How Does Parental Out-migration Affect Left-behind Children's Schooling Outcomes?

Xiaoman Luo *

University of California, Davis

Abstract

In this paper, I investigate how parental out-migration affects the schooling outcomes of left-behind children in rural China. Unlike previous works which almost exclusively focus on the net effect of migration, I analyze three important causal mechanisms — parental absence, child's study time, and investment in the child — simultaneously via a mediation analysis, disentangling the total effect of migration into mechanism-specific effects which are informative for policy makers. The analysis can be justified by the equilibrium solution of a theoretical two-agent model. The identification strategy is based on the rank condition for structural equation models to handle the endogeneity and Heckman selection model to correct for nonrandom missing. Using survey data on rural households from nine provinces, I find that the effects through parental absence and investment are both significantly negative with large sizes, while the effect through child's study time is insignificant with a negligible size. The surprising negative effect through investment is mainly driven by reduced nutrition investment by de facto custodians, who may not have compatible incentives to allocate the remittances on the child. Through a refined subgroup analysis, I find that girls are suffering ten times more from the underinvestment than boys, revealing a shocking gender inequality in rural China. The findings suggest that policies which compensate for underinvestment, especially for girls, tend to be more effective in mitigating the negative effect of migration than other types of policies.

Keywords: migration, education, gender inequality, mediation analysis, structural equation model

*xmluo@ucdavis.edu

1 Introduction

1.1 Background

Labor migrants represent a substantial share of workforce in many developing countries. According to Human Development Report 2009 (Klugman, 2009), there are more than 740 million internal migrants who live and work outside their region of birth within their home country — approximately 2.6 times as many as international migrants, based on the estimates by World Migration Report 2020. Due to various kinds of mobility constraints, family members of migrants are often left behind, facing the potential adverse effects emotionally and physically (Waddington, 2003). The separation could have non-negligible consequences on education and health of left-behind family members.

In this paper, I focus on left-behind children of internal rural migrants in China. According to the 2010 Population Census of China, more than 61 million children are left behind in rural China by migrant parents, accounting for 37.7% of children in rural areas, and 21.88% of children in China overall. The massive number of left-behind children is a consequence of the *hukou* system, the household registration system in China. There are two types of *hukou* in China: rural and urban *hukou*, and it has been difficult to transfer from one type to the other. Prior to the 1970's, people with rural *hukou* were legally prohibited from migrating to urban areas. Since the late 1970's, to meet the huge labor demand in urban areas generated by the Chinese economic reform, the government has gradually relaxed the restriction on the *hukou* system to permit migration from rural to urban areas. Nevertheless, transfers of the *hukou* status remains highly restrictive — rural migrants and their families with rural *hukou* are generally excluded from the social benefits that urban citizens enjoy. In particular, children of rural migrants have limited access to free public schools, health care benefits, housing support, social security, and other resources. Children who migrate with their parents from rural to urban areas can either choose expensive private schools, or less costly "migrant schools" run by local entrepreneurs, typically with unsatisfactory quality of education. As a consequence, most migrant parents choose to leave children behind in their hometown.

Considering the sizable population of left-behind children in rural China, the effect of parental migration on left-behind children's educational outcomes has considerable implications on the accumulation of human capital in China. Although the effect may not be reflected in the short-term household livelihood, it is highly indicative of the future human capital, which directly affects the poverty level. Therefore, it is of paramount importance to assess the impact and design policies to mitigate the negative effects and amplify the positive effects, if any.

Despite the rich literature on internal migration in different developing countries (e.g. Arnold and Shah, 1984; Booth, 1995; Battistella and Conaco, 1998; Ganepola, 2002; Afsar, 2003; Maruja and Baggio, 2003; Edwards and Ureta, 2003; Mendoza, 2004; Adams Jr and Page, 2005; Bryant, 2005; Gupta et al., 2009; Arguillas and Williams, 2010; Antman, 2011; McKenzie and Rapoport, 2011; Graham and Jordan, 2011; Antman, 2013), including China (He et al., 2012; Chang et al., 2011; Chen, 2013; Zhao et al., 2014; Sun et al., 2015), existing works almost exclusively focus on the net effect of parental migration on left-behind children's schooling outcomes. While the net effect is scientifically meaningful, it is less informative for policymakers — even if there is a large negative net effect, it is neither efficient nor ethical to directly impose restrictions on migration because education is not the only important factor for social welfare. A more realistic approach is to design policies targeting specific mechanisms that are more manipulable than the migration per se, through which the migration affects the left-behind children's educational outcomes. To achieve this, the first step is to disentangle the net effect into different causal channels to assess their respective importance (Huber, 2016).

1.2 Contributions

In line with the literature, I investigate three widely-studied mechanisms — parental absence, child's time allocation, and investment in the child (Démurger, 2015). In Section 5.3, I further decompose the investment into two sub-mechanisms — nutritional spending and tuition spending. Each of these mechanism has been partially assessed in various countries. Nevertheless, previous studies in China and other developing countries almost exclusively focus on the net effect of each channel separately without taking their high correlation into account. A significant net effect of an unimportant mechanism may be purely contributed by another important mechanism that is correlated but missing in the analysis. In principle, simultaneous analysis of different mechanisms mitigates the bias due to correlation, thereby providing a more convincing comparison of potential policy targets.

To lay the foundation for the simultaneous analysis, I establish a simple two-agent model to model the decision-making processes of the child and parents jointly. The child's decision variable is the study time; the parents' decision variable is migration; and both agents are maximizing a weighted average of utilities in the present and in the future. Under reasonable assumptions on the utility functions and production function of human capital, both the child and parents face a trade-off between the utilities in two periods, leading to a non-trivial equilibrium. The equilibrium solution reveals how the child's educational outcome is affected by migration directly through parental absence and indirectly through time allocation and monetary investment.

Beyond the lack of simultaneity, most of the existing studies have another important limitation. The effect through an intermediate variable has two components: the effect of migration on that variable, and the effect of that variable on the educational outcome. To the best of my knowledge, all previous works estimate one component only, which provides an incomplete answer. I summarize a selective set of works in Section 1.3. For instance, Chang et al. (2011) and Chen (2013) find that children of migrants tend to spend more time on housework and thus less time on studying. However, this conclusion on its own does not prove the effectiveness of a policy that increases child's study hours, unless one can further show that increased hours has a positive effect on the schooling performance, which may or may not be true depending on the relative effect size of stress or fatigue. Similarly, Kandel and Kao (2000) find that high remittances sent back by migrants may decrease the child's schooling performance. However, this is insufficient to guide policymakers because the negative effect can be either attributed to that the remittances are not allocated to the child by the de facto custodian or that the remittances increase child's desire to work and reduce their aspiration to study.

I overcome this limitation through the mediation analysis, which can provide estimates of both components, thereby providing a fuller description of the effect through different channels. Mediation analysis is a standard technique to decompose the net effect of a treatment into a direct effect and indirect effects through different causal mechanisms. It has been popular for decades in psychology (Baron and Kenny, 1986), and was advocated recently in economics (e.g. Heckman et al., 2013; Heckman and Pinto, 2015; Huber, 2016, 2019; Celli, 2019). In my problem, the treatment is the migration decision of parents, the direct effect is given by the absence of parents, and the indirect effects are given by the child's time allocation and monetary investment in the child. Notably, the equations involved in the mediation analysis coincide with the equilibrium solution of my two-agent model under certain functional specification, rendering the empirical analysis coherent with the theoretical analysis. In this problem, the mediation analysis is complicated by unmeasured confounders and non-random missing mediators, namely child's study time and monetary investment in the child. I propose a generic identification strategy to handle these two sources of endogeneity simultaneously.

Applying the mediation analysis on the Rural-Urban Migration in China (RUMiC) survey, I find significantly negative direct effects of parental migration on both the language and math scores of left-behind children as shown in the literature (Zhao et al., 2014; Meng and Yamauchi, 2015). In particular, the direct effect is -0.524 ($p < 0.01$) for language scores and -0.453 ($p < 0.01$) for math scores, measured in standard deviations. The indirect effects through child's time allocation are insignificant and near zero, with size 0.003 ($p > 0.05$) for language scores and 0.002 ($p > 0.05$) for math scores in standard deviations, implying that intervention through this channel may not be effective for left-behind children in China. In contrast to many studies in other developing countries, I find that the indirect effects through investment in the child are significantly negative for both the language and math scores with larger magnitudes than the direct effects. Specifically, this indirect effect is -0.894 ($p < 0.001$) for language scores and -0.874 ($p < 0.001$) for math scores in standard deviations. The seemingly counterintuitive negativity is caused by underinvestment in the child despite the remittances sent by migrants. This could be driven by incompatible incentives of guardians in the absence of parents (Niimi et al., 2009; Chen, 2013). The net effect of migration on child's schooling performance by adding up three mechanism-specific effects is -1.42 for language scores and -1.33 for math scores in standard deviations. Despite the large net effect, the decomposition into mechanism specific effects is clearly more informative for policy makers.

In addition, the gender subgroup analysis shows that girls' schooling performances are disproportionately affected by migration through underinvestment, revealing a shocking gender inequality in rural China. In fact, the indirect effect through investment for girls is more than 10 times larger than that for boys on both language and math scores. By further decomposing the investment into nutritional spending and educational spending, I find that the underinvestment is driven by both decreasing nutritional spending and decreasing educational spending, and the former has a substantially larger effect sizes. This suggests that a policy which compensates for the nutritional underinvestment tends to be effective in mitigating the negative effects of parental migration on left-behind children.

1.3 Related literature

As mentioned in the last subsection, existing works that I am aware of are almost exclusively focusing on the net effect of a single mechanism without detailed mediation analysis. In this subsection, I will review a selective set of relevant studies and highlight which net effect they estimate.

The first line of studies estimate the direct effect of migration due to the absence of parents (e.g. Graham and Jordan, 2011; He et al., 2012; Antman, 2013). They find that left-behind children's schooling performances are negatively affected by higher levels of emotional disruption, stress, sadness (Ganepola, 2002; Mendoza, 2004), loneliness and abandonment (Battistella and Conaco, 1998; Maruja and Baggio, 2003), and lower self-esteem (Sun et al., 2015). In addition, the absence of parents may disrupt the discipline of children (Arnold and Shah, 1984) and reduce their cognitive preparedness for school (Booth, 1995).

The second line of works focuses on the effect of migration on child's time allocation. It should be noted that it is still different from the effect of migration through child's time allocation since the former is only a component of the latter. Chen (2013) and Chang et al. (2011) examine the effect of children's labor substitution caused by parental migration, concluding that children of migrant households spend more time on housework and thus have less time for studying. Similar evidence has been found in Mexico (Antman, 2011; McKenzie and Rapoport, 2011), although there is no agreement on whether boys or girls suffer more from housework.

The third line of literature studies the effect of remittances sent back by migrants on children's schooling performance. Most studies find that the remittances relax the investment constraints in children, thereby improving children's living conditions, educational spending, and nutrition status (e.g. Afsar, 2003; Adams Jr and Page, 2005; Gupta et al., 2009). The evidence has been found in Indonesia, Thailand (Bryant, 2005), Philippines (Bryant, 2005; Arguillas and Williams, 2010), Bangladesh (Afsar, 2003), Mexico (Hanson and Woodruff, 2003; Alcaraz et al., 2012), and El Salvador (Edwards and Ureta, 2003). Nevertheless, there are also contradictory findings that remittance sent home by migrants may not necessarily increase the investment in the child. Olwig (1999) shows that migrating parents usually leave their children with relatives such as grandparents or foster families. As a consequence, guardians may not have strong incentives to turn remittances into investment in the child due to the potential competition between elderlies and children, or between current and future consumption (Nguyen et al., 2006; Chen, 2013; Niimi et al., 2009; Knodel and Saengtienchai, 2002).

The rest of this paper is organized as follows. Section 2 describes the two-agent model that lays the foundation for empirical analyses and presents some descriptive results of the equilibrium. All mathematical derivations are relegated into Appendix A. Section 3 introduces the RUMiC survey data, as well as the definitions of the treatment, mediators, and outcome. In Section 4, I describe the identification strategy in detail, including the choices of instrumental variables. I present the main empirical findings in Section 5, with results on all samples and subgroups. Section 6 concludes and discusses future directions.

2 Theoretical Modeling Framework

2.1 A two-agent model

To understand the interaction between three mechanisms qualitatively, I consider a simple model with a household of one child and one parent with two time periods but without borrowing or savings. In the first period, I assume that the parent is at working age and the child is at school age, and the household consumption purely relies on the parent's income while the child's income from housework is negligible. In the second period, I assume that the child has grown up and fully entered the labor market while the parent has retired, so the household consumption solely relies on child's income. The child decides on how much time to spend on studying and the parent decides on migration, with both decisions made in the first period. The model is arguably over-simplified since it ignores different roles of the father and mother, behaviors of siblings and de facto custodians, irrationality of decision making from both sides, etc.. Nevertheless, it is complicated enough to reveal how the mechanisms of interest, namely the parental absence, child's study time, and investment in the child, interact and affect the child's schooling performance.

Let u_1^k and u_2^k denote the utility of the child [1] in period 1 and 2, respectively, and s be the share of time that the child spends studying. Therefore, $(1-s)$ denotes the share of time that the child spends on activities other than studying. Furthermore, I denote by h the human capital level of child in period 1, by h_0 the endowment of human capital, by $d \in [0,1]$ the proportion of days that parent migrate away and leave the child behind, by W_p the parent income from work as a function of d, by β_k the child's discount factor of the second-period utility, and by $f(\cdot)$ the production function of human capital, which takes the input of parent migration status, child study time, monetary investment in child, and the endowment in human capital. I assume the child's utility in the first period depends on s and child's consumption c_1^k, while the second-period utility depends solely on the household consumption c_2. Note that c_1^k does not necessarily increase as the parent income W_p increases, because $\gamma(d)$, the proportion of total income spent on the child, is not necessarily increased in d. Essentially, $\gamma(d)$ is the decision variable of the de facto custodian, who can decide how much of the remittances sent back by migrants will be spent on the child. If the custodian has full control of the spending when the parents migrate out, $\gamma(d)$ can be decreasing in d (Nguyen et al., 2006; Chen, 2013; Niimi et al., 2009; Knodel and Saengtienchai, 2002). Due to the lack of data on guardians, I model $\gamma(d)$

[1] We use the letter k for "kid" instead of c for "child" to avoid similarity with consumption.

as an exogenous factor to make the empirical analysis fully aligned with the theoretical model. Nonetheless, I briefly discuss how the guardian can be included into an extended three-agent model in Section 6 where $\gamma(d)$ is determined endogenously. For the second period, I denote by $g(h)$ the return to human capital h. Given all other variables, the child chooses an optimal study time s that maximizes the total utility from two periods, i.e.

$$\begin{aligned} \max_{s} \quad & u_1^k(s, c_1^k) + \beta_k u_2^k(c_2), \\ s.t. \quad & c_1^k = \gamma(d) W_p(d), \\ & c_2 = g(h), \\ & h = f(d, s, c_1^k, h_0). \end{aligned}$$

Similarly, for the parent, let u_1^p, u_2^p be the utility of the parent in period 1 and 2, respectively, and β_p be parent's discounting factor. I assume that parent consumption in the first period c_1^p is a fixed proportion γ_p of parent income with $0 < \gamma_p < 1$ because the parents can decide on their spending regardless of the migration status. Note that $\gamma(d) + \gamma_p \leq 1$ because when parents migrate away, the de facto custodian might not spend all the remittances on the child. The parent maximizes total utility by choosing the optimal migration status d^*, i.e.

$$\begin{aligned} \max_{d} \quad & u_1^p(c_1^p) + \beta_p u_2^p(c_2), \\ s.t. \quad & c_1^p = \gamma_p W_p(d), \\ & c_2 = g(h), \\ & h = f(d, s, c_1^k, h_0). \end{aligned}$$

To derive the equilibrium, I make the following assumptions.

