DISCUSSION PAPER SERIES

IZA DP No. 7859

Sibling Influence on the Human Capital of the Left Behind

Costanza Biavaschi
Corrado Giulietti
Klaus F. Zimmermann

December 2013

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Sibling Influence on the Human Capital of the Left Behind

Costanza Biavaschi
IZA

Corrado Giulietti
IZA

Klaus F. Zimmermann
IZA and University of Bonn

Discussion Paper No. 7859
December 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 7859
December 2013

ABSTRACT

Sibling Influence on the Human Capital of the Left Behind*

While a growing literature has analyzed the effects of parental migration on the educational outcomes of children left behind, this is the first study to highlight the importance of sibling interactions in such a context. Using panel data from the RUMiC Survey, we find that sibling influence on schooling performance is stronger among left- behind children. Hence, parental migration seems to trigger changes in the roles and effects among children. However, it is primarily older sisters who exhibit a positive influence on their younger siblings. We corroborate our results by performing a series of tests to mitigate endogeneity issues. The results from the analysis suggest that sibling effects in migrant households might be a mechanism to shape children's outcomes and success and that adjustments within the family left behind have the potential to generate benefits – or reduce hardship – in response to parental migration.

JEL Classification: O15, J61

Keywords: left behind, siblings, human capital

Corresponding author:

Corrado Giulietti
IZA
P.O. Box 7240
53072 Bonn
Germany
E-mail: giulietti@iza.org

* The authors would like to thank participants at the IZA and Center on Human Capital Workshop on Migration and Human Capital (organized in collaboration with the Journal of Human Capital and the IZA Journal of Migration), the 8th IZA-World Bank Conference on Employment and Development, the 5th CIER-IZA Workshop, as well as seminar participants at Peking University and at the Institute for Population and Economics – Chinese Academy of Social Sciences. This draft has greatly benefitted from comments by Joseph Altonji, Jinyoung Kim, Peter Kuhn and Yaohui Zhao as from editorial suggestions by Victoria Finn.

The Longitudinal Survey on Rural Urban Migration in China (RUMiC) consists of three parts: the Urban Household Survey, the Rural Household Survey and the Migrant Household Survey. It was initiated by a group of researchers at the Australian National University, the University of Queensland and the Beijing Normal University and was supported by the Institute for the Study of Labor (IZA), which provides the Scientific Use Files. The financial support for RUMiC was obtained from the Australian Research Council, the Australian Agency for International Development (AusAID), the Ford Foundation, IZA and the Chinese Foundation of Social Sciences.

1 Introduction

After decades of research on the factors that induce people to migrate and the consequences for the receiving regions, the literature has now reverted its attention to the effects of migration on the sending areas. Within this context, recent studies have focused on the impact of family separation on individuals left behind. A number of papers have analyzed the consequences of a family member's migration on the education outcomes of left-behind children, although no consensus has been reached on the sign and magnitude of such an impact. On the one hand, parental emigration inherently implies their absence from the household and can hence have detrimental effects on children's outcomes (Antman 2013). On the other hand, migration usually entails a flow of remittances that might be invested in children's education. Therefore, migration represents a type of parental absence that is fundamentally different from other types of separation such as death or divorce. Furthermore, the effects of migration on children's outcomes have been shown to be very heterogeneous, varying depending on the age and gender of the child, as well as the gender of the parent (for a review see Antman 2013).

The novel approach of our paper is to incorporate the role of sibling interactions in a context where parents are absent due to migration, and to study changes in sibship correlation in response to parental migration. There is a prominent literature investigating sibling correlations, finding important spillovers in education and income (Solon 1999; Black and Devereux 2011) in several developed countries (Schnitzlein 2014). Such correlations capture "nature" and "nurture" effects within households, common environments as well as the influence of one sibling on another (Black and Devereux 2011). Siblings might affect each other in a number of ways. For instance, sociologists propose that siblings might act as role models (Haynie and McHugh 2003) and psychologists add that they might even directly shape reciprocal outcomes by providing socializing opportunities or directly affecting (younger) siblings' personality and intelligence (Arnold et al. 1975). Economists have shown that older siblings influence younger siblings' high school graduation rates (Oettinger 2000; Rees and Sabia 2009), as well as risk-taking behaviors (Altonji et al. 2010). Furthermore, a long-standing literature shows how birth order, sex composition and family size affect children's outcomes (Behrman et al. 1982; Hanushek 1992; Kaestner 1997).

It is rational to expect that siblings' influence might play a different – if not more important – role during parental absence, particularly in the migration case. Yet, to the best of our knowledge, the literature studying the impact of such disruptive events on children's human capital development is scant. Qualitative findings from the psychology literature suggest that siblings experience increased closeness as a result of the shared experience (Abbey and

Dallos 2004), but very little is known about the interaction between siblings, human capital formation and migration. On the one hand, older siblings might substitute for parents and hence have a stronger influence on the younger children when parents are absent. On the other hand, if parents' and older siblings' supervision are complements, parental absence will still be detrimental for the academic achievements of younger children.

Our paper explores for the first time the role of sibling spillovers on the educational attainment of left-behind children, by focusing on China, a country where parental migration has become a primary concern. Official Chinese statistics report that there were more than 160 million rural migrant workers in the first quarter of 2013 (NBS China 2013). This massive migration has attracted considerable attention regarding its social consequences, since it has created a large number of split households and left-behind individuals. It is often difficult for migrant parents to bring children along, because hukou regulations usually prevent access to quality schools, as well as to urban welfare benefits. The All-Women Federation China estimates that around 40 million children aged 0-15 are left behind in rural areas due to parental migration. This accounts for around 20% of all rural children – and corresponds to the total U.S. population of children roughly the same age. Understanding the effects of parental absence on children left behind in rural areas is therefore important given the size of this phenomenon. Furthermore, learning whether sibling interactions are an intra-household mechanism that can smooth the potential adverse migration effects on left-behind children's human capital development is key to understanding the necessity and scope for public interventions that influence these interactions, such as the one-child policy.

Identification of sibling influence on economic achievement is non-trivial. Methodologically, we add to the literature in several respects. First, we address the problem of unobserved shared influences using a unique panel data on sibling pairs. Second, we explore whether results are driven by the endogeneity of the parents' migration decision. To address this issue, we adopt two strategies. We exploit the longitudinal dimension of the data to compare the performance of left-behind and non-migrant children before their parents migrate, carefully controlling for migration decision reasons. As long as the school performance of children is not driving the migration decision and future migration is not affecting past performance, such a strategy enables purging the endogeneity of migration from our estimates. We subsequently exploit the differential timing of migration to compare the children currently left-behind with those who will be left behind in the future.

