Ex Ante Evaluation of the Effects of Migration Policy on Children's Cognitive Achievements

Bolun Li *†

Abstract

China's large-scale rural-to-urban migration has affected more than 60 million children who have been left behind by their migrant parents. This paper proposes several migration policies that are aimed at improving these children's cognitive achievements and quantifies the impacts of these policies prior to their implementation. I develop a model of household migration that embeds a child's cognitive skill formation process. By exploiting income variation using data from the China Family Panel Studies, I directly evaluate the policy impacts via a nonparametric matching estimator. Counterfactual experiments suggest that a non-migration subsidy is most effective in improving children's cognitive achievements when it targets low-income families and younger children. When associated with middle school graduation, the subsidy induced test score change lifts the probability of graduation by 8.6% for children from low-income households.

*Department of Economics, Rice University. Email: bolun.li@rice.edu

†Many thanks to Ken Wolpin, Flávio Cunha, Nick Frazier, Andrea Salvati, Ajinkya Keskar, Johnny Chau, Shuaizhang Feng, Yuanyuan Chen, John Kennan, Chao Fu, Rossella Calvi, and Chris Taber for helpful comments and suggestions. All remaining errors are my own.

1 Introduction

The rapid economic growth in China has led to roughly 160 million[1] rural residents moving to urban areas in search of higher incomes. This large-scale rural-to-urban migration has affected over 60 million children aged 0-18 who have been left behind by their migrant parents due to institutional barriers, such as China's household registration system and the capacity constraints of local schools in urban areas. A growing consensus in the literature emphasizes the negative association between parental migration and the cognitive development of rural-origin children.[2] However, little is known about how to design migration policies that are aimed at improving the cognitive outcomes of these children.

The main ambition of this paper is to analyze the impacts of hypothetical migration policies on children's cognitive development prior to implementation. In the absence of data on the treatment group, an economic model is often required to evaluate program impacts ex ante. Specifically, I formalize a model in which a child's cognitive skill formation process is nested within household migration decision making. Parents choose whether to move or stay. A child is left behind at the home location if parents choose to migrate. A household receives stochastic income based parental migration decision. Child cognitive outcome is determined by the parental migration decision and the associated household income level. Parents make optimal migration decision to maximize household utility subject to their budget constraints and the production technology of a child's cognitive achievement. I use this model to predict the effects of hypothetical migration policies on the cognitive achievements of rural-origin children who remain in rural areas.

The standard econometric approach for evaluating policy impact ex ante usually involves two stages. A structural model is estimated first; then, policy-invariant model primitives are used to simulate counterfactual experiments. However, this paper sidesteps the process of estimating the full structural model and directly evaluates counterfactual policy impacts by exploiting the specific variation in the data that provides an analogue of policy variation.

Two policy experiments are considered in this paper. The first policy is restricting parents from

[1] China National Bureau of Statistics 2012

[2] See Section 2

migrating for work. This policy is aimed at improving a child's cognitive achievement through increased time investments under the assumption that parents are more likely to spend time with their children when they are forced to stay in rural areas. Although restricting migration seems out of equilibrium, it is in line with a recent policy recommendation from the State Council of China as well as general attitudes of metropolitan government. The policy, State Council Opinions on Efforts to Strengthen Care and Protection of Rural Left-Behind Children (Ministry of Human Resources and Social Security of the People's Republic of China, 2016), states that parents are required to return home when there is no legal guardian or when a legal guardian is incapable of taking care of the children. In addition, migrant schools, which are constructed for children of migrant workers in urban areas, have been shut down by local government in recent years in major cities like Beijing and Shanghai as a way to force migrant workers to move back (Chen & Feng, 2013; Duncan & Mao, 2011; Edes & Inskeep, 2017; Hernández & Zhao, 2017). Due to restrictions on the choice set, the economic model upholds a nonparametric matching estimation strategy absent any data on treated individuals.

The second policy is providing a subsidy to a household if the parents do not migrate for work. The subsidy is aimed at improving a child's cognitive achievement through increased time investments as well as material inputs since parents are incentivized to stay with a higher level of household non-migrant income. This experiment also coincides with the previous policy recommendation on left-behind children (Ministry of Human Resources and Social Security of the People's Republic of China, 2016), which proposes tax cut if migrant workers move back and start business at rural home location. Because the subsidy operates only through the budget constraint, the impact of non-migration subsidy can also be estimated directly using a nonparametric matching technique. However, the variation in household income is required to estimate the effect of the subsidy on migration behavior and children's cognitive outcomes.[3]

However, partial observability of income poses a challenge to the implementation of the match-

[3]Another counterfactual that could be answered by the current framework, see Section 4, is to subsidize parents if they migrate with their children. I do not consider this policy experiment because China's household registration system and the capacity constraints of urban schools prevent children with rural household registration from attending local schools in urban areas.

ing estimator. In the absence of its non-migrant income, a migrant household cannot be matched to a non-migrant household nor to another migrant household with an appropriate level of income. The same argument applies to a non-migrant household. Therefore, the counterfactual non-migrant income for a migrant household if this household had not moved must be imputed, and vice versa for a non-migrant household. I impute the full set of incomes under different sets of assumptions.

I estimate the policy impacts using data from the *China Family Panel Studies* (CFPS). The CFPS provides, in four waves (2010, 2011, 2012, 2014), information on family structure and migration patterns that is key to identifying rural-origin children. Moreover, the CFPS provides information on children's cognitive achievements.

The first counterfactual experiment suggests that migration restriction improves a child's word and math test score by 0.7% and 2.2% standard deviation, respectively, which is associated with a 2.8% increase in middle school graduation probability. The second counterfactual experiment shows that a non-migration subsidy reduces the migration rate and improves children's cognitive achievements. The higher the level of the subsidy is, the less likely it is that a household migrates and the higher the test scores a child achieves. Providing 14000 RMB (2000 USD) per year to a household reduces the migration rate by 10% and increases a child's word and math test score by 5.5% and 3.5% standard deviation, which translates to a 6.5% increase in graduation rate. When the same level of subsidy targets low-income families, it lifts the probability of graduation substantially by 8.6% for children from low-income households.