- $\frac{\partial u_j^i}{\partial c_j^i} > 0$ and $\frac{\partial^2 u_j^i}{\partial (c_j^i)^2} < 0$, where $i \in \{k, p\}$ and $j \in \{1, 2\}$, implying that the utility in each period increases while the marginal utility decreases in consumption in that period.
- $\frac{\partial u_1^k}{\partial s} < 0$ and $\frac{\partial^2 u_1^k}{\partial s^2} < 0$, implying a fatiguing effect of studying that is marginally increasing.
- $\frac{\partial f}{\partial s} \geq 0$, $\frac{\partial f}{\partial c_1^k} \geq 0$, $\frac{\partial^2 f}{\partial s^2} \leq 0$, $\frac{\partial^2 f}{\partial (c_1^k)^2} \leq 0$, implying that the study time and consumption weakly increase the production of human capital, but with decreasing marginal return.
- $\frac{\partial f}{\partial d} < 0$, implying that parental absence worsens child' human capital.

- $\frac{\partial g}{\partial h} \geq 0$ and $\frac{\partial^2 g}{\partial h^2} \leq 0$, implying that higher human capital of the child leads to higher income in the future, though with a decreasing marginal return.
- $\frac{\partial W_p}{\partial d} \geq 0$, implying the existence of monetary incentives to migrate.

For the rest of this section, I will provide qualitative analyses of the equilibrium solution from both the child and parent side. Formal mathematical derivations are relegated into Appendix A. In particular, I derive the closed-form solution for the equilibrium assuming specific functional forms of the utility functions in Appendix A.3, which yields the structural equation model that will be used for empirical analysis in Section 4.

2.2 Optimal decision of the child

For child utility maximization, there is a trade-off between current and future utility. Holding parent migration status d fixed, if study time s increases, the first-period utility decreases due to the fatiguing effect of studying, while the second-period utility increases because the child's human capital will increase due to increased study time, resulting in a higher future consumption c_2.

Intuitively, the child's optimal study time should be at the intersection of the marginal effect of study time on current utility ($\mathrm{MU}_1^k = -\frac{\partial u_1^k}{\partial s}$) and its marginal effect on future utility ($\mathrm{MU}_2^k = \frac{\partial u_2^k}{\partial s}$). The marginal effect of study time on current utility only depends on the level of study time. Holding the study time fixed, if the parent increases the proportion of days of migration, the marginal effect of study time on first-period utility is not affected, but the increased migration status will have a negative direct effect on human capital, and the indirect effect on human capital is undetermined since the change of investment in child is undetermined. Suppose the indirect effect of migration on human capital through investment in child is positive, the final effect on child's human capital still depends on the relative sizes of these direct and indirect effects. If the negative direct effect of migration dominates, then child's human capital worsens, leading to less income and consumption in the second period, so that the marginal utility from future consumption increases [2]. Graphically, the curve MU_1^k remains unchanged, while the curve MU_2^k shifts up, as shown in Figure 1. In general, the optimal child study time is increasing in migration status if the negative direct effect of migration dominates and is decreasing in migration status otherwise.

[2] See Appendix A.1 for mathematical details.

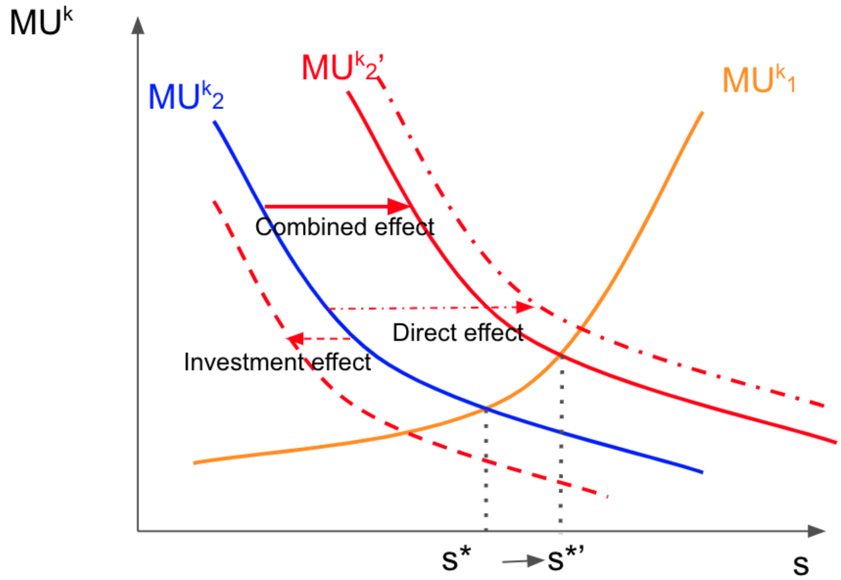


Figure 1: Tradeoff between current and future utility of the child. The optimal study time s is determined by the intersection of two marginal utility curves. When the parental migration status d increases, the MU_1^k curve remains unchanged while the MU_2^k curve shifts up or down, depending on the relative sizes of the direct and indirect effect through investment in the child

Due to the trade-off, under mild assumptions, we can find an interior solution $s^*(d)$. In Appendix A.3, I derive the closed-form solution of $s^*(d)$ assuming certain functional forms of the utility function. In general, assuming the existence of the interior optimal solution, I derive that [3]

$$\frac{\partial s^*}{\partial d} \propto -\Bigg(\overbrace{\frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d}}^{\text{Investment effect}} + \overbrace{\frac{\partial f}{\partial d}}^{\text{Direct effect}} \Bigg), \tag{1}$$

where $W_k(d) = \gamma(d)W_p(d)$, denoting the investment in the child, and $\propto$ denotes "proportional to", which hides a positive multiplicative factor. By the chain rule, $\frac{\partial W_k(d)}{\partial d} = \gamma(d)\frac{\partial W_p(d)}{\partial d} + W_p(d)\frac{\partial \gamma(d)}{\partial d}$. The decomposition of $\frac{\partial s^*}{\partial d}$ shows how the left-behind child's optimal study time changes when the parent migration status changes. $\frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d}$ represents the effect of migration on the child's human capital through investment in the child. $\frac{\partial f}{\partial d}$ measures the effect of parent absence on child's human capital. Since the sign of $\frac{\partial W_k(d)}{\partial d}$ is undetermined, equation (1) shows that the sign of $(\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k}{\partial d})$ is undetermined as

[3] See Appendix A.1 for the mathematical derivation.

well. If the indirect effect through investment in the child is positive, and the negative direct effect of being left-behind is greater in size, then $\frac{\partial s^*}{\partial d} \geq 0$, suggesting that the child will increase study time to compensate for worse human capital, and vice versa. This is consistent with the graphical illustration in Figure 1.

2.3 Optimal decision of the parent

For parent utility maximization, there is also a trade-off between current and future consumption. Holding the child's study time s fixed, if the parental migration status d increases, then the parent's first-period utility increases due to the increased consumption c_1^p, while the second-period utility decreases since the child's human capital will decrease due to the lack of parent accompaniment, resulting in a lower future consumption c_2.

Intuitively, the optimal parental migration status is at the intersection of the marginal effect of migration on current utility ($\text{MU}_1^p = \frac{\partial u_1^p}{\partial d}$) and its marginal effect on future utility ($\text{MU}_2^p = -\frac{\partial u_2^p}{\partial d}$). Parent utility in period 1 only depends on consumption levels. Holding the parental migration status constant, if the child increases the study time, it will not affect parent consumption or utility in period 1, but will decrease the marginal utility in period 2. This is because the increased study time will lead to higher human capital, which translates into higher income and higher consumption in period 2, so that the marginal utility from future consumption shifts down [4]. Graphically, the curve for marginal effect of migration on current utility remains unchanged, and the curve for marginal effect of migration on future utility shifts down, as shown in Figure 2. That is, the optimal migration decision is increasing in child study time. Unlike the child optimal decision process, the MU_2^p curve always shifts down as the proportion of days of migration increases.

Due to the trade-off, under mild assumptions, we can find an interior solution $d^*(s)$. In Appendix A.3, I derive the closed-form solution of $d^*(s)$ assuming certain functional forms of the utility function. In general, assuming the existence of the interior optimal solution, I derive that [5]

$$\frac{\partial d^*}{\partial s} \propto -\left(\overbrace{\frac{\partial f}{\partial c_1^k} \frac{\partial W_k(d)}{\partial d}}^{\text{Investment effect}} + \overbrace{\frac{\partial f}{\partial d}}^{\text{Direct effect}} \right), \tag{2}$$

[4] See Appendix A.2 for mathematical details.

[5] See Appendix A.2 for the mathematical derivation.

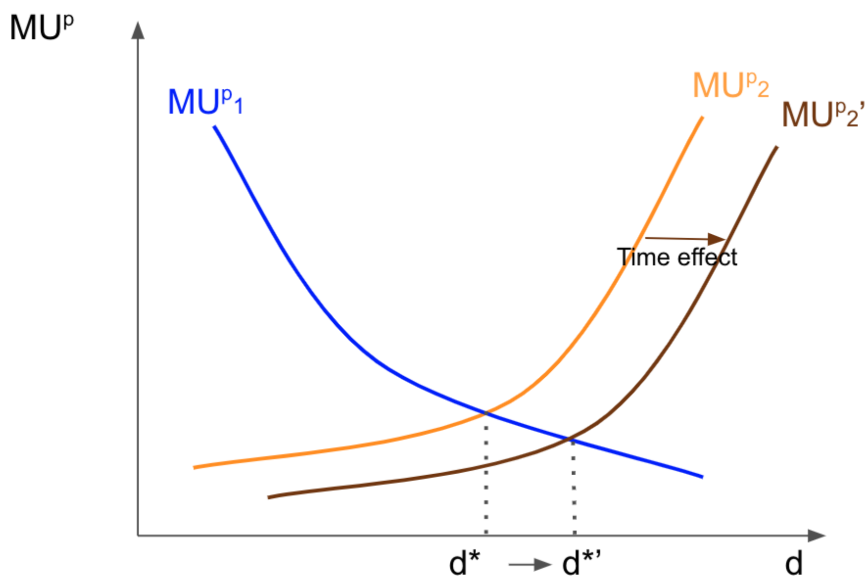


Figure 2: Tradeoff between current and future utility of the parent. The optimal migration status d^* is determined by the intersection of two marginal utility curves. When the child's study time s increases, the MU_1^p curve remains unchanged while the MU_2^p curve shifts down

The decomposition of $\frac{\partial d^*}{\partial s}$ shows how parent's optimal migration decision changes as child study time changes. The meaning of each part of $\frac{\partial d^*}{\partial s}$ is the same as in Equation (1). The marginal effect of parental migration on current utility is $\frac{\partial u_1^p}{\partial c_1^p}\frac{\partial c_1^p}{\partial d}$, and its marginal effect on future utility is $-\beta_p \frac{\partial u_2^p}{\partial h}(\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d})$. To guarantee an interior solution, we need the future marginal effect to be nonnegative, that is, $\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d} \leq 0$. Therefore, $\frac{\partial d^*}{\partial s} \geq 0$. This confirms our intuition that $\frac{\partial d^*}{\partial s}$ has a definite sign, unlike $\frac{\partial s^*}{\partial d}$. This is consistent to the graphical illustration in Figure 2.

2.4 Equilibrium solution

In Section 2.2 and Section 2.3, I show that the child's optimal decision on study time is a function of the parental migration status d, and the parent's optimal decision on migration is a function of the child study time s. Solving both equations will lead to the equilibrium. Under specific functional forms to the utility function, human capital production function, and wage function, I show that there is only one unique equilibrium solution [6]. Figure 3 is an illustration of the equilibrium solution.

[6]See Appendix A.3 for the mathematical details.

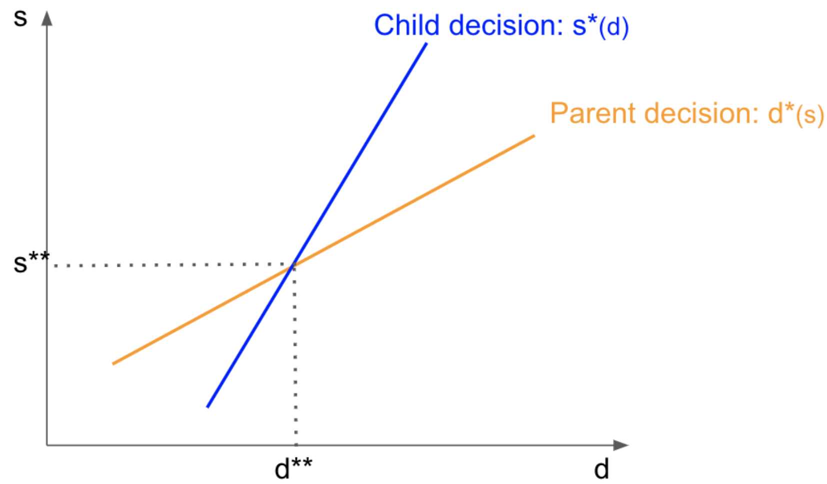


Figure 3: The equilibrium study time and migration status determined by the intersection of the child's optimal study time $s^*(d)$ as a function of d and the parent's optimal migration status $d^*(s)$ as a function of s

Given a migrant family, the theoretical model induces the following relationship among the observables – child's human capital h, investment in child W_k, the proportion of time that child's spent on studying s, the proportion of days of parent migration d, as well as other covariates X that account for the heterogeneity of families:

$$\begin{aligned} h &= f(d, s, W_k; X) \\ W_k &= \gamma(d) W_p(d; X) \\ s &= s^*(d; X) \\ d &= d^*(s; X) \end{aligned} \tag{3}$$

The main goal of this work is to study the effects of migration on child's human capital, i.e. $\frac{\partial h}{\partial d}$. By definition,

$$\frac{\partial h}{\partial d} = \underbrace{\frac{\partial f}{\partial d}}_{\text{Direct effect}} + \underbrace{\frac{\partial f}{\partial s}\frac{\partial s}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k}{\partial d}}_{\text{Indirect effects}}.$$

This yields the decomposition of the total effect into the direct effect and indirect effects. All parameters in the decomposition are necessary for answering my research question, and thus the first three equations in (3) must be estimated. By contrast, the effect of s on d, i.e. $\frac{\partial d}{\partial s}$, does not have to be known. As a result, we can simplify the last equation by computing $d^{**}(X)$ as the solution of $d = d^*(s^*(d, X), X)$ for each X, thereby facilitating the system. In Appendix A.3, I illustrate this step under a specific functional specification. To summarize, I focus on the following system:

$$\begin{aligned} h &= f(d, s, W_k; X) \\ W_k &= \gamma(d) W_p(d; X) \\ s &= s^*(d; X) \\ d &= d^{**}(X) \end{aligned} \tag{4}$$

It is worth emphasizing that the way to simplify (3) hinges on the research question and whether the simplification works depends on whether an effective identification strategy exists. For (4), I find a promising identification strategy as detailed in Section 4. In principle, we can also study how child's human capital affects parent's migration decision, which can be answered by (3) in theory. However, it is arguably more challenging to find a convincing identification strategy.