Our findings indicate that sibling influence on school performance is stronger among left-behind children, although there is only an effect for Chinese and not for Math. The results also highlight remarkable heterogeneous patterns, depending on the gender composition and the age distance within sibling pairs. Although some left-behind children seem to suffer in terms of school performance, the presence of an older sibling completely balances out such a negative outcome. These results therefore speak to several aspects of the literature. First, parental migration not only results in a change in bargaining power among household decision makers (Antman 2013), but also triggers changes in the roles and influences of children. Second, peer effects and sibling effects in migrant households might be worth studying, as they seem to play a particularly important role in shaping children's behavior and outcomes.

The remainder of the paper proceeds as follows. Section 2 briefly outlines existing theoretical perspectives on the role of sibship on human capital development. Sections 3 and 4 describe the data used in the analysis and the empirical strategy, respectively. Benchmark results and the heterogeneity analysis are presented in Sections 5 and 6. Section 7 meticulously explores potential econometric issues and outlines our strategies to mitigate endogeneity. Section 8 concludes the paper.

2 Do Siblings Affect Human Capital Accumulation?

The economic literature provides a few theoretical perspectives on the potential importance of sibling influence on children's human capital accumulation in the migration context. First, parental migration might trigger changes in resources spent on children's education (Hanson and Woodruff 2003). Second, it might vary bargaining powers within the household. Third, it might differentially affect the incentives to invest in children (Antman 2011b). Parents or caretakers might adopt reinforcing, compensating or neutral investment in children, not only in terms of monetary resources, but also in terms of intrahousehold task reallocation (Antman 2011a). The presence of an older sibling might attenuate or reinforce the disruption from migration, hence affecting children's outcomes, since intra-household adjustments might be borne differently by relatively older or younger children. The paper focuses on this mechanism, which has not been studied in the literature.

In fact, most of the existing studies have primarily looked at the effect of parental migration on children, without focusing on sibling interactions (for a review, see Antman 2013). Furthermore, the literature that considers the educational outcomes of children left behind in China remains scant. Using data from two provinces in China, Meyerhoefer and Chen (2011) show that migration is associated with a lag in the educational attainment of girls, which is interpreted as a re-allocation of girls' time towards home production in migrant households. Chen et al. (2009) and work by Wang (2012) look at the effect of parental absence on children's school enrollment, performance and health. While the former study finds no effect on educational achievement in Shaanxi province, the latter finds negative and persistent effects of parental absence on children's school enrollment, especially for boys. We build upon this work and introduce the effect of sibship on the human capital performance of children left behind in rural China.

In order to empirically test the research question of interest, we estimate regression models with the young children's schooling performance (exam scores in Chinese and Math) as the response variable and the oldest sibling's performance, an indicator for being left behind and the interaction between the two latter terms as the key explanatory covariates:

$$Score^{Y}_{ijt} = \beta_0 + \beta_1 \text{Score}^{O}_{jt} + \beta_2 \text{Left-Behind}_{jt} + \gamma(\text{Score}^{O}_{jt} \times \text{Left-Behind}_{jt}) + \beta_3 X^{O}_{ijt} + \beta_4 W_{jt} + \eta_t + c_i + \epsilon_{ijt}. \tag{1}$$

We indicate with the superscript Y the outcomes and characteristics of the young children in the household and with the superscript O the score of the oldest child. Hence, $Score^{Y}_{ijt}$ is the score in Math or Chinese for child i, in household j, in year t. Left-Behind$_{jt}$ is the indicator that equals one if one of the parents has lived outside the household for at least one month in the 12 months preceding the survey. Score$^{O}_{jt}$ measures the oldest sibling's score, and hence β_1 captures the correlation between the scores of the oldest and younger siblings. The interaction term (Score$^{O}_{jt}$ $\times$ Left-Behind$_{jt}$) is the key variable of interest, measuring the additional sibling influence if the children are left behind. In our regressions, we cluster standard errors at the household level, accounting for key variables being measured at this stratum.

The regression model controls for a variety of other factors, captured by the vectors X^{O}_{ijt} and W_{jt}. Such factors measure younger children's characteristics such as gender, age dummies, entry grade dummies, school quality and whether they attend boarding school; additionally, they capture family characteristics such as parents' education and labor market characteristics, households' characteristics such as household size, land and financial assets, as well as village characteristics and indicators for provinces (for OLS models only). Finally, η_t measures time fixed effects and c_i captures unobserved individual heterogeneity.

As a starting point, we estimate equation 1 with Ordinary Least Squares (OLS). However, simple OLS ignores the unobserved individual heterogeneity that might drive young children's performance but can also be correlated with the probability of being left behind or with the oldest sibling's scores. In our context, where the vast majority of observations are represented by sibling pairs, c_i almost corresponds to a household fixed effect – a factor that it is particularly important to control for. In fact, some of the cross-sectional comparison between siblings could be reduced, increased or eliminated once household time-invariant characteristics are taken into consideration. A model that controls for this family effect also seems particularly relevant in this context, where our key variable is a function of a household choice. Individual fixed effects will mitigate the endogeneity of the migration decision stemming from time-invariant family traits that affect educational outcomes as well as the propensity to migrate. We therefore estimate the model in equation (1) by a standard fixed effect estimator that purges such heterogeneity through first differencing.

3 Data

Our analysis is based on the RUMiC, a large-scale project conducted in China, comprising the Rural Household Hurvey (RHS), the Urban Household Survey and a Migrant Household Survey. For our purposes, we extract data from the 2009, 2010 and 2011 waves of the RHS (see Akgüc et al. 2013 for a technical description of the RUMiC panel dataset). The dataset contains detailed information about household members – including those currently migrating out of the village – and comprises sociodemographic characteristics, labor market outcomes, migration history and the family situation prior to leaving the hometown. The data also provide information on educational characteristics and outcomes of children younger than 16 years of age and family members older than 16 who are still in school. We complement these variables with a rich set of controls at the village level, which include the village population, expenses on education, as well as the number of teachers and pupils.

