	0.147	0.158	0.008	0.133	0.149
B: FEMALE sample						
Post×*GZ*	0.020	0.027	0.012	0.005	0.005	-0.022
	(0.017)	(0.017)	(0.026)	(0.017)	(0.017)	(0.031)
GZ	0.023**	0.011	0.256***	0.023**	0.018	0.560**
	(0.010)	(0.010)	(0.060)	(0.010)	(0.011)	(0.211)
post	-0.024	-0.037**	-0.027	0.005	0.009	0.024
	(0.015)	(0.016)	(0.019)	(0.013)	(0.016)	(0.025)
Obs.	5,106	5,106	5,106	3,167	3,167	3,167
Adj. R2	0.006	0.137	0.144	0.004	0.111	0.119
Controls:						
city of origin	Y	Y	Y	Y	Y	Y
quadratic age	N	Y	Y	N	Y	Y
year	N	Y	Y	N	Y	Y
city specific time trend	N	N	Y	N	N	Y
P value for H0:	0.678	0.681	0.798	0.057	0.047	0.005

Notes: H0 is the hypothesis that the coefficient of the interaction term *post*×*GZ* does not differ between the two groups of male and female. Standard errors are in parentheses and are clustered at the city level. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table 8 School closing and the migration of primary school age children in Sichuan and Yunnan

Dependent variable: migrate at the primary school age=1						
	(1)	(2)	(3)	(4)	(5)	(6)
	unrestricted			migrants within hukou city dropped		
A: MALE sample						
Post×*SCH*	0.038**	0.021	0.022	0.044*	0.029	0.080*
	(0.017)	(0.015)	(0.033)	(0.024)	(0.020)	(0.040)
SCH	-0.169***	-0.132***	-1.975***	-0.401***	-0.317***	-1.792***
	(0.013)	(0.012)	(0.046)	(0.019)	(0.017)	(0.039)
post	-0.027*	0.007	0.010	-0.037	-0.008	-0.051
	(0.016)	(0.018)	(0.030)	(0.023)	(0.023)	(0.039)
Obs.	8,523	8,523	8,523	6,915	6,915	6,915
Adj. R2	0.002	0.186	0.199	0.002	0.196	0.213
B: FEMALE sample						
Post×*SCH*	0.032**	0.037**	0.038*	-0.002	-0.000	-0.016
	(0.016)	(0.014)	(0.023)	(0.014)	(0.013)	(0.027)
SCH	0.026***	0.068***	0.331***	0.045***	0.039***	0.540**
	(0.009)	(0.009)	(0.039)	(0.007)	(0.007)	(0.207)
post	-0.024	-0.031**	-0.027	0.005	0.012	0.024
	(0.014)	(0.015)	(0.019)	(0.013)	(0.015)	(0.025)
Obs.	8,617	8,617	8,617	6,241	6,241	6,241
Adj. R2	0.004	0.171	0.180	0.002	0.163	0.178
Controls:						
city of origin	Y	Y	Y	Y	Y	Y
quadratic age	N	Y	Y	N	Y	Y
year	N	Y	Y	N	Y	Y
city specific time trend	N	N	Y	N	N	Y
P value for H0:	0.686	0.331	0.601	0.053	0.084	0.007

Notes: H0 is the hypothesis that the coefficient of the interaction term *post*×*SCH* does not differ between the two groups of male and female. Standard errors are in parentheses and are clustered at the city level. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table 9 Parents' educational expectation for boys vs. girls

Dependent variable: educational expectation, college or above=1/otherwise=0					
	Aged 0-5		Aged 6-15		
	(1)	(2)	(3)	(4)	(5)
male	0.052***	0.051**	0.015	0.024	0.037**
	(0.019)	(0.020)	(0.017)	(0.017)	(0.018)
Age	no	yes	no	yes	yes
parent's education	no	yes	no	yes	yes
Log (family income per capita)	no	yes	no	yes	yes
academic performance	—	—	no	no	yes
County dummies	no	yes	no	yes	yes
Obs.	1,592	1,592	2,116	2,116	2,116
Adj_R^2	0.004	0.123	-0.000	0.088	0.115

Notes: Age is a continuous variable. *parent's education* is the highest education level of the parent who answered the question about educational expectation and could be either the father or mother. *Family income per capita* is in current Yuan in 2016 and is used in log form in the regressions. *Academic performance* refers to the child's average grade in Chinese language/grammar and math tests of the last semester as perceived by the parent (the grades are classified into four levels: excellent, good, average and poor). Standard errors in parentheses are clustered at the family level. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. Data source: CFPS 2016.