The remainder of this paper is organized as follows. In Section 2, I discuss the related literature. In Section 3, I describe the data used for the policy estimation and provide descriptive statistics. In Section 4, I introduce the economic model and discuss conceptually how to evaluate the proposed migration policies. Section 5 presents the estimation strategy and policy analysis, leading to conclusions and directions for future work in Section 6.

2 Related Literature

A substantial body of research in economics and migration studies documents the impact of parental migration on children's well-being, focusing on dimensions of schooling attainment, time allocation, cognitive achievement, and health outcomes. Although a few studies find that paternal migration is associated with an improvement in children's outcomes (Bai et al, 2017; Chen et al, 2009; Lee & Park, 2010), and others find no significant relationship between these two factors (Chen, 2013; Xu & Xie, 2015), a growing consensus among current research agrees on the negative impact of exposure to parental migration on children's well-being (Antman, 2011; McKenzie & Rapoport, 2011; Meng & Yamauchi, 2017; Wen & Lin, 2012; Zhang, Behrman, Fan, Wei, & Zhang, 2014; Zhao, Yu, Wang, & Glauben, 2014; Zhou, Murphy, & Tao, 2014). Many studies conclude that actions must be taken to improve children's outcomes. This paper contributes to the literature by proposing several policy recommendations and quantifying their impacts on the migration rate and children's cognitive outcomes prior to implementation.

Ex ante evaluation of social programs is important because it informs policy makers of the range of effects to expect after programs are implemented. Typically, ex ante policy evaluation first requires to estimate a structural model; then, the estimated model parameters are used to simulate counterfactual outcomes. However, in some cases, policy impacts can be directly estimated without fully estimating the structural model. Following the seminal work of Marschak (1953) and recent papers by Heckman (2000, 2001), Ichimura & Taber (2000, 2002) use a semiparametric reduced-form model to estimate the effects of tuition subsidies on education and labor market outcomes. Although the full model is based on Keane & Wolpin (2001), Ichimura & Taber (2000, 2002) show that the reduced-form estimation is sufficient to recover relevant policy impacts by exploiting tuition variation under the conditional independence assumption with exclusion restrictions. Building on previous literature, Todd & Wolpin (2008) and Wolpin (2013) discuss methods and feasibilities for the ex ante evaluation of social programs. They implement a matching estimator to evaluate the impact of subsidies on school attendance using data from PROGRESSA and show that the predictions are fairly close to the actual impact measured in a randomized experiment. I extend their framework by embedding a child's skill formation function inside the household migration decision, which allows me to investigate the impact on the migration decision as well as children's cognitive outcomes.

3 Data

To answer the policy questions outlined in Section 1, I use data from the *China Family Panel Studies* (CFPS).[4] The CFPS is a nationally representative longitudinal survey that collects important information on a wide range of topics, including family relationships, migration patterns, cognitive measures and child development, which are essential for the purpose of this study.

3.1 Sample Formation

Since many policy impactions concern children with rural origins, including non-migrant children and left-behind children,[5] sample restrictions are made to look more closely at this subpopulation. I define left-behind children as rural-origin children who are left behind by their migrant parents and define non-migrant children as those from families with non-migrant parents. After pooling 4 waves (2010, 2011, 2012, 2014) of the CFPS and dropping observations with missing data, a total of 7800 children remain. The sample consists of 5856 non-migrant children and 1944 left-behind children.

Table 1 reports the summary statistics of the variables used in the estimation. Parents of left-behind children are younger and slightly more educated. For those parents who have migrated to urban areas, the average migration spell is roughly 11 months, which is likely to influence of the cognitive development of left-behind children nontrivially. In addition, left-behind children live closer to bus stations and are more likely to have grandparents living within the household, which provides additional incentives for parents to migrate.

Household income for the group of left-behind children is constructed by combining earnings generated in their rural home location and remittance sent back by their migrant parents. Unsurprisingly, given the common understanding that parents migrate to urban destinations in search of higher earnings, household income for the left-behind group (43378 RMB)[6] is much higher than

[4]See Xu, Zhang, Tu & Ren (2008).

[5]Children with rural origins who migrate with their parents to a city are beyond the scope of this project. Liu (2015) studies the migration behavior of households with migrant children, but child outcomes are not analyzed.

[6]6.7 RMB = 1 USD in 2010

that for the non-migrant group (32565 RMB). As discussed in Wolpin (2013), it is crucial to have the variation in the data that is analogous to policy variation when implementing ex ante policy evaluation. In the case of the non-migration subsidy, variation in income corresponds to the policy under consideration. In Figure 1, I plot the histogram of income for migrant and non-migrant households, which shows a large amount of policy-relevant variability in the data.

3.2 Cognitive Measure

The CFPS systematically collects cognitive measures of surveyed individuals aged 10 years older through cognitive literacy and mathematics tests. The literacy test consists of 34 Chinese characters drawn from primary and secondary school textbooks. Respondents are assigned to different entry points, based on their highest level of education. They are asked to recognize characters one by one in ascending order of difficulty until they fail to recognize three consecutive characters. The final test score is rank order of the last character that a surveyed individual correctly recognizes. The mathematical test consists of 24 mathematical problems and is conducted in a similar fashion.