We restrict the sample to children between 5 and 18 years old currently living in rural areas. In the main analysis, we use pairs of older-younger siblings in which the oldest child in the household is paired with all younger siblings (when there are more than two children in school). To be included in the sample, siblings must be of school age and report scores.

Our outcome variables are young children's test scores in Chinese and Math in the semester prior to the interview, which a parent or guardian self-reports. Accurate knowledge of student performance is guaranteed through recording students' scores in a booklet. Due to varying scoring standards across provinces, we normalize scores by the highest obtainable score on the specific test, through information also reported in the survey.

The key covariate of interest is an indicator for being left behind. In our baseline regressions, we define a left-behind child as one whose father or mother left the hometown for at least one month during the twelve months preceding the survey. Hence, this variable captures children's exposure to the absence of at least one parent during the period of schooling attendance. Each year about one-third of the children have at least one parent who has spent some time outside the household, once again highlighting the relevance of such a phenomenon.

In Table 1 we show characteristics by migration status, i.e., dividing our sample of young children into those who are left behind and those from non-migrant families (columns 1 and 2). The share of males in the sample is around 60%, with little difference between the two groups. This high share reflects the consequences of the one-child policy.[1] The average age of left-behind children is also similar to that of children whose parents have not migrated. There is also rather little variation in the school entry age, averaging around six and a half years old. Around 25%-30% of siblings are in boarding school and are enrolled in 5^{th} grade. There seems to be some variation in the subjective assessment of the children's school quality, with left-behind children less likely to be enrolled in a high-quality school (where the benchmark is the average quality of school in the village).

Table 1 also shows the characteristics of the parents and households in which children live. Mothers and fathers of left-behind children are more likely to engage in wage work. The number of household members and household land size do not vary between the two groups; however, households with children left behind have higher financial assets.

The last rows of the table report the characteristics of the children's village and province distribution. Left-behind children come from smaller villages, which spend much less on education and have fewer teachers. The provincial distribution of children left behind differs from that of children of non-migratory parents, reflecting varying migration propensity depending on the geographic location.[2]

The most interesting patterns emerging from the first two columns of Table 1 are the rather small differences in terms of individual observable traits between the left-behind children and those of non-migrants, with the exception of parental characteristics which, however, we will be able to control for through fixed effect estimations. Such small differences reassure the likely absence of a strong selection between migrant and non-migrant households. On the other hand, children in migrant households seem to come from more impoverished villages. In all analyses, we will therefore control for time-varying village characteristics.

The third and fourth columns of Table 1 compare the characteristics of the children in our sample with those of children who are excluded from our analysis. The latter group is largely composed of children from one-child households, but also includes children who have (younger or older) siblings who are currently not in school, and hence do not report test scores, or

[1] In the majority of rural areas, the policy allows couples to have a second child if their first one is a girl. On aggregate, for families with more than one child, this determines a disproportionate share of males among youngest siblings and females among oldest siblings: In fact, in our sample, nearly 67% of oldest siblings are female.

[2] For the 2011 wave, data from the Guangdong province were not collected, which slightly unbalances aggregate figures on the geographical distribution. For robustness purposes, we have performed our entire analysis on a restricted sample excluding this province, finding essentially identical results to those presented in the paper.

siblings whose scores are missing. The purpose of this comparison is to highlight potential selectivity patterns emerging from the definition of our sample. Unlike what was expected a priori, such differences are rather limited. One peculiar aspect is the sex ratio, which again exhibits a relatively large share of males. Within the sample of “Other children”, mostly from one-child families, the reasons of such an unbalance are related to the consequences of sex-selective abortions (for a recent description, see Wei and Zhang 2011). Other socio-demographic variables are very similar between the children in our sample and other children, with the exception of the village characteristics and the provincial distribution.[3]

Table 2 shows the performance in Chinese and Math by left-behind status. For completeness, we also report exam scores of children out of our estimation sample. Younger siblings perform consistently better than older siblings, with such differences present for both children left behind and those living in non-migrant households. How much of the sibling resemblance is driven by individual heterogeneity, endogeneity of the migration decision or simply observable differences across the two groups? The remainder of the paper will answer this question.

4 Results

The key results of the paper are reported in Tables 3 and 4. Starting with the OLS estimates, some interesting patterns arise. Looking at the performance in Chinese throughout all specifications, there is a high correlation between the oldest and younger child’s score, with estimates in the range of 0.39 to 0.45. Such estimated correlation is remarkably in line with sibling correlations found in other studies (see, e.g., Björklund and Salvanes 2010). As shown in column II, on average the absence of a parent does not seem to have a large effect on children’s performance in school, which might be driven by remittances balancing the potential adverse effect of being left behind. Columns III and IV show the estimates of the interaction between the older sibling’s scores and being left behind, controlling for several characteristics of the child, parents, household and village. Throughout all specifications the results are remarkably stable.

To solidify this idea, let us consider a left-behind child and a child whose parents have not migrated, both of whom have an older sibling who scored zero in Chinese (this example purely serves the scope of comparative statics, as there are virtually no children exhibiting score zero). Compared to children of non-migrants, left-behind young children with poor performing oldest siblings face a score penalty of 12% on average. However, as younger chil-

[3]For example, the one-child policy was particularly enforced in rural areas of Jiangsu, Chongqing and Sichuan. This determines that one-child families are relatively more represented in these provinces.

dren are exposed to better performing siblings, not only are their scores positively correlated to their sibling's scores, but also the disadvantage of being left behind is fully compensated. Therefore, left-behind children with average performing siblings end up only suffering from a 1%-1.5% penalty in their scores. The presence of a top performing sibling is able to fully re-balance the score disadvantage stemming from being left behind.

In conclusion, sibling influence is stronger in left-behind households since sibling correlations are higher than in non-migrant households. All other controls exhibit the expected sign: children enrolled in high-quality schools have higher scores; furthermore, the higher the number of teachers in the village, the better the children perform. Two variables inhibit children's scores: the high number of students in the village and household's land size. Villages with too many students might have lower quality school and education systems less targeted to the children's needs. Children from households owning larger land might face a disincentive in investing in education, as well as being induced to work (Liang and Chen 2007).