Table 10 Share of migrant students and share of male students

Dependent variable: share of male students within class						
	(1)	(2)	(3)	(4)	(5)	(6)
migrant share_1	0.059**	0.077***	0.038			
	(0.024)	(0.027)	(0.052)			
migrant share_2				0.057***	0.069***	0.010
				(0.021)	(0.023)	(0.046)
City effects	N	Y	Y	N	Y	Y
School effects	N	N	Y	N	N	Y
Obs.	438	438	438	438	438	438
R-squared	0.022	0.111	0.438	0.021	0.109	0.437

Note: "migrant share_1" and "migrant share_2" are calculated at the class level. In the first share, migrant students are defined as those born out of county or those whose birthplace was unknown; and in the second, migrant students are defined as those with hukou registered out of the county. Robust standard errors are in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table 11 migrant children and probability of being boys, LPM

Dependent variable: boy (yes=1/no=0)					
			Exclude migrants living with parent(s)	Migrant sample	
	(1)	(2)	(3)	(4)	(5)
Born out of county or unknown	0.043***	0.042***	0.072*		
	(0.012)	(0.014)	(0.037)		
Hukou out of county	-0.031**	-0.016	0.029		
	(0.013)	(0.015)	(0.047)		
None - one-child family		-0.154***	-0.165***		
		(0.012)	(0.014)		
No siblings left behind				-0.066**	-0.065**
				(0.029)	(0.032)
City/school/grade dummies	Y	Y	Y	Y	Y
Other controls	N	Y	Y	N	Y
Obs.	18,795	14,841	11,362	2,118	1,509
R-squared	0.012	0.049	0.041	0.069	0.146

Notes: City and school dummies and the grade dummy are controlled in all columns; columns 2, 3, and 5 also control for students' residence, health, family backgrounds, and the parents' education levels. Column 3 deletes migrant observations that live with parents. Columns 4 and 5 contain only migrant students with hukou registered out of county. Robust standard errors are in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table 12 The extra costs/difficulties faced by migrant households

	The parents have taken measures to enroll the student in the current school (yes=1/no=0)			Did parents face difficulty in preparing required official documents to have the student enrolled (yes=1/no=0)		
	(1)	(2)	(3)	(4)	(5)	(6)
boy	0.041***	0.037***	0.038	0.016**	0.011*	0.021
	(0.007)	(0.007)	(0.025)	(0.007)	(0.006)	(0.022)
migrant		0.138***	0.138***		0.101***	0.105***
		(0.015)	(0.018)		(0.010)	(0.014)
boy×migrant			-0.001			-0.009
			(0.021)			(0.019)
Obs.	14,550	14,550	14,550	14,021	14,021	14,021
Adj. R2	0.002	0.077	0.077	0	0.103	0.103

Notes: City and school dummies, grade dummy, students' residence, health, family backgrounds, and parents' education levels are controlled for in all regressions. Robust standard errors are in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

Table 13 To enroll the student, whether the parents have done the following (yes=1/no=0)

	boy		migrant		boyXmigr.		Adj R2
	Coef.	s.e	Coef.	s.e	Coef.	s.e	
Friends for help	0.031	(0.020)	0.103***	(0.014)	-0.010	(0.017)	0.055
Bribe person in charge	0.018**	(0.008)	0.015***	(0.005)	-0.011	(0.007)	0.011
Pay extra fees	0.006	(0.010)	0.031***	(0.007)	-0.001	(0.009)	0.039
Purchase housing	0.009	(0.013)	0.030***	(0.009)	-0.003	(0.011)	0.095
Hukou transfer	0.010	(0.010)	-0.009	(0.007)	-0.003	(0.008)	0.029
Prepare Hukou registration	-0.004	(0.014)	-0.034***	(0.009)	-0.005	(0.012)	0.047
Property ownership cert.	0.004	(0.018)	0.094***	(0.015)	-0.003	(0.015)	0.385
Temporary resident permit	0.075***	(0.019)	0.425***	(0.019)	-0.063***	(0.019)	0.340
Social security record	0.026	(0.016)	0.126***	(0.015)	-0.020	(0.015)	0.131
Family planning cert.	0.028*	(0.015)	0.098***	(0.013)	-0.020	(0.014)	0.072
Business permit/employ cert.	0.027**	(0.013)	0.118***	(0.012)	-0.023*	(0.013)	0.120
Satisfactory points	0.022**	(0.010)	0.014*	(0.007)	-0.019**	(0.009)	0.015
Other documents	0.001	(0.012)	0.016**	(0.008)	-0.004	(0.010)	0.042