Table 2 provides descriptive statistics of children's cognitive achievements measured in terms of vocabulary and mathematics tests for children aged 10-15 years. Naively comparing the average of test scores between the two groups at different age points, I find that left-behind children perform worse than non-migrant children at most ages. Table 3 reveals additional evidence that left-behind children perform worse in cognitive tests. Naive estimator compares average test score of two groups of children for all ages. The negative sign indicates that the average test scores of left-behind children is less than the one from non-migrant children. However, the difference in mean test scores is not significant. To control for the observed characteristics that are associated with children's cognitive achievements, I apply a propensity score matching method to compare the difference in test scores between the two groups of children. Results show that left-behind children perform significantly worse in both tests once socioeconomic and demographic characteristics are controlled for.[7] By no means the descriptive statistics above indicate casual effect of parental

[7]Using CFPS, Xu & Xie (2015) apply propensity score matching to investigate the effect of parental migration on child development. The difference in findings between their paper and this one is due to data availability and controlling covariates. They use the initial wave from 2010 but

migration on children's cognitive achievements. These suggestive findings, however, confirm the commonality and normality of the data. They also coincide with the empirical evidence found in the literature, which in turn motivates researchers to take further steps and contemplate how to design migration policies to improve the cognitive achievements of left-behind children.

3.3 Anchoring Test Scores

Because cognitive test scores themselves do not provide meaningful interpretation in terms of evaluating social welfare, in the language of Cunha & Heckman (2008), I "anchor" cognitive test scores using middle school graduation of a child, which has a well-defined cardinal metric. Specifically, I consider a Logit model of graduation probability for each test score, that is,

$$G = \mathbb{1}\{\gamma_0 + \gamma_1 Q_{\text{word}} + \gamma_2 Q_{\text{math}} + \omega > 0\} \tag{1}$$

where G equals to 1 if a child graduates from middle school and Q denotes cognitive measure for word and math test score. I estimate Equation (1) using the full sample of CFPS rural origin children with available data on scores and middle school completion. Table 4 provides estimates of score anchoring equation. Both scores are positively and statistically significantly correlated with graduation probability as one would expect. I use these estimates to discuss policy impacts and social welfare in Section 5.

4 Economic Model & Evaluation Method

In this section, I first present a behavioral model of household migration that embeds a production technology of a child's cognitive skill formation process. I then discuss conceptually how to evaluate the hypothetical migration policies proposed above.

I use all 4 waves available.

4.1 The Model

Consider a static model in which parents choose whether to migrate or not. The choice set is defined as $K \equiv \{\ k \in \{0,1\} \mid k = 1$ if at least one parent migrates and $k = 0$ if both parents choose not to migrate $\}$. A household earns migrant income Y_1 if at least one parent migrates and non-migrant Y_0 otherwise. Household utility depends on consumption C, the migration decision k, a child's cognitive achievement Q, observed utility shifter X and preference shock ε. A child's cognitive achievement Q is a function of the migration decision k and associated income Y, additional observed production covariate Z and unobserved component ν. Follow the standard specification in the children's skill formation literature (Cunha & Heckman, 2008; Cunha, Heckman & Schennach, 2010; Todd & Wolpin, 2003, 2007), the production technology considered here encompasses parental investments, namely time and material inputs, in the following way. Parental migration reduces time interaction with a child and is used as a proxy for time inputs since the time inputs are not directly observed. Material inputs, such as purchasing books and toys, depend on household income. I use total household income as a proxy for good inputs and implicitly assume households spend a fixed proportion of income on goods and services for a child. This specification offers a way of thinking the trade-off between decreased time inputs and increased money inputs due to parental migration, similar to the trade-off seen in female labor supply and child development literature (Agostinelli & Sorrenti, 2018; Bernal, 2008; Bernal & Keane, 2011).

Given a budget constraint and a production technology of child's cognitive development, a household solves the following problem:

$$\begin{aligned} \max_{k \in K} \quad & U(C, Q_k, k, X, \varepsilon) \\ \text{s.t.} \quad & C = \mathbb{1}\{k=0\}Y_0 + \mathbb{1}\{k=1\}Y_1 \\ & Q_k = q_k(Y_k, Z, \nu_k). \end{aligned}$$

The household then chooses the migration alternative k that yields the highest utility:

$$U_k \geq U_j, \quad \forall k, j \in K,$$

where $U_k \equiv U(Y_k, q_k(Y_k, Z, \nu_k), k, X, \varepsilon)$. Therefore, the optimal migration decision k^* and corresponding cognitive outcome Q_k^* can be derived as functions of income Y, observables X and Z and unobservables ε and ν, namely,[8]

$$
\begin{aligned}
k^* &= \kappa(Y_0, Y_1, X, Z, \varepsilon, \nu), \\
Q_k^* &= \begin{cases} q_0(Y_0, Z, \nu_0), & \text{if } k^* = 0 \\ q_1(Y_1, Z, \nu_1), & \text{if } k^* = 1. \end{cases}
\end{aligned}
$$

4.2 Evaluation of Migration Restriction

The first policy experiment is to restrict parents from migrating. Since no one is allowed to migrate under this regime, the migration choice and the corresponding cognitive outcome of a child can be stated as:

$$
\begin{aligned}
k^{'} &= 0 \\
Q_k^{'} &= q_0(Y_0, Z, \nu_0).
\end{aligned}
$$

The impact of this restriction is defined as the difference between the expected cognitive outcome when migration is prohibited and the one when migration is permitted, that is:

$$\triangle Q^{res}(Y_0, Y_1, Z) \equiv E[Q_0^{'} \mid Y_0, Z] - E[Q_k^* \mid Y_k, Z]. \tag{2}$$

To recover the policy effect, I assume that when the parents are prohibited from migrating, they make the same investment decisions as the parents who would have stayed regardless, conditional on observed household characteristics. This paper does not explicitly model inputs into production technology because the nonparametric estimation procedure breaks down once heterogeneity in investment decisions are taken into consideration. Therefore, the impact of migration restriction is measured at the extensive margin but not the intensive margin.