Columns V to VIII of Table 3 also show the same regressions for performance in Math, with estimates for the sibling correlation very similar to those for Chinese. However, the estimates for the left-behind indicator and for its interaction with the oldest sibling's score are much smaller in magnitude and statistically insignificant, despite following a similar pattern in terms of sign. These results are not entirely surprising if one considers the different skills – and their transferability – involved in the study of Chinese and Math. We could postulate that good performance in Math is harder to achieve and influence, while Chinese performance – driven by students' ability to read and write a certain number of characters – can be better monitored and influenced by older siblings. This differential pattern between Chinese and Math scores persists throughout most of our analyses.

As previously highlighted, these patterns might suffer from the presence of individual effects that correlate with the covariates of interest, which could stem from individual ability as well as the endogeneity of the migration decision. In Table 4, we tackle this issue by estimating fixed effect regressions. The fixed effect estimator purges out any observed and unobserved variation that is common within individuals. In our context, where the majority of sibling pairs are from a two-sibling household, fixed effect results are also almost identical to results obtained by including household time-invariant effects.[4] Hence, we are able to control for any individual- or household-specific propensity to migrate. Remarkably, the fixed effects estimates reinforce our previous findings.

As in the OLS case, left-behind children seem to suffer from parental absence in their performance in Chinese, while there are little effects in Math. The sibling correlation remains

[4]Results available upon request.

economically sizeable even after controlling for individual heterogeneity. The introduction of the fixed effects exacerbate the older siblings influence in Chinese scores when parents are away: the estimate for the interaction term between the left-behind indicator and the oldest sibling's score doubles compared to the OLS estimates. Achievement in Chinese of the children left behind is around 30% lower than that of the non-migrant children in the most parsimonious specification, independently of the controls added. However, as before, left-behind children seem to resemble their older siblings in such a way that this disadvantage is reduced and brought to zero in the case of top performing older siblings. As before, performance in Math is unaffected by being left behind and sibling influence is the same in the migrant and non-migrant households.

Our baseline analysis suggests two main findings. First, the presence of an older child plays a more important role in terms of schooling performance when the parent has migrated. Second, there is heterogeneity between school subjects in terms of the role that both parental absence and sibling influence play in shaping the human capital development of the left behind.

5 Heterogeneity Analysis

We continue our analysis by showing the heterogeneity of the results by children and parent characteristics in Table 5. The first panel in Table 5 shows performance by the sex composition of sibling pairs. In both Chinese and Math, older male siblings do not exhibit additional effects on the left behind. However, the opposite is true for older female siblings: the positive correlation across scores is amplified among left-behind households. Once again, this pattern is strong in the case of Chinese performance. However, when focusing on female-female pairs, we also find sibling effects for the case of Math performance, albeit these results are statistically significant at the 10% level only. Therefore, it appears that older sisters exhibit a nurturing effect on younger siblings, who seem to have an advantage from the presence of a top performing child, large enough to even revert the penalty of being left behind. The changing role of female children in the household quantifies the speculative conclusion in Meyerhoefer and Chen (2011), who suggest that migrant households re-allocate girls' time towards home production.

We further disentangle the effects by looking at which stage of compulsory education sibling influence matters the most – in earlier or later grades – and at which age difference such effects come into play. The second panel of Table 5 presents results for children with smaller (less or equal to 5 years) and larger (more than 5 years) age distance between the sibling pairs. As one would perhaps expect, the sibling influence is stronger among children whose age distance is further apart. Correlations are particularly stronger in the left-behind group in both Chinese and Math. For children closer in age, correlations are not only statistically insignificant but also economically small. We could conjecture, once again, that this pattern captures a nurturing effect of older siblings on younger ones, with this result also holding in the acquisition of Math skills.

The last four columns of this panel show the effects for younger children in grade 1 to 3 and grades greater than 3. Remarkably, children who have already spent a few years in school are most affected by sibling influence. While results are non-significant for children in grade 1 to 3, children in grade 4 and above exhibit strong correlation across Chinese exam scores, a penalty if left-behind, and stronger dependence on older siblings' performance if left behind – although these effects are weaker than those by age. As before, the left-behind disadvantage is fully compensated for by the average performing older sibling. Overall, it seems that the effect on younger siblings is driven by both age distance and the younger child's stage on the educational ladder.

In the third panel of Table 5, we show whether the impact is stronger in households where the child lives with parents or other individuals (relatives, friends or other people in the village). As expected, the impact of migration and the sibling effects are stronger when parental authority is absent and children live with members outside the restricted familial nucleus. We subsequently check whether sibling correlations are stronger in households with a previous migration experience. We select a subsample of households without and with migration history, i.e. those households who report having migrated before 2009. Remarkably, only children whose parents have not migrated before are those who face a penalty for being left behind. As before, however, this penalty is mitigated by the presence of siblings. Therefore, it appears that the intra-household reallocation of roles (and potentially resources) is only temporary and migrant households are able to adjust over time to the possible disruption of migration.

In the last panel of the table, we test whether sibling effects depend on which parent is absent. We show estimates from models in which we consider left behind only those children whose mother is absent (columns 1 and 4), the father is absent (columns 2 and 5) and both parents are absent (columns 3 and 6). Remarkably, we find similar effects on the left behind, independently of which parent is migrating, although the mother matters somewhat more for Chinese than the father; once again, the impact is essentially nil in the case of Math performance.

6 Econometric Issues

There are two main methodological challenges in our analysis. First, the migration decision might remain endogenous even after controlling for several confounding factors, as we do in our regressions. Migrants self-select in the decision to migrate, and there might be additional unobserved factors correlated with both children's outcomes and migration status. For instance, families with better socioeconomic status, networks and earning potential might be more (or less) likely to migrate, as well as possibly provide an environment in which siblings might interact more (or less) with each other. Although we control for a significant variety of characteristics, it is difficult to rule out a priori that endogeneity affects the results. Second, it might be argued that reverse causality is present: migrants might move and leave children behind due to children's school performance; for example since parents might want to provide better educational opportunities for children or might decide not to move if children are facing particular difficulties in school. Furthermore, younger siblings' performance might in turn affect older sibling performance. We allow for this possibility and admit that the sibling correlation captures an "equilibrium relationship" between the sibling pairs.