3 Data

3.1 Data source

The data used in this paper is collected by the Rural-Urban Migration in China (RUMiC) Project, which is a longitudinal survey (Institute of Labor Economics (IZA) et al., 2014). This project is a joint effort by the Australian University, University of Queensland, Beijing Normal University, and Institute for the Study of Labor (IZA). Starting in 2008, the project covers 9 provinces or province-level municipalities that are major sending or receiving areas of out-migration: Anhui, Chongqing, Guangdong, Hebei, Henan, Hubei, Jiangsu, Sichuan, and Zhejiang. The RUMiC survey includes 8,000 samples in rural household survey (RHS), 5,000 in urban household survey (UHS), and 5,000 in migrant household survey (MHS). Subjects in each category are randomly selected in each province. For detailed information on sampling design and tracking , see Gong et al. (2008); Meng et al. (2010); Kong (2010).

Although survey documents and data for both 2008 and 2009 are available, the 2008 data does not include important outcome variables such as children's exam scores or study time. For this reason, I mainly focus on the cross-sectional data in 2009 survey in this paper, and use the 2008 data for auxiliary purpose.

Since this paper focuses on rural migrants, data from RHS and MHS can both be used for analysis in principle. However, for the purpose of this paper, data from the RHS is preferable for several reasons. First, my paper is to compare left-behind children with children whose parent do not migrate. The RHS involves both groups, while the MHS only involves children of migrants. Second, RHS has substantially higher quality than MHS in terms of both the sample size and attrition rate (0.4% v.s. 58.4% attrition at the individual level, and 0.1% v.s. 63.6% at the household level, according to Akgüç et al. (2014)). Although the main analysis is based on 2009 data, the 2008 data is useful to impute for the missing values of demographic information, thereby increasing the effective sample size. As a result, RHS is more suitable for my analysis in terms of efficiency. Finally, this paper is focused on rural households. The RHS draws random samples from the annual household income and expenditure surveys carried out in rural villages, and tracks subjects having permanent living addresses. This makes the RHS a representative survey for my purpose.

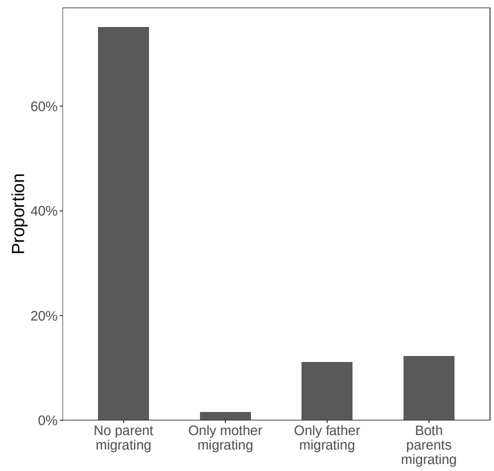
The raw data has 6899 children in 4843 households. To make the analysis meaningful, I only include school-age children (6-15 years old) who have never married and with parents older than 16 years old. After filtering, 2666 children in 2112 households are left in the data. I further exclude households for which the migration status is unreported, resulting in 1971 children in 1593 households. The parents in the data for my analysis come from 68 cities in 9 provinces, and their migration destinations spread over 137 cities in 29 provinces.

3.2 Descriptive statistics

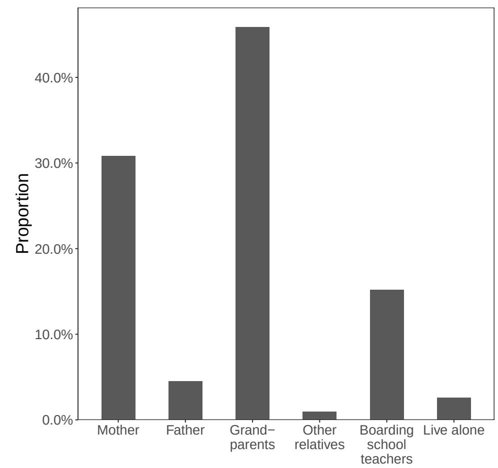
In this section, I provide some basic information of the data. Figure 4a shows the fractions of children with both parents migrating, with father migrating only, with mother migrating only and with neither parents migrating. In this figure, a person is counted as a migrant if she/he migrates for more than 90 days in the past year; see Section 3.3 for a more detailed description and explanation. Adding up the proportions of the first three categories, left-behind children account for roughly 30% of children in rural China.

Figure 4b shows the fractions of different guardians whom the left-behind children live with. When both parents migrate out, grandparents are most common de facto custodians. This shows a potential source of incentive compatibility on monetary investment since grandparents may not want to allocate most remittances on children's education or nutrition, or they may not have a good sense on the appropriate amount of money spent on children. The second most common guardians are boarding school teachers, who may not have strong incentives to take care of any single child.

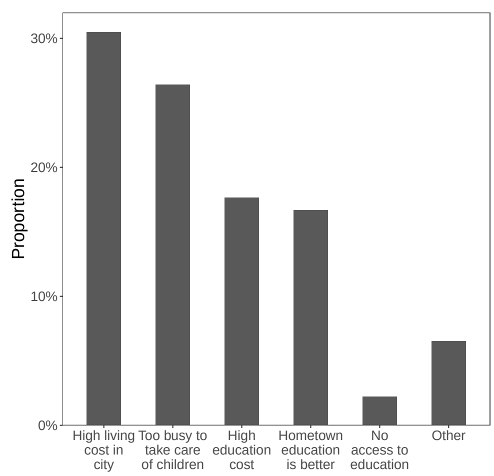
Figure 4c presents different reasons why parents do not bring children when migrating to work in cities. High living and education cost in cities appears to be the driving force to leaving children behind. This is partly because of the *hukou* restriction discussed in Section 1.1 that children with rural *hukou* cannot enjoy the social benefits such as education and housing. The lack of access to the social welfare system increases their living and education cost if they migrate with their parents. Another important motivation to leave children behind is that parents are too busy to take care of their children if they were brought along. This motivation is particularly strong when other family members who can play the role of caregivers, such as grandparents, are unable to migrate together.



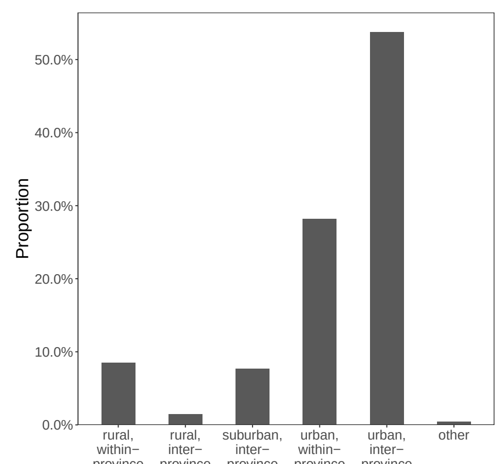
(a) Type of parental migration



(b) De facto custodian



(c) Reason for leaving children behind



(d) Migration destination (work location)

Figure 4: Descriptive statistics of data

Figure 4d shows the types of migration destinations. We can see that a majority of migrants move from rural to urban areas. Among rural-to-urban migrants, around two thirds of them move to an urban city in a different province.

3.3 Treatment variable: migration status

According to Meng and Yamauchi (2015), a good indicator for parental migration is based on very recent migration experience. Thus, I will define the migration status, the treatment variable in my problem, based on the duration of migration in the past year. In principle, I can define the migration status as the fraction of days migrating out in the past year as in the Section 2. However, this measure is not accurate since it is self-reported. In fact, I find that a large fraction of reported measures are multiples of 50, indicating potential lack of reliability of the measurement. Moreover, the estimated effects are harder to interpret and less transparent based on continuous treatment variables.

For these reasons, I will not use a continuous measure of migration but instead define a dummy variable D to represent the migration status where $D = 1$ if at least one parent of the child migrates out and leaves the child behind for at least 90 days in the past year. This measure is more accurate since there are questions that explicitly use 90 days as the threshold [7]. In addition, since the control group in this paper is children in rural areas with non-migrant parents, rather than children who migrate with their migrant parents, I exclude the children of the latter kind from the analysis. Figure 5 provides the diagram definition of D. Note that the dummy variable D is essentially $I(d \geq 90)$ where d is the decision variable of the parent in Section 2.

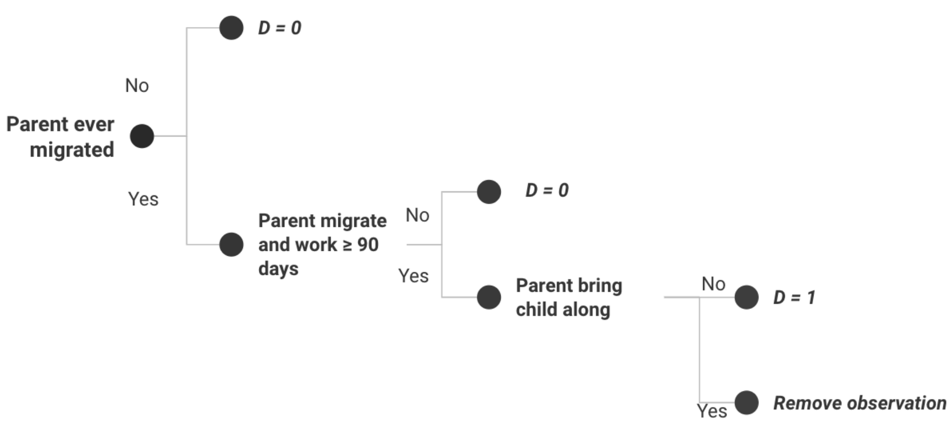


Figure 5: Definition of the treatment variable D

[7] For example, question C07_4 states that how many days did you work outside your hometown in 2008? (If none, please fill in '0', if ≥ 90, skip to C07_6)

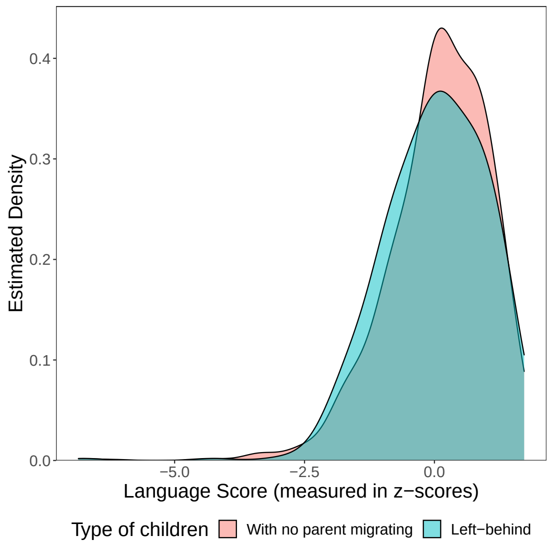
3.4 Mediation variables: study time and investment in the child

I use the weekly study time in hours reported by their guardians as a measure of the child's study time, denoted by T to be distinguished from the share of study time s used in Section 2. For investment in the child that corresponds to c_1^k in the theoretical model, I use the sum of the child's tuition at school, supplemental classes inside and outside of school, and food expenditure in 2008 reported by guardians. To stabilize the variance of the variable, I will use the logarithmic transformation of the total investment measured in Chinese Yuan as the mediation variable and denote it by W.

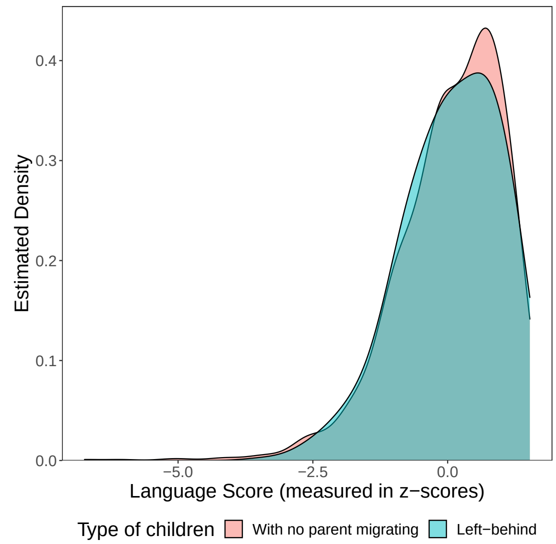
3.5 Dependent variables: standardized language and math scores

To measure the schooling performance, I choose the child exam scores, which are measures of the child human capital h in the model. In particular, I consider the final exam scores in language and math reported by parents or other guardians, who are informed of children's scores during parental meetings at school every semester. In addition, they would receive the hard copy of children's score reports from school at the end of every semester. Therefore, the reported scores are reliable. The test scores are also comparable across children in the sample since 7 out of 9 provinces use the same version of textbooks, while only a few villages in the remaining 2 provinces use another two versions of textbook. All of the three versions of textbooks and exams are designed closely following the curriculum standards designed by the Ministry of Education of China. Particularly, the materials are highly consistent for core subjects such as language and math. To make the scale of scores and estimated effects more interpretable, I convert test scores into z-scores by subtracting the average and dividing the standard error of the sample.

Figure 6 displays the distribution of exam scores. We could see that for left-behind children, the distribution of language scores is more left skewed, suggesting that these children perform worse in language exams on average. But the difference in math score distribution for left-behind children and other children is less pronounced. The marginal averages in two groups are reported in Table 3.1. We can see that left-behind children perform worse than children with non-migrant parents in language exams and slightly better in math exams, though neither of the differences is statistically significant at the 5% significance level. It is worth emphasizing that the marginal difference does reflect the effect of any kind because of the endogeneity and non-random missing values.



(a) Language score



(b) Math score

Figure 6: Distribution of testing scores

3.6 Control variables

As for other covariates, I include the personal characteristics of the child, such as the age, gender, height, weight, and birth weight. I also include parent-level characteristics such as the years of education of their parents. For those parents with missing values in these attributes in 2009, I impute them using the values reported in 2008 survey if available. For those parents who have records in both 2008 and 2009 surveys, if the measurements are inconsistent, I will choose the higher one. Other potentially important variables are excluded because they have a large fraction of missing values in the data for reasons that are hard to pinpoint.

Table 3.1 shows the summary statistics of a set of important control variables. We can see that left-behind children are significantly lighter and shorter than their counterparts. The difference in parent education levels in two groups is not statistically significant. In the empirical analysis, I control for covariates that are significantly different across treatment and control groups, and also include covariates that do not differ significantly to increase estimation efficiency.

Table 3.1: Summary Statistics

Variable	Migrant Parents	Non-migrant Parents	Difference (P-value)
Dependent Variables			
Language score	0.01	0.08	0.14
Math score	0.08	0.06	0.73
Covariates: Child			
Male	0.53	0.55	0.52
Age	11.27	11.52	0.07
Height	135.92	142.44	$< 0.001^{***}$
Weight	38.96	41.46	$< 0.001^{***}$
Birthweight	32.50	32.51	0.94
Covariates: Parents			
Mother edu year	7.50	7.36	0.21
Father edu year	8.23	8.19	0.70
Note:	*p<0.05; **p<0.01; ***p<0.001		

4 Empirical Framework

4.1 Structural equation model (SEM) for mediation analysis

Under specific functional forms as in Appendix A.3, I show that the system of equations (4) have the following linear forms:

$$P_i = \gamma_0 + \gamma_T \cdot T_i + \gamma_W \cdot W_i + \gamma_D \cdot D_i + \xi_P \cdot X_i + \epsilon_{Pi}, \tag{5}$$

$$T_i = a_T + b_T \cdot D_i + \xi_T \cdot X_i + \epsilon_{Ti}, \tag{6}$$

$$W_i = a_W + b_W \cdot D_i + \xi_W \cdot X_i + \epsilon_{Wi}, \tag{7}$$

$$D_i = \mathbb{1}(a_D + \xi_D \cdot X_i + \epsilon_{Di} \geq 0), \tag{8}$$

where P_i denotes the schooling performance of child i, measured by normalized final exam scores in language and mathematics as described in Section 3.5, T_i denotes the weekly study time in hours, and W_i denotes the logarithmic transformation of monetary investment and D_i denotes the dummy variable of the parental migration, as defined in Section 3. Recalling Section 3.3 that D is essentially $I(d \geq 90$, the last equation has a different form compared to the other three. To account for individual heterogeneity, other covariates and error terms are included. X_i is the set of control variables, including characteristics of children and parents introduced in Section 3.6. The error terms ϵ_{Pi}, ϵ_{Ti}, ϵ_{Wi}, and ϵ_{Di} are random errors, and we assume that they are correlated due to unobserved confounders in order to account for the endogeneity. When ϵ_{Di} is normal, the last equation is equivalent to a Probit model.