The expression in Equation (2) resembles what is typically seen in the matching and treat-

[8] $\nu \equiv (\nu_0, \nu_1)$

ment effect literature, however, it is fundamentally different in the following perspectives. First, treatment effect measures the program impact ex post whereas Equation (2) recovers the impact of a policy that has not yet been implemented. Second, the hypothetical migration restriction is enforced onto the entire population rather than the observed migrants from the data because for a different population or a new generation prior to moving, it would be difficult and costly to identify who are potential movers. The policy impact recovered here is similar to the intend-to-treat (ITT) effect rather than the treatment effect on the treated. Third, a matching estimator usually matches on observed characteristics between treatment group and control group, which is infeasible in this case. A migrant household needs to be matched to a non-migrant household based on an appropriate level of non-migrant income and other characteristics but the econometrician only observes the migrant income for a migrant household. One must impute the counterfactual income before implementing a matching estimator. The procedure is discussed in detail in Section 5.

4.3 Evaluation of the Non-migration Subsidy

The second proposed policy experiment is to provide subsidy S to a household if the parents do not migrate. The subsidy is offered to the entire population subject to eligibility for reasons discussed previously. The program eligible groups considered in this paper include low-income families, families by number of children, and children by age and gender. Because the subsidy operates only through the budget constraint such that

$$
\begin{aligned}
C &= \mathbb{1}\{k=0\}\tilde{Y}_0 + \mathbb{1}\{k=1\}Y_1 \\
\tilde{Y}_0 &= Y_0 + S,
\end{aligned}
$$

the optimal migration decision and the corresponding child's cognitive outcome under the subsidy can be written as:

$$
\begin{aligned}
k^{''} &= \kappa(\tilde{Y}_0, Y_1, X, Z, \varepsilon, \nu), \\
Q_k^{''} &= \begin{cases} q_0(\tilde{Y}_0, Z, \nu_0), & \text{if } k^{''} = 0 \\ q_1(Y_1, Z, \nu_1), & \text{if } k^{''} = 1. \end{cases}
\end{aligned}
$$

The model solution with the subsidy is isomorphic to that without the subsidy, which implies that the effect of introducing subsidy S can be analyzed on the basis of the population variation in household income.

The impact of this subsidy is twofold. First, it provides incentives for parents to stay. Because the subsidy is given to the entire population, this impact can be seen as the difference between the population mean migration choice of households with income $\tilde{Y}_0$ and Y_1, observables X and Z and that of households with income Y_0 and same values of Y_1, X and Z, namely:

$$\triangle k^{sub}(Y_0, Y_1, X, Z) \equiv E[k^{''} \mid \tilde{Y}_0, Y_1, X, Z] - E[k^* \mid Y_0, Y_1, X, Z]. \tag{3}$$

The key identifying assumption for recovering the policy impact on the migration decision is that a household with non-migrant income Y_0 subject to subsidy S would, in expectation, make the same migration decision as another household with income $\tilde{Y}_0$, such that $\tilde{Y}_0 = Y_0 + S$, given that the two households have the same characteristics Y_1, X, and Z.

Next, this subsidy affects children's cognitive development through changes in parental migration status that determines the state of production technology and serves as a proxy for time inputs, and changes in income that serves as a proxy for material input in the production process. We are interested in the overall impact on children's cognitive achievements with respect to the entire population subject to eligibility. This effect is defined as the difference between the population mean of children's cognitive outcomes of households with income $\tilde{Y}_0$ and Y_1 and observed production covariates Z and that of households with income Y_0 and the same Y_1 and Z, namely:

$$\triangle Q^{sub}(Y_0, Y_1, Z) \equiv E[Q_k^{''} \mid \tilde{Y}_k, Z] - E[Q_k^{*} \mid Y_k, Z] \tag{4}$$

where

$$\tilde{Y}_k = \begin{cases} \tilde{Y}_0, & \text{if } k^{''} = 0 \\ Y_1, & \text{if } k^{''} = 1. \end{cases}$$

Note that the expected cognitive outcome and its augments are now indexed by k because the production technology depends on the realization of the migration decision. The key insight of identifying the policy's impact on children's cognitive outcomes is that two households that are similar in terms of observables X and Z, one with income Y_0 subject to subsidy S and the other with income $\tilde{Y}_0 = Y_0 + S$, would, on average, make the same investment decisions in the child skill formation process. The assumption is violated, for instance, when the proportion of income that is invested in children changes as the level of income increases given subsidy. Therefore, the impact of the subsidy is also measured at the extensive margin but not the intensive margin.

5 Estimation & Policy Analysis

The estimation of the proposed policy experiments is infeasible due to the partial observability of income. The econometrician observes migrant (non-migrant) income only if a household chooses to move (stay). One cannot match a migrant household to a non-migrant household nor to another migrant household based on an appropriate level of non-migrant income among characteristics. The same idea applies to a non-migrant household. To obtain a full set of income for every household, I specify the income equation as follows:

$$Y_k = g_k(W) + \eta_k$$

where W and $\eta \equiv (\eta_0, \eta_1)$ are the observed and unobserved components that determine income. The optimal migration decision k^* can therefore be rewritten as:

$$\begin{aligned} k^* =& \kappa(Y_0, Y_1, X, Z, \varepsilon, \nu) \\ =& \lambda(W, X, Z, \eta, \varepsilon, \nu). \end{aligned}$$

To estimate the income equations, note that

$$E[Y_k \mid W, k] = g_k(W) + \underbrace{E[\eta_k \mid \lambda(W, X, Z, \eta, \varepsilon, \nu)]}_{\neq 0} \tag{5}$$

where the second term on the right-hand side of the equation is most likely not zero due to the correlation between η_k and everything in $\lambda(\cdot)$. In the following, I first impute the full set of income under two sets of assumptions and then estimate the policy impacts accordingly.