In this section, we proceed with a few robustness checks, which we present in Table 6. We aim to detect the relevance of these problems in our settings. Before proceeding, it should be noted that one of the remarkable advantages of our dataset is the panel structure, with children observed at three points in time. Hence, we have already purged time-invariant heterogeneity in the migration decision and the siblings relationship from the model by using the fixed effect estimator. The OLS estimator appeared to be upward biased (as we would expect if siblings had a positive influence on each other), although results from the OLS and fixed effect models are not very economically diverse. Although not a test, our baseline results suggest that reverse causality and endogeneity in migration might not be fundamental concerns, given the controls and structure of the data.

Exploiting variation in the timing of migration to mitigate endogeneity. Children in our panel are observed as being left behind over different years. Instead of using variation from comparing left-behind and non-migrant households, we use variation in the timing of being left behind, and drop observations of those who never migrate. In this restricted sample, the control group consists of children who are not left behind at time t but were left behind in the past or will be in the future. By only focusing on the left behind for some t, the advantage of such a strategy is that we implicitly control for the endogeneity in the left-behind indicator.[5] The key identifying assumption is the absence of dynamic selection,

[5] Other studies have used such a technique, see for example Bertrand and Mullainathan (1999).

once relevant characteristics are controlled for. The first four columns in the upper panel of Table 6 show the results applying both OLS and fixed-effect estimators, which are similar to those found in the main part of the paper and thus reassure that endogeneity concerns are limited.

Perspective migration to understand selection. As a secondary check to assess the magnitude of self-selection in migration, we estimate the following model:

$$Score^{Y}_{ijt} = \beta_0 + \beta_1 \text{Score}^{O}_{jt} + \beta_2 \text{Left-Behind}_{jt+1} + \gamma(\text{Score}^{O}_{jt} \times \text{Left-Behind}_{jt+1}) + \beta_3 X^{Y}_{ijt} + \beta_4 W_{jt} + \epsilon_{ijt}. \quad (2)$$

An established literature explores how migrants are selected by comparing the pre-migration outcomes of future migrants and stayers (Moraga 2011; Kaestner and Malamud 2013). The model above proceeds in the same spirit: assuming that future migration is unrelated to current shocks, if selection was not a serious concern in our data, we should expect outcomes not to differ depending on migration status. The last two columns of the upper panel show that migration is unrelated to current performance and hence pre-migration outcomes between migrants and stayers are not different. Furthermore, this analysis also suggests that parents do not respond to children's schooling achievement by adjusting their migration plans.

Understanding reverse causality. We further investigate this last point by directly testing whether migration depends on educational success. We relate the probability of being left behind at time t with children's educational achievement, controlling for the same characteristics used in the baseline specification. We therefore correlate the probability of parents moving in response to the children's current educational outcomes. Following the approach adopted throughout, we estimate the model through fixed effects.

Columns 1 and 2 in the lower panel of Table 6 show the estimates from such a model. Current performance in both Chinese and Math is unrelated to the probability of staying behind, not only statistically, but also in terms of magnitude of the coefficients. As an additional test, we run the same analysis, but relating the likelihood of being left behind at $t+1$ with current performance. In other words, we look at whether children's education performance is related with the future migration decision. Results are reported in columns 3 and 4: current children performance in Chinese and Math are unrelated to future parental migration as well. Therefore, concerns of reverse causality should not be strong.

Lastly, as already mentioned, younger siblings' performance might in turn affect older siblings' performance. Although this paper does not aim to pinpoint the "directionality" of sibling interactions, we check whether the oldest sibling resents from his or her changing role in the household. In this test, the outcome variable is the score of the oldest sibling, while the key covariate is an interaction term between the left-behind indicator and the average score of the younger siblings. The last two columns of Table 6 show that the impact on the older sibling is statistically zero and economically very small.

7 Conclusions

While a growing literature has analyzed the effects of parental migration on the educational outcomes of left-behind children, ours is the first study to highlight the importance of sibling influence in such a context. We find consistent results that sibling interactions affect the cognitive development of younger children. The effects are stronger among children left-behind, albeit only in the acquisition of language ability, in which case the positive influence of older siblings compensates the negative effects of being left behind. Parental migration, hence absence, seems to trigger changes in the reciprocal roles and interactions between children. Heterogeneity analysis reveals that such changes primarily involve the role of older sisters and are stronger depending on the current grade and the age distance between siblings. Effects are somewhat stronger when the mother migrates instead of the father.

Our results suggest that sibling effects in migrant households play an important function in shaping children's educational outcomes and success. There are also relevant policy implications. With internal migration expected to increase, the welfare of left-behind individuals will become increasingly central to the Chinese government's policy agenda. Our results suggest that policies fostering sibling interactions will sort – ceteris paribus – positive effects in terms of human capital development. The relaxation of the one-child policy announced by the Chinese government in November 2013 has the potential to create such externalities, since a larger number of rural families will be allowed to have more than one child, hence creating the ground for promoting sibling interactions.

Tables

Table 1: Characteristics of Left-Behind Children and Children in Non-Migrant Households