With an identified model, if we define δ to be the total effect of migration on children's schooling outcomes, then the total effect can be decomposed into the following three part:

$$\delta = \underbrace{\gamma_D(\text{parental absence})}_{\text{Direct effect}} + \underbrace{\gamma_T b_T(\text{study time}) + \gamma_W b_W(\text{investment})}_{\text{Indirect effects}}, \tag{9}$$

where γ_D captures the direct effect of migration, $\gamma_T b_T$ captures the indirect effect of migration through the child's study time, and $\gamma_W b_W$ captures the indirect effect of migration through investment in the child. Figure 7 illustrates the decomposition.

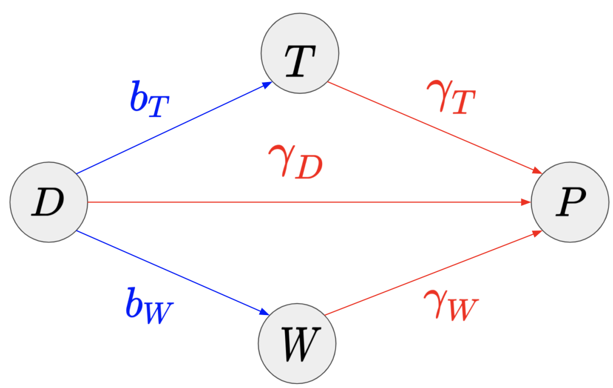


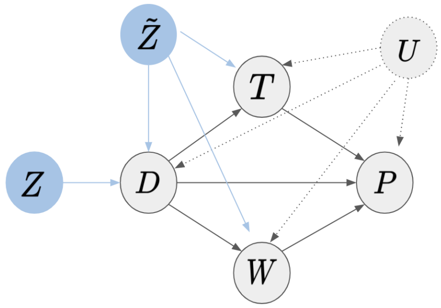
Figure 7: Decomposition of total effect into direct and indirect effects

4.2 Identification of coefficients via order condition

Since there are many unobserved factors that affect the parental migration decision, two mediators, and children's school performance simultaneously, the migration decision and mediators are all endogenous, rendering the standard mediation analysis, which assumes mutual independence between the errors $(\epsilon_{Pi}, \epsilon_{Ti}, \epsilon_{Wi}, \epsilon_{Di})$, implausible due to the omitted variable bias. For instance, variables such as child's self-control ability and parent's attitude toward child's education maybe correlated with both the parent migration decision and child's school performance, so the omission of such variables might lead to considerable bias.

To remove the confounding bias, I resort to an instrumental variable (IV) approach. Unlike the usual IV regression with a single endogenous variable and without mediators, the identification is more complicated for structural equation models with multiple endogenous variables. A necessary condition for identification is the order condition (e.g. Wooldridge, 2010), that is, for each equation in the system, the number of excluded exogenous variables, which includes both instrumental variables and other control covariates, should be larger than or equal to the number of included endogenous variables minus one. Although the order condition is only necessary but not sufficient, it is a simple and transparent condition to decide the necessary structure for identification. In the next subsection I will justify the rank condition, which is sufficient and necessary for identification.

Suppose we are able to find two sets of instrumental variables: Z that can only affect P through D, and $\tilde{Z}$ that can affect P through D or T or W. Both Z and $\tilde{Z}$ may involve multiple variables and we denote by $|Z|$ and $|\tilde{Z}|$ their sizes. Unlike Z which is required to be exogenous to both the outcome and mediators, $\tilde{Z}$ only needs to satisfy a weaker exclusion restriction. For instance, $\tilde{Z}$ is allowed to have direct effects on the mediators.



(a) Path diagram representation

$$\begin{aligned}
&\text{Eq. 1}: P \sim D + T + W \\
&\text{Eq. 2}: T \sim D + \tilde{Z} \\
&\text{Eq. 3}: W \sim D + \tilde{Z} \\
&\text{Eq. 4}: D \sim Z + \tilde{Z}
\end{aligned}$$

(b) Algebraic representation

Figure 8: Representations of the structural equation model for mediation analysis with instrumental variables

The permissible causal paths are illustrated in Figure 8a. The path diagram is a schematic representation of a structural equation model, where each node represents a variable and each directed edge from variable A to variable B represents the inclusion of A into the equation with B being the outcome. Note that the absence of an edge encodes an exclusion restriction while the presence of an edge imposes no constraint. Besides the outcome variable P, mediators T and W, and the treatment variable D, I also include the unmeasured confounders U, characterized by a dashed circle, as well as the instrumental variables Z and $\tilde{Z}$, characterized by blue circles. For simplicity I ignore other control covariates X that are allowed to link to all observed variables in the diagram. The four equations in Figure 8b are the algebraic representation of Figure 8a. More precisely, Z should be mean-independent of all residuals $(\epsilon_P, \epsilon_T, \epsilon_W, \epsilon_D)$ while $\tilde{Z}$ only needs to be mean-independent of ϵ_P.

The order condition can be read off from Figure 8b. For instance, for the equation of P, the number of endogenous variables is 4, given by P, T, W, D, and the number of excluded exogenous variables is $|Z| + |\tilde{Z}|$ since none of those instrumental variables are included in this equation. By contrast, for the equation of T, the number of endogenous variables is 2, given by T and D, while the number of excluded exogenous variables is only $|Z|$ since $\tilde{Z}$ is included. Table 4.1 summarizes these two quantities for each equation.

Table 4.1: Order condition for identification of the SEM in Figure 8b

Outcome variable	# Excluded Exogenous	# Included Endogenous - 1
P	$\|Z\| + \|\tilde{Z}\|$	3
T	$\|Z\|$	1
W	$\|Z\|$	1
D	0	0

Therefore, to identify all coefficients in the SEM, the order condition requires

$$|Z| + |\tilde{Z}| \geq 3, \quad |Z| \geq 1. \tag{10}$$

As a consequence, it suffices to find at least one IV in Z and at least two IVs in $\tilde{Z}$.

4.3 Identification of coefficients via rank condition

Rank condition is sufficient and necessary for identification. For the usual IV regression with one endogenous treatment variable, the order condition is equivalent to the exclusion restriction while the rank condition is equivalent to the exclusion restriction plus the relevance condition that requires the IV to be correlated with the treatment.

For general SEMs, the rank condition is much more complicated. Nonetheless, the SEM in Figure 8b is triangular, enabling a more transparent check of the rank condition. Recall that only five parameters need to be identified: $\gamma_T, \gamma_W, \gamma_D$ from the equation of P, b_T from the equation of T, and b_W from the equation of W. To identify b_T, it is sufficient to focus on the equations of T and D, which form an usual IV regression model with Z being the instrumental variable and $\tilde{Z}$ being a control covariate. The exclusion restriction is guaranteed by the order condition (10). As mentioned above, the rank condition still requires the relevance. In this case, it can be simply verified by the commonly-used F-test that tests if all coefficients of Z in the equation of D are zero. The same strategy can be applied to identify b_W. An asymptotically equivalent test in this case is the Anderson's canonical correlation test, which is designed for multiple endogenous variables and will be introduced in the next paragraph.

To identify γ_T, γ_W and γ_D, we can view T, W, D as three endogenous treatments for P. Then we can regard $Z, \tilde{Z}$ as instrumental variables for T, W, D because $Z, \tilde{Z}$ only affects P through T, W, D by definition. The well-known rule-of-thumb requires no fewer IVs than endogenous treatments. This is essentially the order condition and is guaranteed by (10) in my SEM. However, the relevance condition for multiple endogenous treatments is more involved. In my SEM, it requires the $3 \times (|Z| + |\tilde{Z}|)$ coefficient matrix by regressing (T, W, D) on $(Z, \tilde{Z})$, as well as other control covariates, to be full-rank (i.e. with rank 3). The usual F-test that borrows the heuristics from the single treatment case is flawed since it tests the wrong null hypothesis that the rank is zero instead of the correct null hypothesis that the rank is below 3. A more rigorous test is using the Anderson's canonical correlation LM statistic (Anderson, 1951), which is based on the smallest singular value of the coefficient matrix and thus testing the correct null. It is referred to as an underidentification test in the literature and is the default test in STATA `ivreg2` command (Baum et al., 2007).

In a nutshell, the rank condition can be justified by the order condition plus three underidentification tests, all based on Anderson's canonical correlation LM test. The triangular structure enables a clean interpretation of the identification strategy – when the rank condition holds, b_T, b_W are identified by Z and $\gamma_T, \gamma_W, \gamma_D$ are identified jointly by Z and $\tilde{Z}$. The relevance part of the rank condition is empirically testable so in the next subsection we will focus on justifying the exclusion restriction, which is generally untestable.

4.4 Choice of instrumental variables

I start with the choice of the instrumental variables Z. By definition, Z should be an IV that affects child performance only through the parental migration status. Some popular candidates for Z in the literature are religious preference uncommon in urban locations, dummy variable indicating whether the householder's first occupation was as a farmer, distance from home village to provincial capital, and average migration rate in the village (Fisher, 2005; Xiang et al., 2016; Meng and Yamauchi, 2015). However, these are not appropriate in my case. First, the religious tendencies are generally low in China and uncommon religious preferences are even rarer. Second, the householder's first occupation as a farmer is inappropriate since the share of farmers is predominantly high in rural China, implying a low variation of the occupation indicator. Third, the distance from home village to provincial capital suffers from lack of relevance for migrants in rural China because most people migrate to other provinces and thus the distance is less of a concern in deciding migration. The exclusion restriction of this variable is also likely violated because the general education facilities in regions closer to provincial capital tend to be better, leading to better schooling outcomes. Last, the average migration rate would not only influence the migration decision of each household, but also influence tax revenues and educational investment in that region, thereby influencing the schooling outcomes of children and violating the exclusion restriction.

In this paper, I use the method of Bartik et al. (1991) to construct shift-share instrumental variables. Bartik instruments are widely used in migration literature. They are correlated with migration decision, but are arguably exogenous in the equations of schooling performance, study time, and investment in the child, which makes them appealing valid IVs (Goldsmith-Pinkham et al., 2020). The Bartik-style instrument combines migrants' destination-industry information with changes in employment rate at destination by industry. The migration information is generated based on migrant's origin city, destination city, and the industry they work for using data from China 1% National Population Sample Survey 2005. The employment information is extracted from Urban Statistical Yearbook of China. The change in employment rate is generated using 2007 and 2008 employment data of each industry in all cities in China. These years are chosen such that there is sufficient time for migration flow to change as employment changes, but not too early so that the correlation between migration and employment would fade away. Specifically, the Bartik instrument is generated as below:

$$Z_{\text{ori,2008}} = \frac{\sum_{\text{des}} \sum_{\text{ind}} (\text{Mig}_{\text{ori,des,ind,2005}} \cdot \triangle\text{Employment}_{\text{des,ind,2007-2008}})}{\sum_{\text{des}} \sum_{\text{ind}} \text{Mig}_{\text{ori,des,ind,2005}}},$$

where ori denotes the origin city that the migrant is from, des denotes the destination city that the migrant moves to, ind denotes the industry that the migrant works in, $\text{Mig}_{ori,des,ind,2005}$ denotes the total number of migrant workers from city ori to city des that work in industry ind in 2005, and $\triangle\text{Employment}_{\text{des,ind,2007-2008}}$ denotes the growth rate of employment in industry ind in destination des from 2007 to 2008. Considering the difference between inter-city ($des \neq ori$) and within-city ($des = ori$) migration, I generate Bartik instruments for each, where the inter-city Bartik instrument takes the sum of d over all cities other than o while the within-city Bartik instrument solely takes $d = o$ into account. Note that having two variables in Z does not violate the order condition (10). Later in the empirical analysis, I will conduct Sargan test (Wooldridge, 2010) to test for overidentification.

As for $\tilde{Z}$, I choose three variables – rainfall shocks in previous years, the inherited gold and silver accessories, and the birth order among siblings. Abnormally high or low precipitation is detrimental to agricultural production, which is the main economic activity in rural China. Such weather shocks push rural residents to migrate away in search of more stable job opportunities, and also affect the time allocation and wealth of rural households. Apart from through these channels, weather itself can hardly affect the children's school performance directly. More specifically, the rainfall shock is measured based on the 19-year (between 1991 and 2009) annual precipitation data at the city level obtained from China's National Meteorological Information Center. For each city, data from 1991 to 2003 is used to find the mean and standard deviation of precipitation. For each year in 2004 to 2009, if the annual precipitation is over 1.5 standard deviations below or above the city's historical mean, I define the city to have experienced a rainfall shock. Then I calculate the total number of rainfall shocks experienced in each city during 2004 to 2009. A similar instrumental variable is constructed by Meng and Yamauchi (2015), although they treat high and low levels of precipitation asymmetrically while I treat them equally. The second variable, namely the inherited gold and silver accessories, measures the wealth level of older generations. It is thus unlikely confounded with the treatment, mediators, and outcome in the system. On the other hand, the variable may affect the current wealth level, through which affect the outcome, but it does not hurt identification since $\tilde{Z}$ is allowed to have such effects as shown in Figure 8a. The relaxed exclusion restriction, compared to that for Z, renders it a valid IV in this case. The third variable, namely the birth order among siblings, is random within a household and thus exogenous. Although it may affect the child's schooling performance through the investment due to the potentially unequal allocation between elder and younger children, this doesn't violate the requirement of $\tilde{Z}$ as specified in Figure 8a. Other than through migration, study time, and investment, it's unlikely that birth order among siblings can affect child's schooling performance. This justifies the relaxed exclusion restriction and renders it a valid IV. Finally, the order condition (10) only requires $|\tilde{Z}| \geq 1$ given that Z contains two variables, I choose more than one IV for $\tilde{Z}$. As with Z, I will conduct Sargan overidentification test in empirical analyses for $\tilde{Z}$.

4.5 Nonrandom missing patterns

The above subsections address one common source of endogeneity in variables of interest. In this subsection, I will focus on another source of endogeneity that originates from nonrandom missing patterns in the study time and investment, especially the former. It is unlikely that these two variables are missing at random conditionally on observed covariates because less caring parents may not know the child's education well and thus fail to report the information.

Previous studies simply remove observations with missing values in empirical analysis without accounting for nonrandom missing patterns. However, simply removing the observations with missing values in these variables may yield underestimation or overestimation of the negative effect of migration. Instead, I assume that the guardians reports the study time or investment in child only when a certain utility is above zero. When the utility is a linear function of the covariates with normal errors, this is precisely a Heckman model. In principle, the Heckman model can be added into the structural equation model directly and estimated using methods maximum likelihood. However, the non-standard form will significantly complicate the structure, making the estimation overly challenging. Therefore I apply a two-step procedure in which I first estimate the Heckman model for the study time and investment separately to impute the missing values, and then estimate the SEM using the imputed data.