5.1 Selection on Observables (SO)

I start with the relatively restrictive assumption that the dependence between latent income component η and migration choice k is due to the dependence between η and observed variables X and Z. More formally,

$$\begin{aligned} E[\eta_k \mid W, k] &\neq 0 \\ E[\eta_k \mid W, X, Z, k] &\neq 0 \\ E[\eta_k \mid W, X, Z, k] &= E[\eta_k \mid W, X, Z]. \end{aligned}$$

I refer to this set of assumptions as *Assumption SO*. In this case, controlling for observed selection variables X and Z overcomes the selection bias, that is, it removes the correlation between latent income η and migration choice k. Conditioning on additional observables X and Z, Equation (5) becomes:

$$
\begin{aligned}
E[Y_k \mid W,X,Z,k] &= g_k(W) + E[\eta_k \mid W,X,Z,k] \\
&= g_k(W) + E[\eta_k \mid W,X,Z] \\
&= m_k(W,X,Z).
\end{aligned} \tag{6}
$$

I estimate the conditional expectation $m_k(W,X,Z)$ nonparametrically via a multi-dimensional kernel regression:

$$
\hat{m}_k(W,X,Z) = \frac{\sum_{i\in S_Y} Y_{k_i} K\left(\frac{W-W_i}{h_W}\right) \mathbb{1}(X = X_i)\mathbb{1}(Z = Z_i)}{\sum_{i\in S_Y} K\left(\frac{W-W_i}{h_W}\right) \mathbb{1}(X = X_i)\mathbb{1}(Z = Z_i)}
$$

where the kernel function $K(\cdot)$ is for continuous variables, h is the bandwidth and the indicator function $\mathbb{1}(\cdot)$ is for discrete variables. The kernel estimation is performed on support $S_Y \equiv \{i \mid k_i^* \in \{0,1\}$ and $f(W_i,X_i,Z_i) > 0\}$ because the nonparametric estimator of $m_k(W,X,Z)$ is defined at points where the density $f(W_i,X_i,Z_i)$ is positive. To determine the support S_Y, I first estimate density $f(W,X,Z)$ nonparametrically and then select observations with positive density estimates.

To preserve the sample size, I use a Gaussian kernel function because it has infinite support which gives positive density estimates at all points of evaluation. The optimal bandwidth is computed using the Silverman formula. The choice of kernel function and bandwidth satisfies the standard assumptions in kernel estimation and provides a consistent estimator (Heckman, Ichimura & Todd, 1997). I compute the counterfactual non-migrant income for a migrant household by plugging this household's covariates into the estimate of $m_0(W,X,Z)$. The migrant income for a non-migrant household is computed in a similar fashion. In practice, I use parental age and education as the observed income component W. The production covariate Z includes child age and gender, number of siblings and whether at least one grandparent lives within the household. I use distance to the nearest bus station as a utility shifter X.

Table 5 summarizes the full income distribution estimated the *Assumption SO*. The diagonal elements are observed income whereas the off-diagonal elements are imputed ones (in bold). For instance, had stayed a migrant household would have earned 33937RMB on average, which is much lower than this household receives in urban areas, 42646RMB. The large income gap between urban destinations and rural areas, roughly 7000RMB for the entire sample, is likely to rationalize the parental rural-to-urban migration that we commonly see in China. In addition, migrant households would have earned more than their non-migrant counterparts regardless of location. The difference in earnings between groups at the same location is attributed to observed characteristics under *Assumption SO*.

5.1.1 Effect of Migration Restriction (SO)

Given the complete set of income, I proceed to estimate the impact of migration restriction on child outcomes. Because the only relevant migration decision under this policy regime is to stay, I rewrite the policy impact in Equation (2) as follows:

$$\triangle Q^{res}(Y_0, Y_1, Z) = E[Q_0^{'} \mid Y_0, Z, k^{'} = 0] - E[Q_k^{*} \mid Y_k, Z]. \tag{7}$$

Under the assumption that the distribution of unobserved production heterogeneity ν is independent of the migration decision k, conditional on income Y and covariates Z, one can estimate the conditional child quality of migrant households using the information from non-migrant households.[9] In other words, when a household is not allowed to migrate, its child quality would be the same in expectation as the child quality from an identical household, in terms of Y_0 and Z, that would have stayed regardless of migration restriction. Integrating over the joint distribution of Y_0, Y_1, and Z provides the impact of migration restriction on the expected child quality over the entire population, that is,

$$\triangle Q^{res} = \int \triangle Q^{res}(Y_0, Y_1, Z)\, dG(Y_0, Y_1, Z). \tag{8}$$

[9]The assumption translates into $E[Q_0 \mid Y_0, Z, k = 1] = E[Q_0 \mid Y_0, Z, k = 0]$

The corresponding matching estimator of Equation (8) takes the form:

$$\hat{\triangle}Q^{res} = \frac{1}{N_{S_R}} \sum_{i \in S_R} \left\{ \hat{E}[Q_{i0} \mid Y_{i0}, Z_i, k_i = 0] - Q_{ik}(Y_{ik}, Z_i) \right\} \tag{9}$$

where $\hat{E}[Q_{i0} \mid Y_{i0}, Z_i, k_i = 0]$ is the predicted cognitive achievement under the migration restriction and $Q_{ik}(Y_{ik}, Z_i)$ is the observed cognitive achievement from the data. The matches are performed on the support $S_R \equiv \{i \mid f(Y_{i0}, Z_i) > 0\}$. Because this policy concerns the entire population, support S_R includes all observations in the sample subject to trimming. The expected child outcome is estimated nonparametrically by a kernel regression of Q_{j0} on Y_{j0} and Z_j evaluated at points Y_{i0} and Z_i using only observations whose observed decision is to stay, that is,

$$\hat{E}[Q_{i0} \mid Y_{i0}, Z_i, k_i = 0] = \frac{\sum\limits_{j \in S_N} Q_{j0} K\left(\frac{Y_{i0} - Y_{j0}}{h_0}\right) \mathbb{1}(Z_i = Z_j)}{\sum\limits_{j \in S_N} K\left(\frac{Y_{i0} - Y_{j0}}{h_0}\right) \mathbb{1}(Z_i = Z_j)}$$

where $S_N \equiv \{j \mid k_j^* = 0 \text{ and } f(Y_{j0}, Z_j) > 0\}$. Note that evaluation point Y_{i0} is the observed non-migrant income for a non-migrant household but is the imputed non-migrant income for a migrant household. I use the child outcome of all non-migrant households, matched on observed characteristics and imputed income level, as the counterfactual child outcome of migrant household.