Variable	Young Siblings (in sample) Left-behind	Non-migrant	All	Other Children (not in sample)
Male (D)	0.597	0.618	0.611	0.619
	(0.491)	(0.486)	(0.488)	(0.486)
Age	11.041	11.370	11.258	11.685
	(3.072)	(3.104)	(3.097)	(3.725)
Age at entry	6.560	6.746	6.683	6.661
	(0.735)	(0.803)	(0.785)	(1.278)
Boarding school (D)	0.253	0.298	0.283	0.368
	(0.435)	(0.458)	(0.451)	(0.482)
High quality school (D)	0.191	0.251	0.231	0.292
	(0.394)	(0.434)	(0.421)	(0.455)
Grade	4.968	5.269	5.167	6.347
	(2.730)	(2.997)	(2.912)	(3.572)
Father is farmer or out of labor force (D)	0.103	0.444	0.328	0.321
	(0.304)	(0.497)	(0.470)	(0.467)
Father is employee (D)	0.758	0.375	0.506	0.558
	(0.428)	(0.484)	(0.500)	(0.497)
Father is self-employed (D)	0.139	0.181	0.167	0.122
	(0.346)	(0.385)	(0.373)	(0.327)
Mother is farmer or out of labor force (D)	0.489	0.726	0.645	0.545
	(0.500)	(0.446)	(0.479)	(0.498)
Mother is employee (D)	0.429	0.201	0.278	0.391
	(0.495)	(0.401)	(0.448)	(0.488)
Mother is self-employed (D)	0.082	0.074	0.077	0.064
	(0.275)	(0.261)	(0.266)	(0.245)
Household size	5.155	4.997	5.051	4.305
	(1.278)	(1.183)	(1.218)	(1.267)
Household financial asset (ln RMB)	6.983	7.126	7.077	6.824
	(2.813)	(3.356)	(3.181)	(3.469)
Household land size (Mu/1000)	0.160	0.147	0.152	0.166
	(0.276)	(0.303)	(0.294)	(0.337)
Lag village population (/100)	2.605	3.156	2.968	2.508
	(2.009)	(3.043)	(2.746)	(1.753)
Lag village expenditure on education (Mill RMB)	0.005	0.016	0.012	0.009
	(0.034)	(0.131)	(0.109)	(0.076)
Lag village number of teachers (/100)	0.109	0.240	0.195	0.095
	(0.153)	(0.921)	(0.755)	(0.264)
Lag village number of students (/100)	2.590	3.651	3.289	1.789
	(3.880)	(6.256)	(5.583)	(3.749)
Province: Hebei	0.026	0.072	0.056	0.055
	(0.160)	(0.258)	(0.230)	(0.229)
Province: Jiangsu	0.111	0.080	0.090	0.166
	(0.314)	(0.271)	(0.286)	(0.373)
Province: Zhejiang	0.021	0.065	0.050	0.127
	(0.142)	(0.247)	(0.218)	(0.333)
Province: Anhui	0.212	0.155	0.174	0.118
	(0.409)	(0.362)	(0.380)	(0.322)
Province: Henan	0.131	0.135	0.134	0.102
	(0.338)	(0.342)	(0.340)	(0.302)
Province: Hubei	0.056	0.067	0.063	0.120
	(0.231)	(0.250)	(0.243)	(0.325)
Province: Guangdong	0.200	0.349	0.298	0.064
	(0.401)	(0.477)	(0.458)	(0.245)
Province: Chongqing	0.026	0.023	0.024	0.073
	(0.160)	(0.151)	(0.154)	(0.260)
Province: Sichuan	0.217	0.055	0.111	0.175
	(0.413)	(0.229)	(0.314)	(0.380)
Observations	534	1032	1566	4224

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Standard deviations in parentheses.
Left-behind children are those whom either parent left the hometown for at least one month during the twelve months preceding the survey. Other children refer to the group of children not included in the analysis. See text for details.

Table 2: Test Scores in Chinese and Migrants by Migration Status

	Chinese		Math	
	Left-Behind	Non-Migrant	Left-Behind	Non-Migrant
Young siblings' score	0.802	0.811	0.818	0.821
	(0.119)	(0.110)	(0.122)	(0.116)
Oldest siblings' score	0.776	0.781	0.790	0.795
	(0.120)	(0.119)	(0.127)	(0.132)
Observations	534	1032	534	1032
Other children's score	0.821	0.816	0.836	0.833
(not in sample)	(0.120)	(0.122)	(0.122)	(0.125)
Observations	1495	2729	1495	2729

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Standard deviations in parentheses. Scores refer to exam scores in Chinese and Math in the semester prior to the interview as reported by the parent or guardian of the child.

Table 3: Performance in Chinese and Math, OLS Results

	Chinese				Math			
	I	II	III	IV	V	VI	VII	VIII
Score^O_{jt}	0.454***		0.453***	0.393***	0.410***		0.409***	0.378***
	(0.036)		(0.035)	(0.039)	(0.028)		(0.028)	(0.032)
Left-Behind_{jt} (D)		–0.014**	–0.010*	–0.120**		–0.008	–0.004	–0.059
		(0.007)	(0.006)	(0.059)		(0.007)	(0.006)	(0.052)
$\text{Left-Behind}_{jt} \times \text{Score}^O_{jt}$				0.140*				0.072
				(0.072)				(0.061)
Male (D)	–0.005	–0.007	–0.005	–0.008	0.000	0.001	0.000	–0.001
	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)
Age	–0.003*	–0.005**	–0.003*	–0.004**	–0.002	–0.005**	–0.002	–0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Age at entry	–0.006	–0.012**	–0.007*	–0.010**	–0.009**	–0.015***	–0.009**	–0.011***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
Boarding school	–0.007	–0.010	–0.007	–0.006	–0.008	–0.007	–0.008	–0.004
	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.008)	(0.007)	(0.007)
High quality school	0.034***	0.048***	0.034***	0.028***	0.037***	0.052***	0.037***	0.031***
	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)
Grade	–0.002	–0.002	–0.002	–0.001	–0.003	–0.002	–0.003	–0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Father is employee				0.002				0.002
				(0.006)				(0.007)
Father is self-employed				–0.004				0.000
				(0.008)				(0.009)
Mother is employee				0.012*				0.009
				(0.007)				(0.007)
Mother is self-employed				0.009				0.012
				(0.012)				(0.014)
Household size				–0.003				–0.005**
				(0.003)				(0.003)
Household financial asset				–0.001				–0.001
				(0.001)				(0.001)
Household land size				–0.028**				–0.030**
				(0.012)				(0.012)
Lag village population				0.000				0.001
				(0.002)				(0.002)
Lag village expenditure on education				0.005				–0.018
				(0.021)				(0.020)
Lag village number of teachers				0.015***				0.013**
				(0.004)				(0.006)
Lag village number of students				–0.002**				–0.001
				(0.001)				(0.001)
R^2	0.31	0.10	0.31	0.35	0.29	0.09	0.29	0.32
N	1566	1566	1566	1566	1566	1566	1566	1566

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses.
Column IV and VIII include province fixed effects.
All regressions include year fixed effects.