Notes: Each row corresponds to a regression. City and school dummies, grade dummy, students' residence, health, family backgrounds, and parents' education levels are controlled for in all regressions. Robust standard errors are in parentheses. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively.

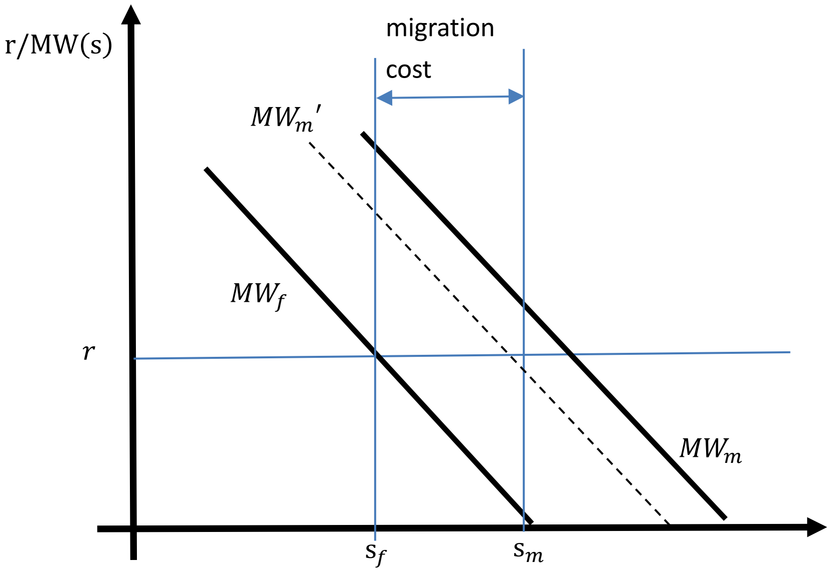

Figure 1 Son preference, migration, and human capital investment

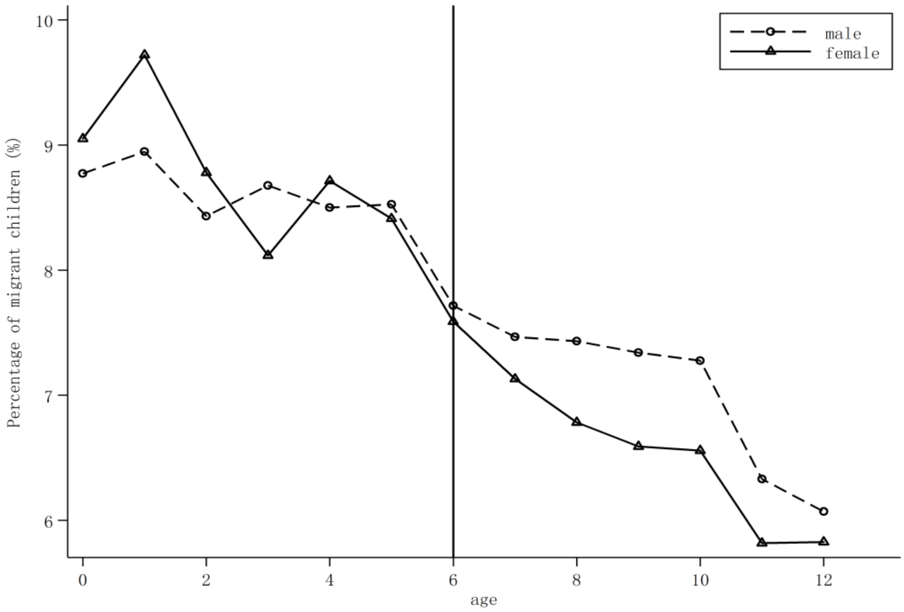

Figure 2 Percentage of migrant children in number of rural children 0-12 age (%)