The first two rows in Table 7 report the impact of migration restriction on children's cognitive achievements under *Assumption SO*. The second column provides predicted changes in mean grade points whereas the last column reports the standard deviation change from the mean score to provide the magnitude and relevance of the impact. When anchored with middle school graduation, the migration restriction increases graduation probability by 2.8%.

5.1.2 Effect of the Non-migration Subsidy (SO)

Next, I turn to the estimation of the subsidy impact in two steps. First, the non-migration subsidy incentivizes households to stay. Under the assumption that conditional on X and Z, the joint distribution of unobservables ε and ν does not depend on income Y_0 and Y_1, the policy impact in

Equation (3) becomes:

$$\triangle k^{subsidy}(Y_0, Y_1, X, Z) = \int_{\varepsilon}\int_{\nu} \kappa(\tilde{Y}_0, Y_1, X, Z, \varepsilon, \nu) f(\varepsilon, \nu \mid X, Z)\, d\varepsilon d\nu \\ - \int_{\varepsilon}\int_{\nu} \kappa(Y_0, Y_1, X, Z, \varepsilon, \nu) f(\varepsilon, \nu \mid X, Z)\, d\varepsilon d\nu. \tag{10}$$

The difference in Equation (10) is exactly the effect of introducing the non-migration subsidy on the expected migration behavior of a household with characteristics Y_0, Y_1, X, and Z. Integrating over the joint distribution of Y_0, Y_1, X, and Z provides the impact of the subsidy on the expected migration behavior over the entire population, that is,

$$\triangle k^{subsidy} = \int \triangle k^{subsidy}(Y_0, Y_1, X, Z)\, dG(Y_0, Y_1, X, Z). \tag{11}$$

The matching estimator of Equation (11) aims to estimate the overall mean impact of the non-migration subsidy on the migration choice, and it takes the form:

$$\hat{\triangle} k^{subsidy} = \frac{1}{N_{S_k}} \sum_{i \in S_k} \left\{ \hat{E}[k_i \mid \tilde{Y}_{i0}, Y_{i1}, X_i, Z_i] - k_i(Y_{i0}, Y_{i1}, X_i, Z_i) \right\} \tag{12}$$

where $k_i(Y_{i0}, Y_{i1}, X_i, Z_i)$ is the migration decision of household i in the absence of the subsidy, and $\hat{E}[k_i \mid Y_{i0} + S_i, Y_{i1}, X_i, Z_i]$ is the predicted migration choice of household i if faced with the subsidy. The matches are performed for N_{S_k} observations that lie in the support $S_k \equiv \{i \mid f(Y_{i0}, Y_{i1}, X_i, Z_i) > 0\}$. If the subsidy is provided unconditionally to all households, support S_k includes the entire sample subject to trimming based on the empirical density estimate of $f(Y_1, Y_0, X, Z)$. If the subsidy is given to a program-eligible group, S_k takes only observations in the eligible group subject to trimming. I estimate the predicted migration decision $\hat{E}[k_i \mid Y_{i0} + S_i, Y_{i1}, X_i, Z_i]$ nonparametrically using a kernel regression of k_j on Y_{j0}, Y_{j1}, X_j, and Z_j evaluated at points $\tilde{Y}_{i0} = Y_{i0} + S, Y_{i1}, X_i$, and Z_i for all observations $j \in S_k$. The idea is to select all individuals j in support S_k whose characteristics Y_{j0}, Y_{j1}, X_j, and Z_j are similar to the characteristics of individual i and use these individuals to predict individual i's migration behavior.

The top two rows in Table 8 present the predicted impact of the non-migration subsidy on migration rate under *Assumption SO*. In particular, I fix subsidy level S_i to be the same across households, i.e., $S_i = S$, and choose subsidy levels of 3000 RMB, 6000 RMB, 9000 RMB and 12000 RMB if a household does not migrate. The estimates are negative and significant, which indicates a reduction in migration when subsidy is provided. I plot the migration rate per 500 RMB increment in Figure 2. I find that the migration rate decreases as the subsidy increases. The migration rate is inelastic at lower subsidy levels and becomes more elastic when the subsidy exceeds 5000 RMB. This is likely due to the fact that the smaller subsidy does not generate enough incentive for households to stay due to the huge income gap between urban and rural areas.

In addition to migration behavior, the introduction of the non-migration subsidy affects the cognitive achievements of children. I continue to assume that the distribution of unobserved production heterogeneity is independent of the migration decision, conditional on income and production covariates. The corresponding matching estimator of Equation (4) takes the form:

$$\hat{\triangle}Q^{subsidy} = \frac{1}{N_{S_Q}} \sum_{i \in S_Q} \left\{ \hat{E}[Q_{ik} \mid \tilde{Y}_{ik}, Z_i] - Q_{ik}(Y_{ik}, Z_i) \right\} \tag{13}$$

where the support S_Q is defined as $S_Q \equiv \{i \mid f(Y_{ik}, Z_i) > 0\}$. Support S_Q contains all households subject to trimming if the subsidy is provided unconditionally and only eligible households if there are eligibility criteria. Because the predicted migration choice in Equation (12) is a probability between 0 and 1, the expected cognitive achievement given subsidy is a weighted average:

$$\hat{E}[Q_{ik} \mid \tilde{Y}_{ik}, Z_i] = \hat{E}[Q_{i0} \mid \tilde{Y}_{i0}, Z_i, k_i = 0]\hat{Pr}(k_i = 0) + \hat{E}[Q_{i1} \mid Y_{i1}, Z_i, k_i = 1]\hat{Pr}(k_i = 1).$$