5 Empirical Results

5.1 Main results on all samples

Since my goal is to investigate the effect of migration on both language and math scores, I consider the extension of (5) - (8) that includes two scores simultaneously:

$$\begin{aligned}
P_\ell &= \gamma_{0,\ell} + \gamma_{T,\ell} \cdot T + \gamma_{W,\ell} \cdot W + \gamma_{D,\ell} \cdot D + \xi_{P,\ell} \cdot X + \epsilon_{P,\ell}, \\
P_m &= \gamma_{0,m} + \gamma_{T,m} \cdot T + \gamma_{W,m} \cdot W + \gamma_{D,m} \cdot D + \xi_{P,m} \cdot X + \epsilon_{P,m}, \\
T &= a_T + b_T \cdot D + \xi_T \cdot X + \epsilon_T, \\
W &= a_W + b_W \cdot D + \xi_W \cdot X + \epsilon_W, \\
D &= \mathbb{1}(a_D + \xi_D \cdot X + \epsilon_D \geq 0),
\end{aligned}$$

where ℓ and m in the subscripts are short for "language" and "math", and the subscript i for each unit is suppressed for notational convenience. Recall that P_ℓ and P_m are standardized z-scores, T measures the study time in hours, W is the logarithmic transformation of spending measured in Chinese Yuan, D is the binary migration decision. Therefore, $\gamma_{D,\ell}, \gamma_{D,m}, b_T, b_W$ measure the average difference of P_ℓ, P_m, T, W between left-behind children and non-left-behind children, $\gamma_{T,\ell}, \gamma_{T,m}$ measure the improvement of P_ℓ, P_m when the study time increases by one hour, and $\gamma_{W,\ell}, \gamma_{W,m}$ measure the improvement of P_ℓ, P_m multiplied by 100 when the investment increases by 1%.

Under this specification, the direct and indirect effects are $(\gamma_{D,\ell}, \gamma_{T,\ell} b_T, \gamma_{W,\ell} b_W)$ and $(\gamma_{D,m}, \gamma_{T,m} b_T, \gamma_{W,m} b_W)$ for language and math scores, respectively. Here I allow the error terms $\epsilon_{P,\ell}$ and $\epsilon_{P,m}$ to be correlated. This expanded SEM can capture the high correlation between the language and math scores. It is easy to see that the order condition remains the same as (10), and the rank condition can be tested exactly in the same way as in Section 4.3. Later on I will suppress the subscripts ℓ and m when no confusion can arise.

The top panel of Table 5.1 shows the direct and indirect effects of migration using all samples. The two columns report the effect on normalized language scores and math scores estimated with the strategy introduced in Section 4. In the bottom panel, I report the underidentification, overidentification, and endogeneity tests separately for study time, investment, language score, and math score. Note that the underidentification results are the same for study time and investment because they are modelled simultaneously, and the same reasoning applies for the underidentification results for language and math scores. All underidentification tests reject the null at the 0.1% significance level, suggesting strong evidence that the rank condition holds. The test results for overidentification suggest that no evidence has been found against the null hypothesis that over-identifying restrictions are valid. As for the endogeneity test, although there is no evidence against the exogeneity of migration decision in the equation for study time, there is strong evidence against it in the equation for investment. In addition, there is strong evidence against the exogeneity of migration, study time, and investment jointly in the equations for exam scores. This marks the importance of accounting for the endogeneity.

For indirect effects, we report p-values from the joint significance test. Although this test cannot be inverted to a confidence interval as opposed to Sobel test, it is valid for testing the null effect. In addition, it is found to be powerful compared to alternative methods (e.g. Fritz and MacKinnon, 2007; Hayes and Scharkow, 2013).

Recall that the net effect of migration is the summation of the following three effects: direct effect of migration γ_D, which is the effect of parent absence; indirect effect of migration through the child's study time $\gamma_T b_T$; and the indirect effect through investment in the child $\gamma_W b_W$. Adding up the direct and indirect effects, left-behind children perform 1.42 standard deviations worse than children with non-migrating parents in language, and 1.33 standard deviations worse in math.

From the first column of Table 5.1, the direct effect of migration on language score is -0.52 standard deviations and significant at the 1% level. From the second column of Table 5.1, the direct effect on math score is -0.45 standard deviations and significant at the 1% level.

As for the indirect effects shown in Table 5.1, the effect of migration on exam scores through the child's study time is almost zero and not significant. This might be because migration has little effect on child's study time, or because the child's study time doesn't change test scores by a lot. To further investigate the cause, I decompose this indirect effect

Table 5.1: Effect of Parental Migration on Child Schooling Outcomes (All Sample)

			(1) Language	(2) Math
Direct Effect				
Parental Accompany			-0.524**	-0.453**
			(0.002)	(0.006)
Indirect Effect				
Study time			0.003	0.002
			(0.096)	(0.406)
Investment in children			-0.894***	-0.874***
			(0.000)	(0.001)
Sepecification Tests	*(1) Study time*	*(2) Investment*	*(3) Language*	*(4) Math*
Underidentification test				
(Anderson canon. corr. LM statistic)	29.868***	29.868***	29.859***	29.859***
	(0.000)	(0.000)	(0.000)	(0.000)
Overidentification test (Sargan statistic)	2.752	0.607	0.360	0.306
	(0.097)	(0.436)	(0.835)	(0.858)
Endogeneity test	2.366	10.199***	26.587***	30.409***
	(0.124)	(0.001)	(0.000)	(0.000)
Obs.		1971		

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

into γ_T and b_T in the top panel of Table 5.2. The first column shows the estimate of b_T, namely the effect of migration on child's study time. The second and third columns show the estimate of γ_T, namely the effect of study time on child's language and math scores respectively. The coefficient b_T shows that left-behind children spend less time studying ($b_T = -0.29$ with $p = 0.083$) than children with non-migrant parents on average, but the difference is not statistically significant. The effect of study time on test scores is neither large nor significant ($|\gamma_T| \leq 0.012$ with $p \geq 0.096$ for both language and math scores).

Table 5.1 also show that parental migration has significant (at 0.1% significance level) and large negative indirect effects on both scores through investment, which are almost doubled in size compared to the corresponding direct effects. This finding is perhaps surprising since most of existing works conclude that remittances have positive effects on the child's educational outcomes. To dig into it, I show the decomposition of this indirect effect into γ_W and b_W in the bottom panel of Table 5.2. The first column shows the estimate of b_W, namely the effect of migration on investment in the child, and the second and third columns show the estimates of γ_W, namely the effects of investment on child's language and math scores respectively. It turns out that the investment in left-behind children is significantly lower ($b_W = -2.38$ with $p < 0.001$) than that in children who are not left behind, despite that investment is beneficial to child's school performance ($\gamma_W \geq 0.37$ with $p \leq 0.001$ for both language and math scores). This implies that the negative indirect effect through investment is driven by underinvestment.

The results in Table 5.1 are based on a careful account of both the endogeneity of

Table 5.2: Decomposition of Indirect Effects of Migration (All Sample)

	(1) Mediator	(2) Language	(3) Math
Through Study Time			
Migration (b_T)	-0.294		
	(0.083)		
Study time (γ_T)		-0.012	-0.006
		(0.096)	(0.406)
Through Investment			
Migration (b_W)	-2.375***		
	(0.000)		
Investment (γ_W)		0.376***	0.368***
		(0.000)	(0.001)

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

migration/mediators and non-random missing mediators as discussed in Section 4. In Appendix B, I present the results without accounting for one of them or both of them as sanity checks, showing that failure of addressing these issues tends to underestimate the direct effect and indirect effect through investment drastically, despite that all analyses are consistent in the signs of the effects.

5.2 Exploring heterogeneous treatment effects

In this subsection, I investigate the heterogeneous treatment effects in different subgroups. In particular, I am interested in subgroups stratified by gender and child birth order. The different attitude of guardians toward boys and girls, as well as the role that the eldest and younger children play in multiple-children families will probably lead to heterogeneous treatment effects when parents migrate away. For each subgroup, I estimate the same SEM and present the results in Table 5.3 and Table 5.5.

Table 5.3 shows the effect of parental migration on left-behind boys and girls separately. Note that for migration in the equations with study time and investment being the outcome and Z being the IVs, the underidentification test statistic is marginally significant ($p = 0.052$) for girls, suggesting a reasonably strong evidence for the rank condition. The other underidentification tests all show strong evidence for the rank condition. The overidentification tests and endogeneity tests yield qualitatively the same results as those for all samples.

In terms of direct effects, left-behind boys are more negatively affected in language scores, which is consistent with the finding for all samples in Table 5.1. By contrast, left-behind girls are almost equally affected in language scores and math scores with significant and large negative effect sizes.

In terms of indirect effects, neither left-behind boys nor girls are largely or significantly

Table 5.3: Effect of Parental Migration on Child Schooling Outcomes (Subgroup by Gender)

	Girl		Boy	
	(1) Language	(2) Math	(3) Language	(4) Math
Direct Effect				
Parental Accompany	-0.413*	-0.424*	-0.351**	-0.207
	(0.015)	(0.030)	(0.008)	(0.074)
Indirect Effect				
Study time	0.002	0.003	-0.006	0.000
	(0.602)	(0.474)	(0.340)	(0.974)
Investment in children	-1.393*	-1.621*	-0.124**	-0.115**
	(0.010)	(0.010)	(0.008)	(0.003)
Sepecification Tests	*(1) Study time*	*(2) Investment*	*(3) Language*	*(4) Math*
Girl				
Underidentification test				
(Anderson canon. corr. LM statistic)	5.904	5.904	13.599**	13.599**
	(0.052)	(0.052)	(0.004)	(0.004)
Overidentification test (Sargan statistic)	0.903	1.938	1.486	2.573
	(0.342)	(0.164)	(0.476)	(0.276)
Endogeneity test	0.800	9.201**	9.436*	15.019**
	(0.371)	(0.002)	(0.024)	(0.002)
Obs.		887		
Boy				
Underidentification test				
(Anderson canon. corr. LM statistic)	28.100***	28.100***	20.772***	20.772***
	(0.000)	(0.000)	(0.000)	(0.000)
Overidentification test (Sargan statistic)	2.622	0.001	1.794	2.798
	(0.105)	(0.981)	(0.408)	(0.247)
Endogeneity test	1.156	1.696	18.078***	12.723**
	(0.282)	(0.193)	(0.000)	(0.005)
Obs.		1084		

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

affected through study time. The decomposition of this indirect effect is presented in the top panel of Table 5.4. Left-behind girls experience large and significant reductions in study time, while left-behind boys tend to spend more time studying, though the effect is insignificant and cannot be viewed as positive on this data. As with the analysis for all samples, study time has no significant effect on either test scores.

Unlike the effect through study time, left-behind boys and girls are all significantly affected through investment. Notably, the girls are suffering ten times more than boys through this mechanism, with huge effects that are more than 1 standard deviation in sizes. The decomposition of this indirect effect is presented in the bottom panel of Table 5.4. Compared with left-behind boys, left-behind girls are suffering from a much severer reduction in investment, and their scores are also more vulnerable to underinvestment. This finding reveals a shocking gender inequality in rural China, at least among the left-behind children.

Table 5.4: Decomposition of Indirect Effects of Migration (Subgroup by Gender)

	Girl			Boy		
	(1) Mediator	(2) Language	(3) Math	(4) Mediator	(5) Language	(6) Math
Through Study Time						
Migration (b_T)	-0.524*			0.400		
	(0.022)			(0.340)		
Study time (γ_T)		-0.004	-0.006		-0.016	0.000
		(0.602)	(0.474)		(0.096)	(0.974)
Through Investment						
Migration (b_W)	-3.514*			-0.967***		
	(0.010)			(0.001)		
Investment (γ_W)		0.397**	0.461**		0.128**	0.119**
		(0.008)	(0.008)		(0.008)	(0.003)

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These results suggest that in order to increase left-behind girls' school performances, policies targeted at increasing the monetary investment in them should in general be much more effective than policies targeting at increasing their study time.

Similar to the gender subgroup analysis, I also investigate the heterogeneous treatment effects with respect to the birth order. In particular, I conduct separate mediation analyses for the eldest and younger children in households with more than one child. Since the birth order is used as an instrumental variable, I replace this IV with the number of regular residents in the household. The number of regular residents is relatively exogenous. On the other hand, it unlikely affects child's schooling performance through mechanisms other than migration, household wealth and child's study time. Thus it also meets the requirements for $\tilde{Z}$, as shown in Figure 8a.

For households with multiple children, usually the eldest child takes care of the younger

Table 5.5: Effect of Parental Migration on Child Schooling Outcomes (Subgroup by Birth Order)

	First child		Subsequent children	
	(1) Language	(2) Math	(3) Language	(4) Math
Direct Effect				
Parental Accompany	-0.696*	-0.696*	-0.289	-0.052
	(0.012)	(0.013)	(0.056)	(0.696)
Indirect Effect				
Study time	-0.014	-0.013	-0.001	-0.011*
	(0.399)	(0.399)	(0.779)	(0.019)
Investment in children	-0.295	-0.359*	-0.470*	-0.205
	(0.092)	(0.047)	(0.043)	(0.310)
Sepecification Tests	*(1) Study time*	*(2) Investment*	*(3) Language*	*(4) Math*
First child				
Underidentification test				
(Anderson canon. corr. LM statistic)	18.535***	18.535***	7.175*	7.175*
	(0.000)	(0.000)	(0.028)	(0.028)
Overidentification test (Sargan statistic)	0.050	2.671	0.037	0.612
	(0.823)	(0.102)	(0.848)	(0.434)
Endogeneity test	2.011	0.959	12.934**	16.820***
	(0.156)	(0.327)	(0.005)	(0.001)
Obs.		891		
Subsequent children				
Underidentification test				
(Anderson canon. corr. LM statistic)	7.183*	7.183*	7.598*	7.598*
	(0.028)	(0.028)	(0.022)	(0.022)
Overidentification test (Sargan statistic)	0.619	1.179	0.077	0.247
	(0.431)	(0.278)	(0.781)	(0.619)
Endogeneity test	6.058*	9.804**	6.815	4.326
	(0.014)	(0.002)	(0.078)	(0.228)
Obs.		860		

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.6: Decomposition of Indirect Effects of Migration (Subgroup by Birth Order)

	First child			Subsequent children		
	(1) Mediator	(2) Language	(3) Math	(4) Mediator	(5) Language	(6) Math
Through Study Time						
Migration (b_T)	0.528			-0.641**		
	(0.399)			(0.003)		
Study time (γ_T)		-0.027	-0.025		0.002	0.017*
		(0.110)	(0.167)		(0.779)	(0.019)
Through Investment						
Migration (b_W)	-1.226*			-2.251*		
	(0.044)			(0.012)		
Investment (γ_W)		0.241	0.293*		0.209*	0.091
		(0.092)	(0.047)		(0.043)	(0.310)

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

children and provide their younger siblings with emotional support. As a consequence, I expect that the role of parent partially shifts to the eldest child when parents migrate away, so the subsequent children may suffer less than the eldest child due to parent absence. This is confirmed by the results in Table 5.5. Due to the absence of parents, the eldest children have approximately 0.7 standard deviations lower scores on average in language and math exams than their non-migrant counterparts, and these effects are significant at the 1% level approximately. They are much larger than the direct effects of migration on the subsequent children.

The indirect effects through study time are small in sizes for both subgroups. Even though the effect on subsequent children's math scores is significant, I will not over-interpret it due to the tiny effect size. The top panel of Table 5.6 shows the decomposition of this effect into b_T and γ_T. We can observe that the eldest child tend to spend more time studying, though the effect is insignificant and cannot be interpreted as positive, while the subsequent children experience significant and large reductions in study time. In line with the results for all samples, the study time has tiny effect on test scores.