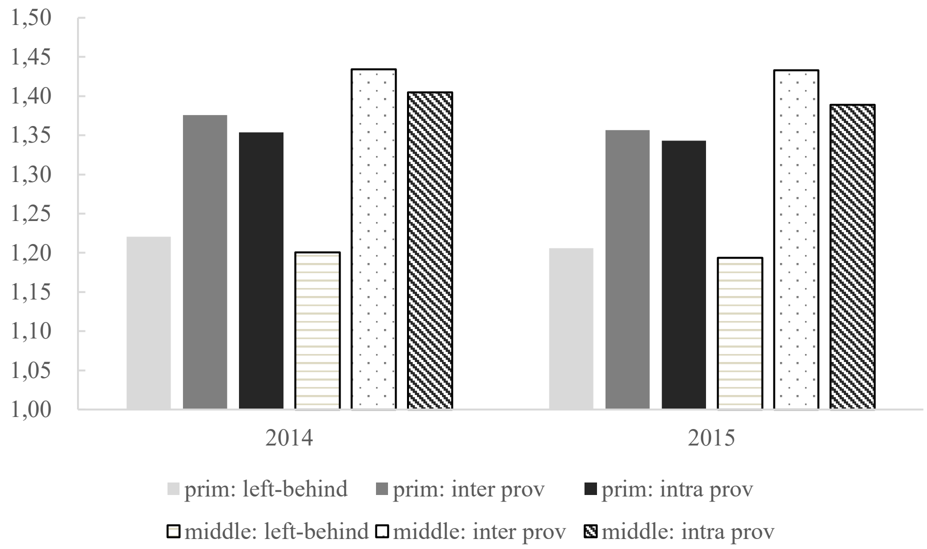

Figure 3 The proportion of boys in the migrant and non-migrant children population

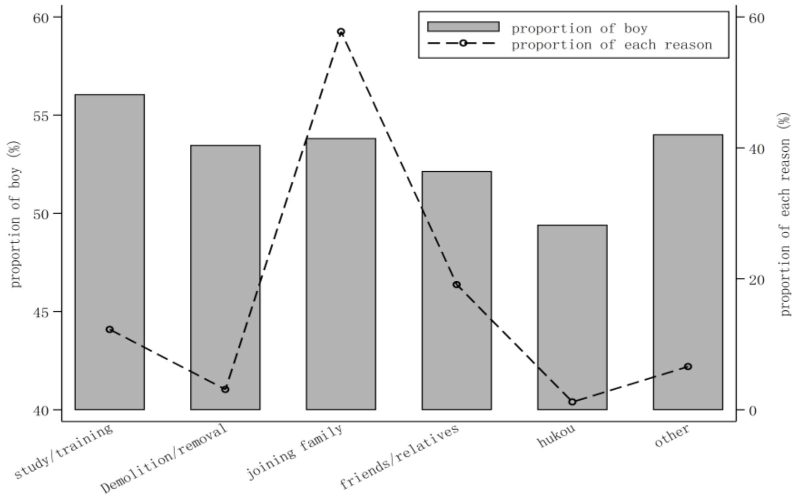

Figure 4 Boys' proportion in different migration reason groups and the proportion of each reason