The choice probabilities $\hat{Pr}(k_i = 0)$ and $\hat{Pr}(k_i = 1)$ are obtained from the previous step for each observation. Next, I estimate the expected cognitive outcome when a household stays, i.e., $\hat{E}[Q_{i0} \mid \tilde{Y}_{i0}, Z_i, k_i = 0]$, using all observed stayers in support $S_N \equiv \{j \mid k_j^* = 0 \text{ and } f(Y_{j0}, Z_j) > 0\}$ and then evaluate at point $\tilde{Y}_{i0}$ and Z_i for each individual i in support S_Q. I construct $\hat{E}[Q_{i1} \mid Y_{i1}, Z_i, k_i = 1]$ in a similar way but only using observed movers.

Table 9 presents the ex ante evaluation of the impact of the subsidy on children's cognitive achievements under *Assumption SO*. The first two rows present the predicted cognitive test scores measured in grade point. The next two rows provide the standard deviation change from the mean score. The last row shows middle school graduation probability once test scores are anchored using Equation (1). The program effect on test scores and graduation probability increases with the subsidy level because parents are incentivized to stay in the rural areas with higher non-migrant income and hence likely to invest more time and material inputs in their children.

Panel 1 in Figure 3 compares the predicted program effect when using the full sample with the low-income subsample, with a low-income household defined as whose non-migrant income is below the median of the full non-migrant income distribution. The observed migrant income in urban sector do not necessarily reflect the true economic status of a migrant household in rural area. Therefore, using observed income combining both urban and rural sectors as a selection criterion mistargets the actual low-income population in rural regions. More importantly, for a different population such that no one has yet moved, it would be impossible, from a policy maker's point of view, to target low-income families based on migrant income. The impact is greater for the low-income subsample at all subsidy levels. For instance, a subsidy of 14000RMB (2000 USD) increases word and math test score by 7.1% and 3.5% standard deviation respectively, in contrast to 3.1% increase in word score and 2.8% increase in math score for the full sample. This level of subsidy translates into a sizable increase, 6.9%, in terms of graduation probability. These findings indicate that a non-migration subsidy that targets low-income households leads to the greater increase in test scores and improvement in socioeconomic outcomes such middle school graduation rate.

Panel 1 & 2 in Figure 4 plot the word and math test scores against the subsidy level between different age groups. The impact of subsidy on the younger group dominates the impact on the older group at every subsidy level for both scores.[10] For example, a subsidy of 14000RMB (2000 USD) is associated with a 6.2% increase in graduation probability for the younger group in contrast to a 3.8% increase for the older group. Panel 1 & 2 in Figure 5 compares the impact of subsidy on child quality between boys and girls. The effect of subsidy on girls is slightly higher than the one on boys for word score at every subsidy level. On the other hand, the pattern is not clear when it

[10] I do not consider finer age groups because the sample sizes are too small for credible nonparametric estimation.

comes to math score. At a low level of subsidy, the impact is almost the same between the two. The impact on boys dominates the one on girls once the subsidy reaches 8000RMB. Under the assumption that fertility is exogenous, Panel 1 & 2 in Figure 6 demonstrate the impact of subsidy based on the number of children per household. The impact is greater for households with more than one child, particularly when the subsidy level is high.

5.2 Selection on Unobservables (SU)

In the previous section, when subsidy reaches 10000 RMB, which exceeds the difference in mean income between urban sector and rural area, the migration rate is only reduced to 18%. It is likely that the conditional independence assumption is violated, which naturally leads to the discussion on selection on unobservables. Specifically, note that the dependence between latent income η and the migration choice k may not be eliminated even after controlling for observed variables X, W, and Z. That is,

$$E[\eta_k \mid W, k] \neq 0$$
$$E[\eta_k \mid W, X, Z, k] \neq E[\eta_k \mid W, X, Z].$$

Selection now depends on unobservables. I assume that (ε, ν, η) have a joint distribution and utility is linear in all error terms.[11] I refer to this set of assumption as *Assumption SU*. In this case, the migration decision is given by:

$$\begin{aligned} k &= 0 \quad \text{iff} \quad U_0 \geq U_1, \\ k &= 1 \quad \text{otherwise}. \end{aligned}$$

[11]The assumption that utility is in all error terms requires that the arguments of the utility function $U(\cdot)$ and the arguments of the production function $q_k(\cdot)$ are additively separable. Thus, the utility function can be expressed as $U(C, Q_k, k, X, \varepsilon) = V_k(X, W, Z) + \eta_k + \nu_k + \varepsilon_k$.

The probability that a household with observables W, X, and Z chooses to stay is:

$$\begin{aligned}\Pr(k=0 \mid X,W,Z) &= \Pr(U_0 \geq U_1 \mid X,W,Z) \\ &= \Pr(\underbrace{\eta_1+\nu_1+\varepsilon_1-(\eta_0+\nu_0+\varepsilon_0)}_{\equiv\delta} \leq \underbrace{V_0(X,W,Z)-V_1(X,W,Z)}_{\equiv V(X,W,Z)}).\end{aligned}$$

The conditional expected income of Equation (6) can be expressed as:

$$\begin{aligned} E[Y_0 \mid X,W,Z,k=0] &= g_0(W) + \underbrace{\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{V(X,W,Z)} \eta_0 f(\eta_0,\delta)\, d\eta_0 d\delta}{\int_{-\infty}^{\infty}\int_{-\infty}^{V(X,W,Z)} f(\eta_0,\delta)\, d\eta_0 d\delta}}_{\equiv G_0(X,W,Z)} \\ E[Y_1 \mid X,W,Z,k=1] &= g_1(W) + \underbrace{\frac{\int_{-\infty}^{\infty}\int_{V(X,W,Z)}^{\infty} \eta_1 f(\eta_1,\delta)\, d\eta_1 d\delta}{\int_{-\infty}^{\infty}\int_{V(X,W,Z)}^{\infty} f(\eta_1,\delta)\, d\eta_1 d\delta}}_{\equiv G_1(X,W,Z)}. \end{aligned} \tag{14}$$

With exclusion restrictions X and Z that affect the migration choice but do not influence income directly, one can nonparametrically identify functions $g_0(\cdot), g_1(\cdot), G_0(\cdot)$ and $G_1(\cdot)$ including constant terms under the argument of identification at infinity. The constant terms in $g_k(\cdot)$ and $G_k(\cdot)$ must be identified separately because the matching estimator requires one to compare the migration decision and child's cognitive outcome between a household with income Y_0 and a household with income $\tilde{Y}_0$.