As opposed to the effects through study time, the effects through investment are large in sizes for both subgroups. In particular, the effect on math scores is significant for the eldest child and the effect on language scores is significant for subsequent children. The decomposition of the indirect effect through investment is presented in the bottom panel of Table 5.6. Compared with the eldest child, subsequent children suffer from a more severe reduction in investment. From the estimates of γ_W's, we see that the eldest child's math performance is more vulnerable to underinvestment, while the subsequent child's language performance is more vulnerable to underinvestment. This might be partially explained by the different cognitive development stages of these children.

5.3 Extended analysis

Previous results reveal notable effects of migration through investment in the child. To decouple the contributions of different types of investment, I further decompose it into nutrition spending and course-related spending, which includes expenditure on tuition and remedial classes at school and outside school. I refer to the latter as tuition spending for simplicity. In this case, three mediators are involved — the child's study time and two types of investment. The SEM to estimate becomes slightly more complicated:

$$\begin{aligned}
P_\ell &= \gamma_{0,\ell} + \gamma_{T,\ell} \cdot T + \gamma_{W,\text{tu},\ell} \cdot W_{\text{tu}} + \gamma_{W,\text{nu},\ell} \cdot W_{\text{nu}} + \gamma_{D,\ell} \cdot D + \xi_{P,\ell} \cdot X + \epsilon_{P,\ell}, \\
P_m &= \gamma_{0,m} + \gamma_{T,m} \cdot T + \gamma_{W,\text{tu},m} \cdot W_{\text{tu}} + \gamma_{W,\text{nu},m} \cdot W_{\text{nu}} + \gamma_{D,m} \cdot D + \xi_{P,m} \cdot X + \epsilon_{P,m}, \\
T &= a_T + b_T \cdot D + \xi_T \cdot X + \epsilon_T, \\
W_{\text{tu}} &= a_{W,\text{tu}} + b_{W,\text{tu}} \cdot D + \xi_{W,\text{tu}} \cdot X + \epsilon_{W,\text{tu}}, \\
W_{\text{nu}} &= a_{W,\text{nu}} + b_{W,\text{nu}} \cdot D + \xi_{W,\text{nu}} \cdot X + \epsilon_{W,\text{nu}}, \\
D &= \mathbb{1}(a_D + \xi_D \cdot X + \epsilon_D \geq 0),
\end{aligned}$$

where the subscript i for each unit is suppressed for notational convenience. Using a similar argument as in Section 4.2 see that the order condition becomes

$$|Z| + |\tilde{Z}| \geq 4, \quad |Z| \geq 1.$$

It is clear that the set of instrumental variables in Section 4.4 satisfies this condition. The rank condition can be tested using the same strategy as in Section 4.3.

The results are presented in Table 5.7. The model specification results, direct effects and indirect effects through the child's study time are all consistent with Table 5.1. Decomposition of the effects through study time is presented in the top panel of Table 5.8, which shows qualitatively the same results as previous analyses. As for the investment, the effect through both tuition spending and nutrition spending are significant with large negative effect sizes, and the effect through nutrition spending is larger in sizes. Further decomposing the indirect effects into γ_W and b_W, as shown in the bottom panel of Table 5.8, the left-behind children are suffering from underinvestment in both tuition and nutrition, and the latter is more severe. Therefore, policies targeting at increasing investment in the child should put more weights on nutrition. For instance, conditional cash transfer programs that improve these children's food intakes and nutrition status could be more effective in increasing their school performances than tuition waivers.

6 Conclusion and Discussions

In this paper, I disentangle the total effect of parental out-migration to the child's schooling performance into three mechanism-specific effects through parental absence, child's study time, and investment in the child via a mediation analysis. Using the RUMiC data on rural households from nine provinces in China, I find that the effects through parental absence and investment are both significantly negative with large sizes, while the effect through child's study time is insignificant with a negligible size. The surprising

Table 5.7: Effect of Parental Migration on Child Schooling Outcomes (Nutrition and Tuition)

				(1) Language	(2) Math
Direct Effect					
Parental Accompany				-0.431**	-0.357*
				(0.003)	(0.011)
Indirect Effect					
Study time				0.004	0.002
				(0.092)	(0.426)
Tuition				-0.495**	-0.490**
				(0.007)	(0.007)
Nutrition				-0.950***	-0.895***
				(0.001)	(0.001)
Sepecification Tests	*(1) Study time*	*(2) Tuition*	*(3) Nutrition*	*(4) Language*	*(5) Math*
Underidentification test					
(Anderson canon. corr. LM statistic)	29.868***	29.868***	29.868***	14.696***	14.696***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Overidentification test (Sargan statistic)	2.752	1.189	0.559	0.287	0.713
	(0.097)	(0.276)	(0.454)	(0.592)	(0.398)
Endogeneity test	2.366	2.865	13.350***	26.117***	29.262***
	(0.124)	(0.091)	(0.000)	(0.000)	(0.000)
Obs.			1971		

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.8: Decomposition of Indirect Effects of Migration (Nutrition and Tuition)

	(1) Study Time	(2) Tuition	(3) Nutrition	(4) Language	(5) Math
Through Study Time					
Migration (b_T)	-0.390*				
	(0.012)				
Study Time (γ_T)				-0.012	-0.005
				(0.092)	(0.426)
Through Investment					
Migration (b_W)		-2.472***	-4.128***		
		(0.001)	(0.001)		
Tuition ($\gamma_{W,\text{tu}}$)				0.200**	0.198**
				(0.007)	(0.007)
Nutrition ($\gamma_{W,\text{nu}}$)				0.230***	0.217***
				(0.001)	(0.001)

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

negative effect through investment is mainly driven by reduced nutrition investment by de facto custodians, who may not have compatible incentives to allocate the remittances on the child. The subgroup analysis reveals a shocking gender inequality that girls are suffering ten times more from the underinvestment than boys. These mechanism-specific effects show relative importance of each policy interventional target, thereby providing stronger policy implications than the net effect estimated in previous works. For example, the findings in this paper suggest that policies which compensate for underinvestment, especially for girls and younger children in the household, tend to be more effective in mitigating the negative effect of migration than other types of policies. In particular, policies that increase the nutrition spending on left-behind children also tends to be effective in improving the human capital of left-behind children.

There are a few extensions that are worth discussing. First, the negative effect through investment corroborates the importance to study the role of de facto custodians. As mentioned in Section 2.1, we can consider a three-agent model with the guardian included. Instead of taking $\gamma(d)$ as an exogeneous factor, we can view it as the decision variable of the guardian. Denote it by γ for simplicity. The first-period utility function of the guardian depends on the his/her own consumption, which is $(1-\gamma-\gamma_p)W_p(d)$ by definition. The equilibrium solution for γ is determined by how much the guardian care about the future human capital of the child, which can be characterized by the discount factor for the second-period utility. We can study whether the equilibrium proportion of household income invested in the child is increasing or decreasing in d. Empirically, one needs to collect more information about the de facto custodians in order to estimate this part of the model.

Second, the other facet of the problem, namely the effect of the child's schooling performance on parent migration decision, is also interesting and of no less policy importance, as pointed out in Section 2.4. For this research question, the child's schooling performance becomes the major endogenous variable. However, compared to parental migration decision, it is much more challenging to find valid instrumental variables. For instance, the school quality may affect both the performance and study time, implying that school-level information is needed. However, such information is not available in the RUMiC survey, which is focused on adult migrants rather than their children. A well-designed child-centered survey is needed to address this question.

Finally, as shown in Section 5.3, it is straightforward to decompose the pathways or add other pathways into the empirical framework. For instance, one can also investigate the effect through the child's time spent on housework since this may have a negative effect due to fatigue or a positive effect due to the aspiration to leave rural areas. Furthermore, the methodology is quite flexible and can be applied to evaluate the effect of other types of parents' labor market participation on child education.

Appendix

A Mathematical Details of The Two-agent Model

A.1 Child utility maximization

The utility of child is

$$\begin{aligned}\max_{s} \quad & u_1^k(s, c_1^k) + \beta_k u_2^k(c_2), \\ s.t. \quad & c_1^k = \gamma(d) W_p(d), \\ & c_2 = g(h), \\ & h = f(d, s, c_1^k, h_0).\end{aligned} \tag{11}$$

Plugging constraints to utility function

$$L^k = u_1^k(s, \gamma(d)W_p(d)) + \beta_k u_2^k(g(f(d, s, c_1^k, h_0)))$$

Taking the derivative with respect to s and obtain the first order condition

$$\frac{\partial L^k}{\partial s} = \frac{\partial u_1^k}{\partial s} + \beta_k \frac{\partial u_2^k}{\partial c_2} \frac{\partial g}{\partial h} \frac{\partial f}{\partial s} = 0.$$

The marginal effect of studying time on current utility is $\text{MU}_1^k = -\frac{\partial u_1^k}{\partial s}$, and its marginal effect on future utility is $\text{MU}_2^k = \beta_k \frac{\partial u_2^k}{\partial c_2} \frac{\partial g}{\partial h} \frac{\partial f}{\partial s}$. Note that we assume $c_1^k = \gamma(d) W_p(d) = W_k(d)$ for simplicity here, so we don't assume a certain sign on $\frac{\partial W_k(d)}{\partial d}$. The goal is to study the effect of d on s^*, so further take the derivative of $\frac{\partial L^k}{\partial s}$ with respect to d,

$$\begin{aligned}\frac{\partial^2 L^k}{\partial s \partial d} =& \frac{\partial^2 u_1^k}{\partial s^2} \frac{\partial s}{\partial d} + \frac{\partial^2 u_1^k}{\partial s \partial c_1^k} \frac{\partial c_1^k}{\partial d} + \beta_k A (\frac{\partial f}{\partial d} + \frac{\partial f}{\partial s} \frac{\partial s}{\partial d} + \frac{\partial f}{\partial c_1^k} \frac{\partial W_k(d)}{\partial d}) + \\ & \beta_k \frac{\partial u_2^k}{\partial c_2} \frac{\partial g}{\partial h} \frac{\partial^2 f}{\partial s^2} \frac{\partial s}{\partial d} + \beta_k \frac{\partial u_2^k}{\partial c_2} \frac{\partial g}{\partial h} (\frac{\partial^2 f}{\partial s \partial d} + \frac{\partial^2 f}{\partial s \partial c_1^k} \frac{\partial W_k(d)}{\partial d}) = 0,\end{aligned}$$

where

$$A = \frac{\partial^2 u_2^k}{\partial c_2^2} (\frac{\partial g}{\partial h})^2 \frac{\partial f}{\partial s} + \frac{\partial u_2^k}{\partial c_2} \frac{\partial f}{\partial s} \frac{\partial^2 g}{\partial h^2} < 0.$$

Therefore,

$$\frac{\partial s^*}{\partial d} = -\frac{\beta_k A \left(\overbrace{\frac{\partial f}{\partial c_1^k} \frac{\partial W_k(d)}{\partial d}}^{\text{Investment effect}} + \overbrace{\frac{\partial f}{\partial d}}^{\text{Direct effect}} \right) + \beta_k \frac{\partial u_2^k}{\partial c_2} \frac{\partial g}{\partial h} (\frac{\partial^2 f}{\partial s \partial d} + \frac{\partial^2 f}{\partial s \partial c_1^k} \frac{\partial W_k(d)}{\partial d}) + \frac{\partial^2 u_1^k}{\partial s \partial c_1^k} \frac{\partial W_k(d)}{\partial d}}{\frac{\partial^2 u_1^k}{\partial s^2} + \beta_k \frac{\partial u_2^k}{\partial c_2} \frac{\partial g}{\partial h} \frac{\partial^2 f}{\partial s^2} + \beta_k A \frac{\partial f}{\partial s}}.$$

If we further assume the separability of the child utility function and human capital production function, we will get rid of terms of $\frac{\partial^2 u_1^k}{\partial s \partial c_1^k}$,$\frac{\partial^2 f}{\partial s \partial d}$, and $\frac{\partial^2 f}{\partial s \partial c_1^k}$, then $\frac{\partial s}{\partial d}$ is simplified to

$$\frac{\partial s^*}{\partial d} = -\frac{\beta_k A\Bigg(\overbrace{\frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d}}^{\text{Investment effect}} + \overbrace{\frac{\partial f}{\partial d}}^{\text{Direct effect}}\Bigg)}{\frac{\partial^2 u_1^k}{\partial s^2} + \beta_k \frac{\partial u_2^k}{\partial c_2}\frac{\partial g}{\partial h}\frac{\partial^2 f}{\partial s^2} + \beta_k A \frac{\partial f}{\partial s}}.$$

The denominator of $\frac{\partial s^*}{\partial d}$ is negative, so the sign of $\frac{\partial s^*}{\partial d}$ depends on its numerator, and specifically depends on the relative size of $\frac{\partial f}{\partial d}$ and $\frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d}$. If assume that $\frac{\partial W_k(d)}{\partial d} \geq 0$ so that $\frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d} \geq 0$, and the negative direct effect of being left-behind is larger than the positive indirect effect through income, then $\frac{\partial s^*}{\partial d} \geq 0$, suggesting that the child will increase study time to compensate for worse performance due to the absence of parent, and vice versa.

Graphically, the original equilibrium of child study time should be at the intersection of the marginal utility of studying in period 1 and period 2, which is s^* in Figure 1. For the child, holding study time fixed, if there is a change in parent migration status d, then the child's human capital will change due to the direct effect of migration and its indirect effect through income/consumption, thereby affecting the income and consumption in the future. However, it will not affect consumption in the first period. This is equivalent to stating that holding s fixed, when d changes, h and c_2 will be affected. Recall that the marginal effect of study on current utility only depends on study time s, so even if d changes, the current marginal effect won't change because s is held fixed. The marginal utility of migration in period 2 depends on the discounted marginal effect of study time on future utility through consumption in that period ($\beta_k \frac{\partial u_2^k}{\partial c_2}\frac{\partial g}{\partial h}\frac{\partial f}{\partial s}$). When d changes, the discount rate β_k won't change, and education production through the indirect effect of study time won' change either because s is held constant. However, education production will be directly affected by migration status and indirectly affected by migration through investment in the child, and thus the returns to education will change. In addition, the change in human capital leads to change in the consumption level, which will affect the marginal utility of consumption in the second period. Considering the changes in returns to education and marginal utility of consumption, the marginal utility of study time in period 2 will be affected.

If d increases, on the one hand, the direct effect of migration will lead to lower human capital. Since the return to education decreases as human capital increases, we would expect an increase in the return to education. As human capital worsens, future income and future consumption will drop, which leads to an increase in the marginal utility of consumption since it's decreasing in consumption levels. Therefore, considering the direct effect of migration, we expect the return to education and the marginal utility of consumption to increase as d increases, and thus the marginal utility of study time in period 2 increases as d increases. Graphically, the curve for marginal utility of study time in period 2 shifts up since the new marginal effect of studying on future utility becomes higher for every level of s. If d increases, on the other hand, the indirect effect of migration through investment in child will lead to higher human capital. Since the return to education decreases as human capital increases, we would expect to see a drop in the return to education. As the child human capital becomes higher, future income and future consumption will increase, leading to a decrease in the marginal utility of consumption since it's decreasing in consumption levels. Due to the indirect effect of migration through income, as d increases, we expect the return to education and the marginal utility of consumption to decrease, and thus the marginal effect of study time on future utility decreases. Graphically, the future marginal effect curve shifts down since the new marginal effect of study on future utility becomes lower for every level of s.