Table 4: Performance in Chinese and Math, Fixed Effects Results

	Chinese				Math			
	I	II	III	IV	V	VI	VII	VIII
Score^{O}_{jt}	0.432***		0.431***	0.332***	0.403***		0.402***	0.368***
	(0.071)		(0.071)	(0.075)	(0.042)		(0.042)	(0.053)
Left-Behind_{jt} (D)		–0.018	–0.013	–0.211**		–0.009	–0.004	–0.083
		(0.013)	(0.011)	(0.091)		(0.016)	(0.013)	(0.068)
$\text{Left-Behind}_{jt} \times \text{Score}^{O}_{jt}$				0.252**				0.099
				(0.116)				(0.080)
Boarding school	0.009	0.002	0.009	0.007	0.004	0.001	0.003	0.004
	(0.009)	(0.011)	(0.009)	(0.009)	(0.010)	(0.012)	(0.011)	(0.011)
High quality school	0.024**	0.026**	0.024**	0.022**	0.021*	0.023*	0.021*	0.020
	(0.011)	(0.012)	(0.011)	(0.011)	(0.012)	(0.013)	(0.012)	(0.012)
Grade	–0.003	–0.000	–0.003	–0.003	–0.005*	–0.004	–0.005*	–0.006*
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)
Household size				0.003				–0.006
				(0.010)				(0.013)
Household financial asset				–0.002				–0.000
				(0.002)				(0.002)
Household land size				–0.016				–0.036**
				(0.014)				(0.015)
Lag village population				0.003				–0.008
				(0.007)				(0.007)
Lag village expenditure on education				0.067				0.010
				(0.080)				(0.055)
Lag village number of teachers				0.015				0.017
				(0.016)				(0.016)
Lag village number of students				–0.002				0.001
				(0.001)				(0.001)
R^2	0.25	0.02	0.25	0.28	0.21	0.02	0.21	0.23
N	1566	1566	1566	1566	1566	1566	1566	1566

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses. All regressions include year fixed effects.

Table 5: Heterogeneity

	By Sex of Sibling Pairs							
	Chinese				Math			
	MO-MY	MO-FY	FO-MY	FO-FY	MO-MY	MO-FY	FO-MY	FO-FY
Score^{O}_{jt}	0.1744	0.5123***	0.2798*	0.2700**	0.3247***	0.4385***	0.4522***	0.2164*
	(0.1375)	(0.0651)	(0.1470)	(0.1328)	(0.0996)	(0.0632)	(0.0929)	(0.1209)
Left-Behind_{jt} (D)	–0.1972	0.0096	–0.3141**	–0.3131**	0.0247	0.0195	–0.0836	–0.3103**
	(0.1314)	(0.1146)	(0.1595)	(0.1458)	(0.1177)	(0.1700)	(0.1218)	(0.1532)
$\text{Left-Behind}_{jt} \times \text{Score}^{O}_{jt}$	0.2428	–0.0427	0.3987*	0.3586**	–0.0126	–0.0169	0.1075	0.3489*
	(0.1821)	(0.1392)	(0.2065)	(0.1755)	(0.1449)	(0.2038)	(0.1461)	(0.1788)
N	264	248	693	361	264	248	693	361

	By Age Distance				By Grade of Young Sibling			
	Chinese		Math		Chinese		Math	
	$\leq$ 5 years	$>$ 5 years	$\leq$ 5 years	$>$ 5 years	Grade$\leq$ 3	Grade$>$ 3	Grade$\leq$ 3	Grade$>$ 3
Score^{O}_{jt}	0.3555***	0.1615**	0.4240***	0.1197	0.2735***	0.3579***	0.2306*	0.3606***
	(0.1054)	(0.0695)	(0.0644)	(0.0839)	(0.0933)	(0.0980)	(0.1328)	(0.0623)
Left-Behind_{jt} (D)	–0.1715*	–0.4261**	–0.0423	–0.3192**	–0.1868	–0.2153*	–0.1141	–0.1248*
	(0.0944)	(0.1928)	(0.0757)	(0.1341)	(0.1272)	(0.1213)	(0.1343)	(0.0755)
$\text{Left-Behind}_{jt} \times \text{Score}^{O}_{jt}$	0.1934	0.5313**	0.0589	0.3659**	0.2137	0.2722*	0.1226	0.1465
	(0.1217)	(0.2428)	(0.0949)	(0.1502)	(0.1683)	(0.1531)	(0.1646)	(0.0952)
N	1053	513	1053	513	498	1068	498	1068

	Lives with Parents or Others				Migrated Before			
	Chinese		Math		Chinese		Math	
	Parents	Others	Parents	Others	Yes	No	Yes	No
Score^{O}_{jt}	0.2523***	0.3500*	0.3618***	0.2353	0.3406***	0.2213	0.3783***	0.3363***
	(0.0871)	(0.1877)	(0.0581)	(0.1625)	(0.0817)	(0.1477)	(0.0630)	(0.0758)
Left-Behind_{jt} (D)	–0.0698	–0.3055*	0.0914	–0.2244	–0.0981	–0.3882**	–0.0933	–0.0902
	(0.1089)	(0.1646)	(0.1614)	(0.1463)	(0.0834)	(0.1559)	(0.0982)	(0.0838)
$\text{Left-Behind}_{jt} \times \text{Score}^{O}_{jt}$	0.0755	0.4075*	–0.1210	0.3083*	0.1152	0.4484**	0.1161	0.0933
	(0.1345)	(0.2154)	(0.1895)	(0.1834)	(0.1054)	(0.2014)	(0.1169)	(0.0997)
N	991	575	991	575	1232	334	1232	334

	By Absent Parent					
	Chinese			Math		
	Mother	Father	Both	Mother	Father	Both
Score^{O}_{jt}	0.3493***	0.3389***	0.3529***	0.3865***	0.3723***	0.3893***
	(0.0631)	(0.0735)	(0.0625)	(0.0480)	(0.0524)	(0.0478)
Left-Behind_{jt} (D)	–0.2918***	–0.1970**	–0.2814***	–0.0835	–0.0745	–0.0736
	(0.1007)	(0.0909)	(0.1026)	(0.0734)	(0.0666)	(0.0762)
$\text{Left-Behind}_{jt} \times \text{Score}^{O}_{jt}$	0.3590***	0.2431**	0.3642***	0.0946	0.1005	0.1035
	(0.1259)	(0.1174)	(0.1296)	(0.0871)	(0.0801)	(0.0901)
N	1566	1566	1566	1566	1566	1566

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses.
All regressions include year fixed effects and the full set of regressors in column IV and VIII of Table 4.
MO=Male older sibling; MY=Male younger sibling; FO=Female older sibling; FY=Female younger sibling.
Migrated before refer to whether either parent had migrated before the year 2009.