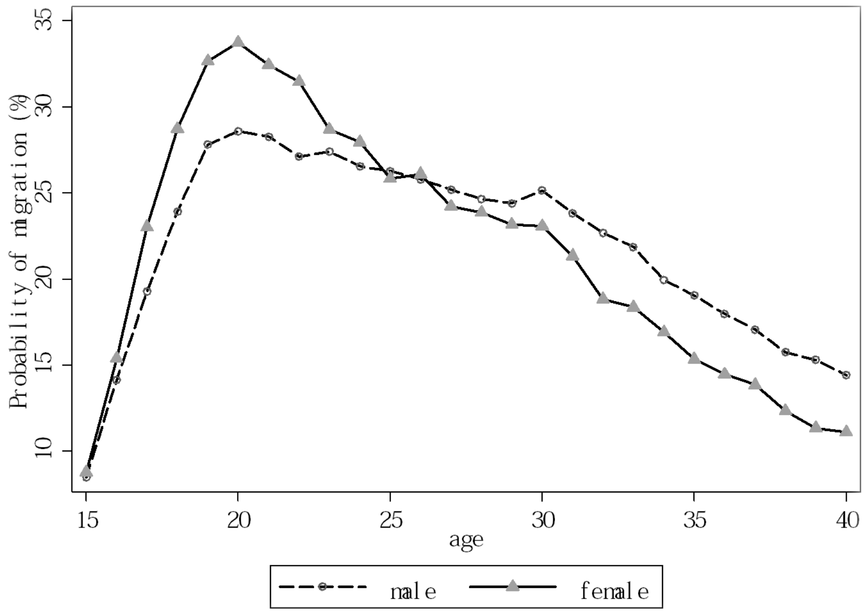

Figure 5 The probability of migration by age and gender of rural residents, 2005

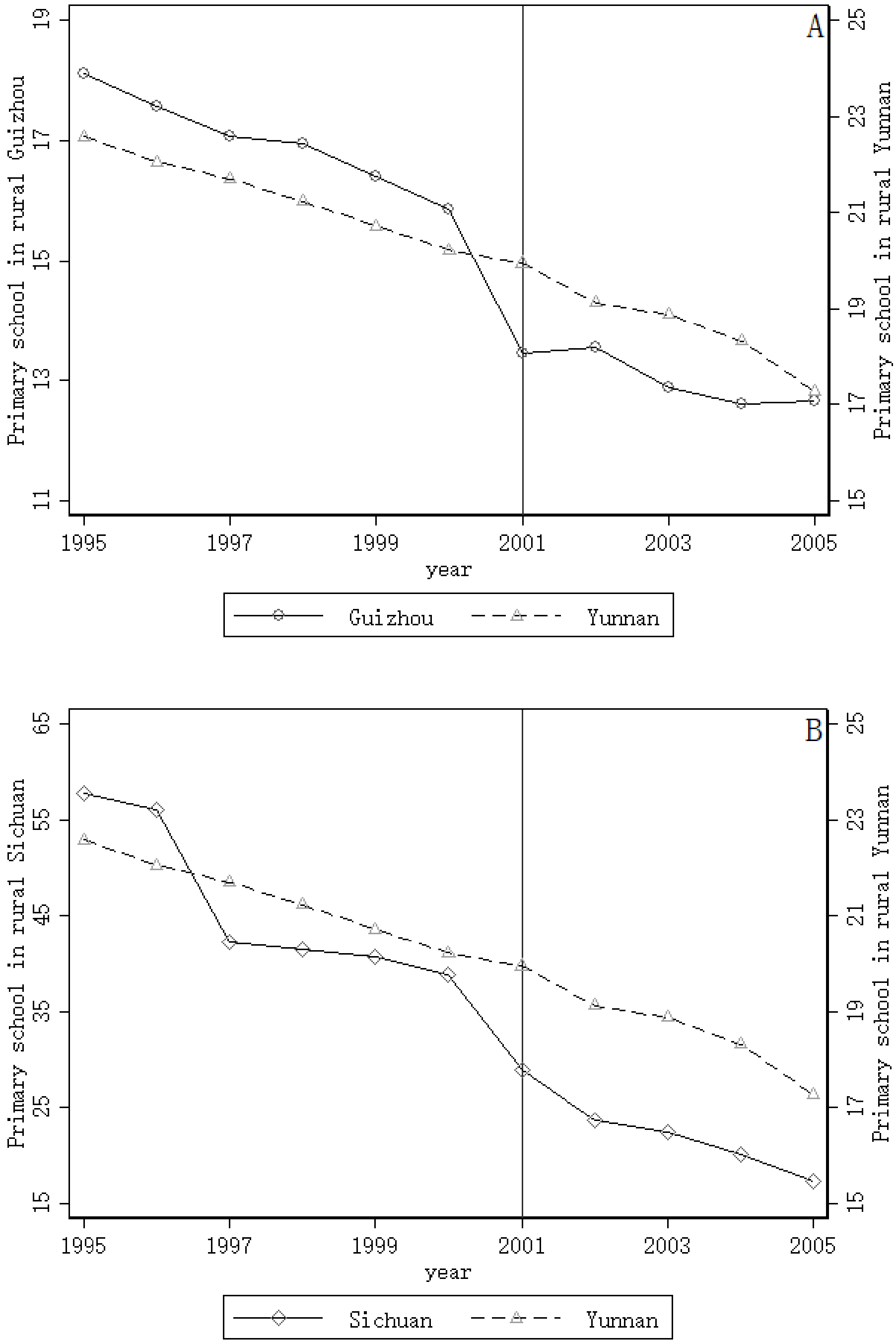

Figure 6 The number of primary schools in rural areas for Guizhou, Sichuan, and Yunnan, 1995-2005, unit: 1000