In summary, when migration status increases, although the current marginal effect curve of study time remains unchanged, the shift of the future marginal effect curve will depend on the relative sizes of the two forces from the direct and indirect effect of migration. If the two effects add up to be negative, then the curve will finally shift up and the new equilibrium study time will increase to $s^{*'}$, suggesting that if d increases, s^* is expected to increase, as shown in Figure 1. This suggests that the child will have to study longer to compensate for the large detrimental effect of migration on their school performances. This is consistent to our findings in Appendix A.1.

A.2 Parental utility maximization

The utility of parent is

$$\begin{aligned} \max_{d} \quad & u_1^p(c_1^p) + \beta_p u_2^p(c_2), \\ s.t. \quad & c_1^p = \gamma_p W_p(d), \\ & c_2 = g(h), \\ & h = f(d, s, c_1^k, h_0). \end{aligned} \tag{12}$$

Since γ_p is a positive fixed number, I simply omit it from the first period consumption.

Plugging constraints to the utility function:

$$L^p = u_1^p(W_p) + \beta_p u_2^p(g(h))$$

With a slight abuse of notation, we write $u_2(h)$ for $u_2(g(h))$. Then we know that

$$\frac{\partial u_2^p}{\partial h} = \frac{\partial u_2^p}{\partial c_2}\frac{\partial g}{\partial h} \geq 0,$$
$$\frac{\partial^2 u_2^p}{\partial h^2} = \frac{\partial^2 u_2}{\partial c_2^2}(\frac{\partial g}{\partial h})^2 + \frac{\partial^2 g}{\partial h^2}\frac{\partial u_2^p}{\partial c_2} \leq 0.$$

Taking the derivative with respect to d and obtain the first order condition

$$\frac{\partial L^p}{\partial d} = \frac{\partial u_1^p}{\partial c_1^p}\frac{\partial c_1^p}{\partial d} + \beta_p \frac{\partial u_2^p}{\partial h}(\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d}) = 0.$$

For the parent, when they migrate out, they are potentially benefiting from higher consumption due to higher income in the first period, but at the cost of their child's human capital and thus their future income and consumption. From the first-order condition, we know the marginal effect of parental migration on current utility is $\mathrm{MU}_1^p = \frac{\partial u_1^p}{\partial c_1^p}\frac{\partial c_1^p}{\partial d}$, and its marginal effect on future utility is $\mathrm{MU}_2^p = -\beta_p \frac{\partial u_2^p}{\partial h}(\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d})$. To guarantee an interior solution, we need the marginal effect on future utility to be nonnegative, that is, $\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d} \leq 0$.

From the first-order condition, we can derive parent's optimal migration decision d^* as a function of child's study time s. Our goal is to study the effect of s on d^*, so further take the derivative of $\frac{\partial L}{\partial d}$ with respect to s,

$$\begin{aligned}\frac{\partial^2 L^p}{\partial s \partial d} =& \frac{\partial u_1^p}{\partial c_1^p}\frac{\partial^2 c_1^p}{\partial d^2}\frac{\partial d}{\partial s} + \frac{\partial^2 u_1^p}{\partial {c_1^p}^2}(\frac{\partial c_1^p}{\partial d})^2\frac{\partial d}{\partial s} + \\ & \beta_p(\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d})\frac{\partial^2 u_2^p}{\partial h^2}(\frac{\partial f}{\partial d}\frac{\partial d}{\partial s} + \frac{\partial f}{\partial s} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d}\frac{\partial d}{\partial s}) + \\ & \beta_p \frac{\partial u_2^p}{\partial h}[\frac{\partial W_k(d)}{\partial d}(\frac{\partial^2 f}{\partial c_1^k \partial d}\frac{\partial d}{\partial s} + \frac{\partial^2 f}{\partial c_1^k \partial s} + \frac{\partial^2 f}{\partial {c_1^k}^2}\frac{\partial W_k(d)}{\partial d}\frac{\partial d}{\partial s}) + \frac{\partial f}{\partial c_1^k}\frac{\partial^2 W_k(d)}{\partial d^2}\frac{\partial d}{\partial s}] = 0.\end{aligned}$$

Since we assume the separability of human capital production function, i.e., $\frac{\partial^2 f}{\partial s \partial d} = \frac{\partial^2 f}{\partial s \partial c_1^k} = \frac{\partial^2 f}{\partial c_1^k \partial d} = 0$, the second-order condition can be simplified, and thus

$$\frac{\partial d^*}{\partial s} = \frac{-\beta_p \frac{\partial^2 u_2^p}{\partial h^2}\frac{\partial f}{\partial s}(\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d})}{\frac{\partial u_1^p}{\partial c_1^p}\frac{\partial^2 c_1^p}{\partial d^2} + \frac{\partial^2 u_1^p}{\partial {c_1^p}^2}(\frac{\partial c_1^p}{\partial d})^2 + \beta_p \frac{\partial^2 u_2^p}{\partial h^2}(\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d})^2 + \beta_p \frac{\partial u_2^p}{\partial h}[\frac{\partial^2 f}{\partial d^2} + \frac{\partial^2 f}{\partial {c_1^k}^2}(\frac{\partial W_k(d)}{\partial d})^2 + \frac{\partial f}{\partial c_1^k}\frac{\partial^2 W_k(d)}{\partial d^2}]}.$$

Since $\frac{\partial u_1^p}{\partial c_1^p} \geq 0, \frac{\partial^2 u_1^p}{\partial {c_1^p}^2} \leq 0; \frac{\partial u_2^p}{\partial h} \geq 0, \frac{\partial^2 u_2^p}{\partial h^2} \leq 0; \frac{\partial c_1^p}{\partial d} \geq 0, \frac{\partial^2 c_1^p}{\partial d^2} \leq 0; \frac{\partial f}{\partial c_1^k} \geq 0, \frac{\partial^2 f}{\partial {c_1^k}^2} \leq 0; \frac{\partial f}{\partial d} \leq 0, \frac{\partial^2 f}{\partial d^2} \leq 0,$ and $\beta_p > 0$, the denominator of $\frac{\partial d^*}{\partial s}$ is negative. The numerator is also negative since $\frac{\partial f}{\partial s} \geq 0$ and $\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d} \leq 0$. Thus, $\frac{\partial d^*}{\partial s} \geq 0$ as long as there is an interior solution. This suggests that if the child is willing to study for longer times, parent will be more "assured" and more likely to migrate out.

Graphically, the original equilibrium of parent migration decision should be at the intersection of the marginal utility of migration in the first period and the marginal utility in the second period, which is d_0^* in Figure 2. Holding parent migration status constant, if there is a change in child's study time, then the child's human capital will be affected, thereby affecting the income and consumption in the second period. However, it will not affect the consumption in the first period. This is equivalent to stating that when holding d fixed and changing s, c_1^p will remain unchanged but h and c_2 will be affected. Recall that the marginal utility from migration in the first period is the marginal effect of migration status on current utility through consumption in that period ($\frac{\partial u_1^p}{\partial c_1^p}\frac{\partial c_1^p}{\partial d}$), so even if s changes, the marginal utility in the first period won't change because d and c_1^p remain the same. The marginal utility in the second period is the discounted marginal effect of migration status on future utility through consumption in that period ($-\beta_p \frac{\partial u_2^p}{\partial h}(\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d})$). future consumption depends only on child human capital. When child study time s changes, education production will be affected [8], the marginal utility in the second period from consumption/human capital will be affected, so the marginal utility of migration in the second period will be affected. If s increases, h will increase as the marginal effect of study time on human capital production is positive. Therefore, consumption in the second period increases as child human capital increases. Since the marginal utility is decreasing in consumption, we expect to see a decrease in marginal utility from future consumption, so the marginal utility of migration in Period 2 will decrease. Therefore, when the child increases study time, although parent's marginal utility in Period 1 will not change, the curve for marginal utility in Period 2 will shift down since it becomes lower for every level of d. This results in the new equilibrium of migration status to increase, as shown in Figure 2. That is, d^* is increasing in s. This is consistent to our findings in Appendix A.2.

[8] Education production through the direct effect of migration or the indirect effect through current consumption ($\frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k}\frac{\partial W_k(d)}{\partial d}$) won't change because c_1^k and d remain the same.

A.3 A specific functional form

There might be some concern in the above decision making process since I am assuming simultaneous decisions. In this section, I will use specific functional forms to show that the joint decision process of parent and child will lead to one unique equilibrium. In that case, it makes no difference if we are assuming a simultaneous decision process or a sequential one. In addition, the specific functional forms I choose is also consistent with my empirical model.

For the child decision process, the utility maximization satisfying the previous assumptions could be depicted by:

$$\begin{aligned} \max_{s} \quad & \log[(1-s)T_0] + \log(c_1^k) + \beta_k \log(c_2), \\ s.t. \quad & c_1^k \leq a + w_1 \cdot d, \\ & c_2 \leq w_2 \cdot e, \\ & h \leq \gamma_T \cdot s \cdot T_0 + \gamma_W (a + w_1 d) + \gamma_D \cdot d, \end{aligned}$$

where T_0 is total weekly time available to the child, and d is a measure of parent migration status. For simplicity, I assume $\gamma(D)$ in Equation (11) to be constant.

Plugging the constraints into the objective function, we have

$$L^k = \log[(1-s)T_0] + \log(a + w_1 \cdot d) + \beta_k \log[w_2(\gamma_T \cdot s \cdot T_0 + \gamma_W(a + w_1 d) + \gamma_D d)].$$

Taking its first-order derivative with respect to s, we have

$$\frac{\partial L^k}{\partial s} = -\frac{1}{1-s} + \frac{\gamma_T \cdot T_0 \cdot \beta_k}{\gamma_T \cdot s \cdot T_0 + \gamma_W(a + w_1 d) + \gamma_D d}$$

Setting the first-order condition to 0, we could solve for s^*, the optimal time decision of children:

$$s^* = \frac{\gamma_T \cdot T_0 \cdot \beta_k + a \cdot \gamma_W + (\gamma_D + w_1 \cdot \gamma_W) \cdot d}{\gamma_T T_0 (\beta_k - 1)} \tag{13}$$

Since $\gamma_D + \gamma_W w_1 = \frac{\partial f}{\partial d} + \frac{\partial f}{\partial c_1^k} \frac{\partial W_k(d)}{\partial d} \leq 0$, and $\gamma_T T_0(\beta_k - 1) < 0$ due to the fact that discount factor $0 \leq \beta_k < 1$, we know that s^* is non-decreasing as d increases, which is consistent to our findings in Appendix A.1.

For the parent decision process, the utility maximization process is depicted by:

$$\begin{aligned} \max_{d} \quad & \log(c_1^p) + \beta_p \log(c_2), \\ s.t. \quad & c_1^p \leq a + w_1 \cdot d, \\ & c_2 \leq w_2 \cdot e, \\ & h \leq \gamma_T \cdot s \cdot T_0 + \gamma_W (a + w_1 d) + \gamma_D \cdot d. \end{aligned}$$

Plugging in the constraints to the objective function,

$$L^p = \log(a + w_1 \cdot d) + \beta_p \log[w_2(\gamma_T \cdot s \cdot T_0 + \gamma_W(a + w_1 d) + \gamma_D \cdot d)].$$

Taking the first-order derivative with respect to D, we have

$$\frac{\partial L^p}{\partial d} = \frac{w_1}{a + w_1 d} - \frac{\gamma_W w_1 + \gamma_D \beta_p}{\gamma_T \cdot s \cdot T_0 + \gamma_W(a + w_1 d) + \gamma_D \cdot d}.$$

Setting the first-order condition to 0, we have

$$d^* = \frac{-a\gamma_D\beta_p + w_1\gamma_T T_0 s}{w_1\gamma_D(\beta_p - 1)}. \tag{14}$$

Since $w_1\gamma_T T_0 \geq 0$, and $w_1\gamma_D(\beta_p - 1) > 0$ because $\gamma_D < 0$ and $0 \leq \beta_p < 1$, we know d^* is non-decreasing as s increases, which is consistent to our finding in Appendix A.2.

Next I will show there is a unique equilibrium (s^{**}, d^{**}). If we draw the reaction function of the parent and the child on one graph with d on the horizontal axis and s on the vertical axis, this is equivalent to show that the reactions curves have different slopes. The slope of child's reaction function is $\frac{\gamma_D + w_1\gamma_W}{\gamma_T T_0(\beta_k - 1)}$, and the slope of parent's reaction function is $\frac{w_1\gamma_D(\beta_p - 1)}{w_1\gamma_T T_0}$:

$$\begin{aligned}
&\frac{\gamma_D + w_1\gamma_W}{\gamma_T T_0(\beta_k - 1)} - \frac{w_1\gamma_D(\beta_p - 1)}{w_1\gamma_T T_0} \\
=&\frac{w_1(\gamma_D + w_1\gamma_W) - w_1\gamma_D(\beta_p - 1)(\beta_k - 1)}{w_1\gamma_T T_0(\beta_k - 1)} \\
=&\frac{w_1^2\gamma_W - w_1\gamma_D(\beta_p\beta_k - \beta_p - \beta_k)}{w_1\gamma_T T_0(\beta_k - 1)} \\
=&\frac{w_1[w_1\gamma_W - \gamma_D(\beta_p\beta_k - \beta_p - \beta_k)]}{w_1\gamma_T T_0(\beta_k - 1)}
\end{aligned}$$

We already know the denominator of the difference is negative since $\beta_k - 1 < 0$, so the value of the difference only depends on the numerator. As long as we have $w_1\gamma_W \neq \gamma_D(\beta_p\beta_k - \beta_p - \beta_k)$, the slopes will be different. Since $-1 < \beta_p\beta_k - \beta_p - \beta_k \leq 0$, if the negative direct effect γ_D is very large, then the numerator of the difference would be negative so the difference would be positive, suggesting that the slope of child's reaction function would be steeper. This also makes intuitive sense because if γ_D is very large, then based on the graph of marginal utility, to compensate for the negative direct effect, the child tends to increase study time by a lot, and the reaction is stronger than parent's. The graph for this case is depicted in Figure 3. The equilibrium study time s^{**} and equilibrium migration decision d^{**} is unique. Since the observed data are in equilibrium, we have s^{**} is given by (13) with $d = d^{**}$ and d^{**} is the solution of (13) and (14). This provides a concrete example for the abstract system (4).

B Complementary empirical results

In this appendix, I will present supplementary results by imposing stronger yet potentially invalid assumptions like exogeneity of treatments/mediators or missing-at-random mediators. These results should be viewed as robustness checks or even sanity checks, which highlight the issues of failure to handle endogeneity and non-random missing values carefully.

B.1 Results without accounting for non-random missing values (no imputation)

When the mediators are missing at random, there is no non-random missing issue and thus one can estimate the model on units without missing values. Table B.1 - B.3 present the results under the same setting as Table 5.1 - 5.5, except that Heckman model is not applied to impute the missing study time and investment in child.

We can observe that for both all sample and subgroup analysis, the direct effect shrinks slightly and the indirect effect through investment shrinks drastically, despite that the sample size only drops by 35%. This corroborates my speculation that simply removing these observations tend to underestimate the impact of migration, because those whose parents or guardians fail to report their study time or investment in them are likely suffering more from parental migration.