Table 6: Robustness

	Left Behind Only				Perspective Mig.	
	OLS		Fixed Effects		Fixed Effects	Fixed Effects
	Chinese	Math	Chinese	Math	Chinese	Math
Score^O_{jt}	0.3749***	0.3956***	0.2796***	0.3871***	0.4183***	0.4088***
	(0.0419)	(0.0362)	(0.0879)	(0.0637)	(0.0624)	(0.0481)
Left-Behind_{jt} (D)	–0.1368**	–0.0457	–0.2552***	–0.0690	–0.1039	0.0170
	(0.0602)	(0.0538)	(0.0975)	(0.0729)	(0.0783)	(0.0660)
$\text{Left-Behind}_{jt} \times \text{Score}^O_{jt}$	0.1605**	0.0560	0.3094**	0.0834	0.1316	–0.0107
	(0.0744)	(0.0636)	(0.1243)	(0.0867)	(0.0966)	(0.0785)
N	1420	1420	1420	1420	843	843
	$\text{Pr(Left behind}_t)$		$\text{Pr(Left behind}_{t+1})$		Oldest Sibling	
	Chinese	Math	Chinese	Math	Chinese	Math
Score^O_{jt}	–0.0039	–0.0917	0.3218	–0.1452	0.5139***	0.4972***
	(0.1385)	(0.1386)	(0.3165)	(0.2400)	(0.1281)	(0.0952)
Left-Behind_{jt} (D)					–0.1085	–0.0852
					(0.1156)	(0.1035)
$\text{Left-Behind}_{jt} \times \text{Score}^O_{jt}$					0.1244	0.0857
					(0.1429)	(0.1248)
Own score	–0.1841	–0.0425	–0.0408	0.4347		
	(0.1516)	(0.1647)	(0.3208)	(0.3395)		
N	1566	1566	843	843	1198	1198

Source: RUMiC data, RHS waves 2009, 2010 and 2011. Robust standard errors clustered at the household level in parentheses.
All regressions include year fixed effects.
The group of Left Behind Only refers to children who at any year in the panel are left behind.
Perspective migrants refer to the estimation of equation 2, in which scores at time t are regressed on an indicator for being Left-Behind at time $t+1$.
Oldest sibling refers to estimating equation 1 for the sample of oldest siblings.

References

Abbey, C. and R. Dallos (2004). The experience of the impact of divorce on sibling relationships: A qualitative study. *Clinical Child Psychology and Psychiatry 9*(2), 241–259.

Akgüc, M., C. Giulietti, and K. F. Zimmermann (2013). The RUMiC longitudinal survey: Fostering research on labor markets in China. *IZA Discussion Paper –*.

Altonji, J. G., S. Cattan, and I. Ware (2010). Identifying sibling influence on teenage substance use. *NBER Discussion Paper 16508*.

Antman, F. M. (2011a). The intergenerational effects of paternal migration on schooling and work: What can we learn from children's time allocations? *Journal of Development Economics 96*(2), 200–208.

Antman, F. M. (2011b). International migration and gender discrimination among children left behind. *The American Economic Review 101*(3), 645–649.

Antman, F. M. (2013). The impact of migration on family left behind. In A. F. Constant and K. F. Zimmermann (Eds.), *International Handbook on the Economics of Migration*. Cheltenham: Edward Edgar.

Arnold, J., A. Levine, and G. Patterson (1975). Changes in sibling behavior following family intervention. *Journal of Consulting and Clinical Psychology 43*(5), 683–688.

Behrman, J. R., R. A. Pollak, and P. Taubman (1982). Parental preferences and provision for progeny. *The Journal of Political Economy 90*(1), 52–73.

Bertrand, M. and S. Mullainathan (1999). Is there discretion in wage setting? A test using takeover legislation. *The Rand Journal of Economics 30*(3), 535–554.

Björklund, A. and K. G. Salvanes (2010). Education and family background: Mechanisms and policies. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 3. Amsterdam: North Holland.

Black, S. E. and P. J. Devereux (2011). Recent developments in intergenerational mobility. In O. Ashenfelter and D. Card (Eds.), *Handbook of labor economics*, Volume 4B. Amsterdam: Elsevier.

Chen, X., Q. Huang, S. Rozelle, Y. Shi, and L. Zhang (2009). Effect of migration on children's educational performance in rural China. *Comparative Economic Studies 51*(3), 323–343.

Hanson, G. H. and C. Woodruff (2003). Emigration and educational attainment in Mexico. Unpublished manuscript, University of California at San Diego.

Hanushek, E. A. (1992). The trade-off between child quantity and quality. *Journal of Political Economy 100*(1), 84–117.

Haynie, D. L. and S. McHugh (2003). Sibling deviance: In the shadows of mutual and unique friendship effects? *Criminology 41*(2), 355–392.

Kaestner, R. (1997). Are brothers really better? Sibling sex composition and educational achievement revisited. *Journal of Human Resources 32*(2), 250–84.

Kaestner, R. and O. Malamud (2013). Self-selection and international migration: New evidence from Mexico. *Review of Economics and Statistics, forthcoming.*

Liang, Z. and Y. P. Chen (2007). The educational consequences of migration for children in China. *Social Science Research 36*(1), 28–47.

Meyerhoefer, C. D. and C. Chen (2011). The effect of parental labor migration on children's educational progress in rural China. *Review of Economics of the Household 9*(3), 379–396.

Moraga, J. F.-H. (2011). New evidence on emigrant selection. *The Review of Economics and Statistics 93*(1), 72–96.

NBS China (2013). Accessed at `http://www.stats.gov.cn/was40/gjtjj_en_detail.jsp?searchword=migrants&channelid=9528&record=3` on 12 December 2013.

Oettinger, G. S. (2000). Sibling similarity in high school graduation outcomes: Causal interdependency or unobserved heterogeneity? *Southern Economic Journal 66*(3), 631–648.

Rees, D. I. and J. J. Sabia (2009). The effect of breast feeding on educational attainment: Evidence from sibling data. *Journal of Human Capital 3*(1), 43–72.

Schnitzlein, D. D. (2014). How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark. *Journal of Population Economics 27*(1), 69–89.

Solon, G. (1999). Intergenerational mobility in the labor market. In O. Ashenfelter and D. Card (Eds.), *Handbook of labor economics*, Volume 3. Amsterdam: Elsevier.

Wang, S. X. (2012). The effect of parental migration on the educational attainment of their left-behind children in rural China. Unpublished manuscript, Central University of Finance and Economics Beijing.

Wei, S.-J. and X. Zhang (2011). The competitive saving motive: Evidence from rising sex ratios and savings rates in China. *Journal of Political Economy 119*(3), 511–564.