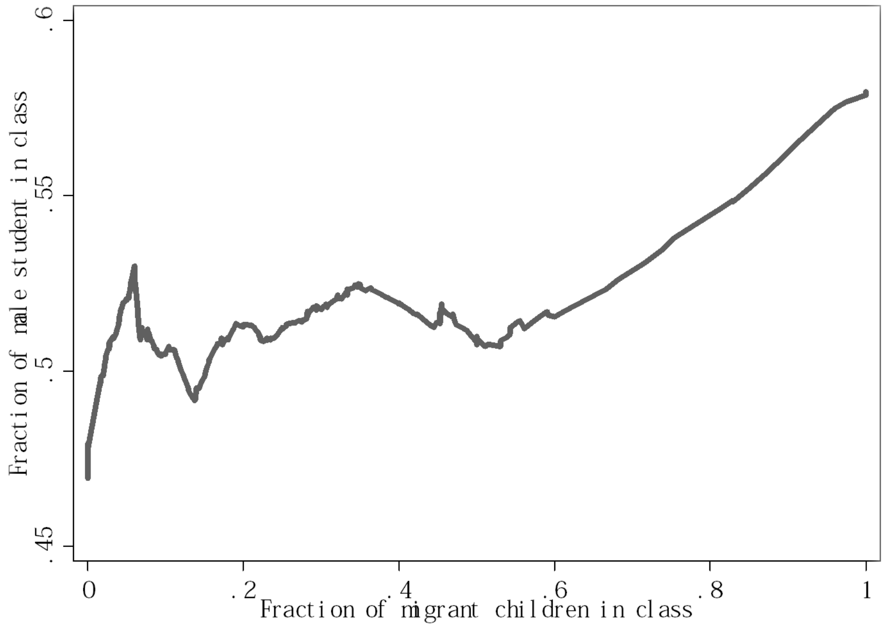

Figure 7 The share of boy students in the total share of migrant students in the class

Table 14 Placebo test (Guizhou and Yunnan, 2000 census data)

	Dependent variable: migrate at the primary school age=1					
	(1)	(2)	(3)	(4)	(5)	(6)
	unrestricted			migrants within hukou city dropped		
A: MALE sample						
Post×GZ	-0.014	-0.026	-0.119	-0.032	-0.028	-0.066
	(0.055)	(0.052)	(0.091)	(0.055)	(0.049)	(0.094)
GZ	-0.108***	-0.098***	-0.028	-0.093***	-0.087***	-0.038
	(0.027)	(0.021)	(0.088)	(0.028)	(0.016)	(0.057)
post	0.074*	0.108*	0.137*	0.078*	0.099	0.100
	(0.039)	(0.056)	(0.078)	(0.046)	(0.059)	(0.085)
Obs.	1,224	1,224	1,224	958	958	958
Adj. R2	0.037	0.219	0.258	0.032	0.208	0.261
B: FEMALE sample						
Post×GZ	0.001	-0.000	-0.004	0.020	0.009	0.040
	(0.036)	(0.031)	(0.062)	(0.034)	(0.033)	(0.059)
GZ	-0.068**	-0.045*	0.010	-0.079***	-0.051**	-0.030
	(0.030)	(0.025)	(0.053)	(0.027)	(0.024)	(0.050)
post	0.009	-0.031	-0.032	-0.003	-0.044*	-0.061
	(0.021)	(0.026)	(0.046)	(0.021)	(0.026)	(0.042)
Obs.	1,283	1,283	1,283	941	941	941
Adj. R2	0.006	0.141	0.197	0.040	0.146	0.156
Controls:						
city of origin	Y	Y	Y	Y	Y	Y
quadratic age	N	Y	Y	N	Y	Y
year	N	Y	Y	N	Y	Y
city specific time trend	N	N	Y	N	N	Y
P value for H0:	0.812	0.677	0.324	0.381	0.542	0.373

Notes: H_0 is the hypothesis that the coefficient of the interaction term *post×GZ* does not differ between the two groups of male and female. Standard errors are in parentheses and are clustered at the city level. *, **, and *** represent significance levels of 10%, 5%, and 1%, respectively. post=1(1998-2000), post=0(1996-1997).