Table B.1: Effect of Parental Migration on Child Schooling Outcomes (IV, All Sample, Not Imputed)

			(1) Language	(2) Math
Direct Effect				
Parental Accompany			-0.458***	-0.411**
			(0.001)	(0.003)
Indirect Effect				
Study time			-0.008	-0.006
			(0.096)	(0.203)
Investment in children			-0.402**	-0.437***
			(0.003)	(0.001)
Sepecification Tests	*(1) Study time*	*(2) Investment*	*(3) Language*	*(4) Math*
Underidentification test				
(Anderson canon. corr. LM statistic)	34.769***	34.769***	18.155***	18.155***
	(0.000)	(0.000)	(0.000)	(0.000)
Overidentification test (Sargan statistic)	5.183*	3.629	1.561	1.487
	(0.023)	(0.057)	(0.458)	(0.475)
Endogeneity test	0.281	5.175*	19.310***	30.792***
	(0.596)	(0.023)	(0.000)	(0.000)
Obs.		1277		

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.2: Effect of Parental Migration on Child Schooling Outcomes (IV, Subgroup by Gender, Not Imputed)

	Girl		Boy	
	(1) Language	(2) Math	(3) Language	(4) Math
Direct Effect				
Parental Accompany	-0.350*	-0.371	-0.340**	-0.236*
	(0.035)	(0.057)	(0.009)	(0.029)
Indirect Effect				
Study time	-0.000	-0.001	-0.026	-0.007
	(0.762)	(0.721)	(0.056)	(0.586)
Investment in children	-0.600*	-0.780*	-0.121*	-0.129***
	(0.043)	(0.030)	(0.011)	(0.001)
Sepecification Tests	*(1) Study time*	*(2) Investment*	*(3) Language*	*(4) Math*
Girl				
Underidentification test				
(Anderson canon. corr. LM statistic)	5.242	5.242	7.516	7.516
	(0.073)	(0.073)	(0.057)	(0.057)
Overidentification test (Sargan statistic)	2.518	5.040*	0.959	1.256
	(0.113)	(0.025)	(0.619)	(0.534)
Endogeneity test	0.021	3.314	3.684	10.138*
	(0.885)	(0.069)	(0.298)	(0.017)
Obs.		571		
Boy				
Underidentification test				
(Anderson canon. corr. LM statistic)	36.416***	36.416***	9.968*	9.968*
	(0.000)	(0.000)	(0.019)	(0.019)
Overidentification test (Sargan statistic)	3.559	0.531	1.081	0.781
	(0.059)	(0.466)	(0.582)	(0.677)
Endogeneity test	0.069	0.871	19.510***	21.147***
	(0.793)	(0.351)	(0.000)	(0.000)
Obs.		706		

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.3: Effect of Parental Migration on Child Schooling Outcomes (IV, Subgroup by Birth Order, Not Imputed)

	First child		Subsequent children	
	(1) Language	(2) Math	(3) Language	(4) Math
Direct Effect				
Parental Accompany	-0.552	-0.647*	-0.469*	-0.282
	(0.075)	(0.040)	(0.014)	(0.091)
Indirect Effect				
Study time	-0.038	-0.046	0.001	-0.001
	(0.303)	(0.303)	(0.766)	(0.766)
Investment in children	-0.168	-0.260	-0.624*	-0.496*
	(0.227)	(0.114)	(0.025)	(0.028)
Sepecification Tests	*(1) Study time*	*(2) Investment*	*(3) Language*	*(4) Math*
First child				
Underidentification test				
(Anderson canon. corr. LM statistic)	14.970***	14.970***	2.693	2.693
	(0.001)	(0.001)	(0.260)	(0.260)
Overidentification test (Sargan statistic)	0.469	3.359	0.001	0.291
	(0.493)	(0.067)	(0.980)	(0.590)
Endogeneity test	0.636	0.119	6.198	14.192**
	(0.425)	(0.730)	(0.102)	(0.003)
Obs.		590		
Subsequent children				
Underidentification test				
(Anderson canon. corr. LM statistic)	7.995*	7.995*	4.635	4.635
	(0.018)	(0.018)	(0.099)	(0.099)
Overidentification test (Sargan statistic)	3.313	0.160	0.974	1.797
	(0.069)	(0.690)	(0.324)	(0.180)
Endogeneity test	1.288	9.416**	9.994*	6.080
	(0.256)	(0.002)	(0.019)	(0.108)
Obs.		535		

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B.2 Results without accounting for endogeneity (no IV)

Recall the structural equation model (5) - (8) in Section 4. If D, T, W were all exogenous, $\epsilon_{Pi}, \epsilon_{Ti}, \epsilon_{Wi}, \epsilon_{Di}$ would be mutually independent, in which case no instrumental variable is needed for identification because the coefficients can be identified separately from each single equation. The results are presented in Table B.4 - 5.5.

Nevertheless, the endogeneity tests in Table 5.1 - 5.5 show strong evidence against exogeneity. Therefore, the results in Table B.4 - 5.5 should be taken as a robustness check. We can observe that both the direct effect and indirect effect through investment have the same signs as the results in Section 5, but they are underestimated drastically for both all samples and subgroups. This highlights the importance to handle the endogeneity in this problem.

Table B.4: Effect of Parental Migration on Child Schooling Outcomes (No IV, All Sample, Imputed)

	Language Score	Math Score
Direct Effect		
Parental Accompany	-0.074**	-0.015
	(0.007)	(0.610)
Indirect Effect		
Study time	-0.002	-0.005**
	(0.159)	(0.008)
Investment in children	-0.004*	-0.005*
	(0.035)	(0.035)
Obs.	1971	

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.5: Effect of Parental Migration on Child Schooling Outcomes (No IV, Subgroup by Gender, Imputed)

	Girl		Boy	
	(1) Language	(2) Math	(3) Language	(4) Math
Direct Effect				
Parental Accompany	-0.126***	-0.077*	-0.018	0.050
	(0.001)	(0.045)	(0.650)	(0.253)
Indirect Effect				
Study time	-0.003	-0.003	-0.002	-0.010**
	(0.370)	(0.403)	(0.428)	(0.003)
Investment in children	-0.002	-0.002	-0.007*	-0.012**
	(0.387)	(0.387)	(0.026)	(0.009)
Obs.	887		1084	

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.6: Effect of Parental Migration on Child Schooling Outcomes (No IV, Subgroup by Birth Order, Imputed)

	First child		Subsequent children	
	(1) Language	(2) Math	(3) Language	(4) Math
Direct Effect				
Parental Accompany	-0.132***	-0.067	-0.042	0.037
	(0.000)	(0.094)	(0.340)	(0.437)
Indirect Effect				
Study time	0.001	-0.002	-0.007*	-0.013***
	(0.743)	(0.464)	(0.016)	(0.000)
Investment in children	-0.001	-0.007*	-0.005	-0.005
	(0.713)	(0.026)	(0.217)	(0.217)
Obs.	891		860	

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B.3 Results without accounting for both

Finally I present the results without accounting for either endogeneity or non-random missing values in Table B.7 - B.9 as an additional sanity check. This analysis even fails to capture the significant negative effect through investment.

Table B.7: Effect of Parental Migration on Child Schooling Outcomes (No IV, All Sample, Not Imputed)

	Language Score	Math Score
Direct Effect		
Parental Accompany	-0.074*	-0.013
	(0.016)	(0.688)
Indirect Effect		
Study time	-0.001	-0.002
	(0.392)	(0.392)
Investment in children	-0.002	-0.004
	(0.301)	(0.301)
Obs.	1277	

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.8: Effect of Parental Migration on Child Schooling Outcomes (No IV, Subgroup by Gender, Not Imputed)

	Girl		Boy	
	(1) Language	(2) Math	(3) Language	(4) Math
Direct Effect				
Parental Accompany	-0.110*	-0.052	-0.045	0.021
	(0.013)	(0.254)	(0.281)	(0.659)
Indirect Effect				
Study time	-0.001	-0.001	-0.000	-0.003
	(0.501)	(0.501)	(0.700)	(0.391)
Investment in children	-0.001	-0.001	-0.005	-0.009
	(0.806)	(0.806)	(0.169)	(0.169)
Obs.	571		706	

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.9: Effect of Parental Migration on Child Schooling Outcomes (No IV, Subgroup by Birth Order, Not Imputed)

	First child		Subsequent children	
	(1) Language	(2) Math	(3) Language	(4) Math
Direct Effect				
Parental Accompany	-0.084*	-0.024	-0.107*	-0.018
	(0.036)	(0.601)	(0.035)	(0.730)
Indirect Effect				
Study time	0.001	-0.001	0.005	0.008
	(0.751)	(0.716)	(0.174)	(0.174)
Investment in children	-0.000	-0.007	0.010	0.011
	(0.899)	(0.101)	(0.200)	(0.200)
Obs.	590		535	

p-values in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

Adams Jr, R. H. and Page, J. (2005). Do international migration and remittances reduce poverty in developing countries? *World development*, 33(10):1645–1669.

Afsar, R. (2003). Internal migration and the development nexus: the case of bangladesh. In *Regional conference on migration, development and pro-poor policy choices in Asia*, pages 22–24.

Akgüç, M., Giulietti, C., and Zimmermann, K. F. (2014). The rumic longitudinal survey: Fostering research on labor markets in china. *IZA Journal of Labor & Development*, 3(1):5.

Alcaraz, C., Chiquiar, D., and Salcedo, A. (2012). Remittances, schooling, and child labor in mexico. *Journal of Development Economics*, 97(1):156–165.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, 22(3):327–351.

Antman, F. M. (2011). 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. (2013). 16 the impact of migration on family left behind. *International handbook on the economics of migration*, page 293.

Arguillas, M. J. B. and Williams, L. (2010). The impact of parents' overseas employment on educational outcomes of filipino children. *International Migration Review*, 44(2):300–319.

Arnold, F. and Shah, N. M. (1984). Asian labor migration to the middle east. *International Migration Review*, 18(2):294–318.

Baron, R. M. and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173.

Bartik, T. J. et al. (1991). Who benefits from state and local economic development policies? *Books from Upjohn Press*.

Battistella, G. and Conaco, M. C. G. (1998). The impact of labour migration on the children left behind: a study of elementary school children in the philippines. *SOJOURN: Journal of Social Issues in Southeast Asia*, pages 220–241.

Baum, C. F., Schaffer, M. E., and Stillman, S. (2007). Enhanced routines for instrumental variables/generalized method of moments estimation and testing. *The Stata Journal*, 7(4):465–506.

Booth, M. Z. (1995). Children of migrant fathers: The effects of father absence on swazi children's preparedness for school. *Comparative Education Review*, 39(2):195–210.

Bryant, J. (2005). Children of international migrants in indonesia, thailand, and the philippines: A review of evidence and policies.

Celli, V. (2019). Causal mediation analysis in economics: objectives, assumptions, models. *Working Papers from Sapienza University of Rome, DISS*.

Chang, H., Dong, X.-y., and MacPhail, F. (2011). Labor migration and time use patterns of the left-behind children and elderly in rural china. *World Development*, 39(12):2199–2210.

Chen, J. J. (2013). Identifying non-cooperative behavior among spouses: child outcomes in migrant-sending households. *Journal of Development Economics*, 100(1):1–18.

Démurger, S. (2015). Migration and families left behind. *IZA World of Labor*.

Edwards, A. C. and Ureta, M. (2003). International migration, remittances, and schooling: evidence from el salvador. *Journal of development economics*, 72(2):429–461.

Fisher, M. (2005). On the empirical finding of a higher risk of poverty in rural areas: Is rural residence endogenous to poverty? *Journal of Agricultural and Resource Economics*, pages 185–199.

Fritz, M. S. and MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological science*, 18(3):233–239.

Ganepola, V. (2002). The psychosocial wellbeing of families left behind by asylum migration–the sri lankan experience. In *WIDER Conference on Poverty, International Migration and Asylum*, pages 27–28.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624.

Gong, X., Kong, S. T., Li, S., and Meng, X. (2008). Rural-urban migrants: A driving force for growth. *China's Dilemma: Economic Growth, the Environment and Climate Change*, pages 110–152.

Graham, E. and Jordan, L. P. (2011). Migrant parents and the psychological well-being of left-behind children in southeast asia. *Journal of Marriage and Family*, 73(4):763–787.

Gupta, S., Pattillo, C. A., and Wagh, S. (2009). Effect of remittances on poverty and financial development in sub-saharan africa. *World development*, 37(1):104–115.

Hanson, G. H. and Woodruff, C. (2003). Emigration and educational attainment in mexico. Technical report, Mimeo., University of California at San Diego.

Hayes, A. F. and Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological science*, 24(10):1918–1927.

He, B., Fan, J., Liu, N., Li, H., Wang, Y., Williams, J., and Wong, K. (2012). Depression risk of ?left-behind children?in rural china. *Psychiatry research*, 200(2-3):306–312.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.

Heckman, J. J. and Pinto, R. (2015). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric reviews*, 34(1-2):6–31.

Huber, M. (2016). Disentangling policy effects into causal channels. *IZA World of Labor*.

Huber, M. (2019). A review of causal mediation analysis for assessing direct and indirect treatment effects. Technical report, Université de Fribourg.

Institute of Labor Economics (IZA), Australian National University, Pthe University of Queensland, and the Beijing Normal University (2014). Longitudinal survey on rural urban migration in china (rumic) (2008-2009). *International Data Service Center of IZA (IDSC). Version 1.0*.

Kandel, W. and Kao, G. (2000). Shifting orientations: How us labor migration affects children's aspirations in mexican migrant communities. *Social science quarterly*, pages 16–32.

Klugman, J. (2009). Human development report 2009. overcoming barriers: Human mobility and development. *Overcoming Barriers: Human Mobility and Development (October 5, 2009). UNDP-HDRO Human Development Reports*.

Knodel, J. E. and Saengtienchai, C. (2002). *AIDS and older persons: The view from Thailand*. Population Studies Center, University of Michigan.

Kong, S. T. (2010). Rural–urban migration in china: Survey design and implementation. *Chapters*.

Maruja, M. and Baggio, A. F. (2003). The other face of migration: children and families left behind. *Workshop "Taking the lead: successful partnership initiatives for the delivery of settlement services"*.

McKenzie, D. and Rapoport, H. (2011). Can migration reduce educational attainment? evidence from mexico. *Journal of Population Economics*, 24(4):1331–1358.

Mendoza, D. (2004). Migrant workers? children left behind, left out. *IPS news*.

Meng, X., Kong, S. T., and Zhang, D. (2010). How much do we know about the impact of the economic downturn on the employment of migrants? Technical report, ADBI Working Paper.

Meng, X. and Yamauchi, C. (2015). Children of migrants: The impact of parental migration on their children's education and health outcomes. *IZA Discussion Paper No. 9165*.

Nguyen, L., Yeoh, B. S., and Toyota, M. (2006). Migration and the well-being of the ?left behind?in asia: Key themes and trends. *Asian Population Studies*, 2(1):37–44.

Niimi, Y., Pham, T. H., and Reilly, B. (2009). Determinants of remittances: Recent evidence using data on internal migrants in vietnam. *Asian Economic Journal*, 23(1):19–39.

Olwig, K. F. (1999). Narratives of the children left behind: Home and identity in globalised caribbean families. *Journal of ethnic and migration studies*, 25(2):267–284.

Sun, X., Tian, Y., Zhang, Y., Xie, X., Heath, M. A., and Zhou, Z. (2015). Psychological development and educational problems of left-behind children in rural china. *School Psychology International*, 36(3):227–252.

Waddington, C. (2003). Livelihood outcomes of migration for poor people. In *Development Research Centre on Migration Globalisation and Poverty*. Citeseer.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Xiang, A., Jiang, D., and Zhong, Z. (2016). The impact of rural–urban migration on the health of the left-behind parents. *China Economic Review*, 37:126–139.

Zhao, Q., Yu, X., Wang, X., and Glauben, T. (2014). The impact of parental migration on children's school performance in rural china. *China Economic Review*, 31:43–54.