In the estimation, I use the distance to the nearest bus station included in the utility shifter X as an exclusion restriction because the shorter the distance is, the less difficult it is to move to an urban destination for work, but the distance itself does not directly affect income. The distance to a bus stop is a valid exclusion restriction under the assumptions that rural residents have lived in their current location for decades and therefore residential location is exogenous. This assumption is empirically validated from the data. On average, rural households have resided in their current home location for over 15 years for the given sample. It is unlikely that these households select their residential location with respect to transportation opportunities. In addition, the bus station under consideration is a specific type of station that offers long distance buses that allow rural residents to travel to urban destinations rather than local buses that offer transportation within a village. Given the fact that rural residents usually engage in local agricultural and farming activities, roughly 65% of the population per village, it is unlikely that the availability of long distance bus impacts non-migrant income directly. Lastly, the number of children in Z can also be used as an additional exclusion restriction if fertility is assumed to be exogenous.

Given the strong requirement on the support under identification at infinity, I impose a distributional assumption on the error terms and estimate the income equations semiparametrically. Formally, I assume that (ε, η, ν) follow a joint normal distribution $N(0, \Sigma)$.[12] In this case, the migration probability becomes:

$$\Pr(k = 0 \mid X, W, Z) = \Phi(\underbrace{\frac{V(X, W, Z)}{\sigma_\delta}}_{\equiv V_\delta}). \tag{15}$$

The conditional income from Equation (14) becomes:

$$\begin{aligned} E[Y_0 \mid X, W, Z, k = 0] &= g_0(W) + \alpha_0 \underbrace{\frac{-\phi(V_\delta)}{\Phi(V_\delta)}}_{\equiv H_0} \\ E[Y_1 \mid X, W, Z, k = 1] &= g_1(W) + \alpha_1 \underbrace{\frac{\phi(V_\delta)}{1 - \Phi(V_\delta)}}_{\equiv H_1}. \end{aligned} \tag{16}$$

Note that function $g_k(\cdot)$, including its intercept, and parameter α_k are identified without a confounding constant component in H_k. To obtain estimates of $g_k(\cdot)$ and α_k, I estimate the following partially linear model in several steps:

$$Y_k = g_k(W) + \alpha_k H_k + u_k \tag{17}$$

First, because the left-hand side of the Equation (15) is directly identifiable and thus estimable from the data, I obtain an estimate of V_δ from the inverse of a normal CDF of the estimated $\Pr(k = 0 \mid X, W, Z)$. The correction term H_k is then constructed from the estimate of V_δ. Next,

[12] Although not all terms in variance-covariance matrix Σ are identified, it is sufficient to identify a combination of them to estimate the income equations.

using the technique proposed by Robinson (1988), I subtract Equation (17) by its conditional expectation:

$$\underbrace{Y_k - E[Y_k \mid W]}_{\equiv \tilde{Y}_k} = \alpha_k \underbrace{[\hat{H}_k - E[H_k \mid W]]}_{\equiv \tilde{H}_k} + \underbrace{u_k - E[u_k \mid W]}_{\equiv \tilde{u}_k}.$$

I run nonparametric regressions of Y_k and H_k on W to obtain $\tilde{Y}_k$ and $\tilde{H}_k$. Parameter α_k is estimated from an OLS regression of $\tilde{Y}_k$ on $\tilde{H}_k$ subject to trimming. Finally, given the knowledge of Y_k, H_k and α_k, I construct residual $\hat{e}_k \equiv Y_k - \hat{\alpha}_k H_k$ and estimate $g_k(\cdot)$ from a nonparametric regression of $\hat{e}_k$ on W. After all the estimates are obtained, I compute the migrant income for a non-migrant household and the non-migrant income for a migrant household using Equation (16).

Table 6 reports the descriptive statistics of the full set of income estimated under *Assumption SU*. In comparison to the income distribution estimated under *Assumption SO*, the income gap between urban and rural areas is much greater, roughly 40% increase for the entire sample. This finding partially rationalizes the previous puzzle estimated under *Assumption SO* such that a subsidy level covering rural urban income gap only leads to a small change in migration rate since the actual income gap is greater once selection on unobservables is controlled for.

5.2.1 Effect of Migration Restriction (SU)

The matching estimator for child's cognitive achievement under migration restriction takes the same form as Equation (9), but the predicted cognitive outcome is estimated differently due to selection on the latent component ν in the production technology. I first estimate the following partially linear model to obtain estimates of $q_k(\cdot)$ and β_k:

$$Q_k = q_k(Y_k, Z) + \beta_k H_k + r_k.$$

I then construct the expected child's cognitive outcome as follows:

$$\hat{E}[Q_{i0} \mid Y_{i0}, Z_i, k_i = 0] = \begin{cases} \hat{q}_0(Y_{i0}, Z_i) + \hat{\beta}_0 \frac{-\phi(\hat{V}_{\delta i})}{\Phi(\hat{V}_{\delta i})}, & \text{if } k_i^* = 0 \\ \hat{q}_0(Y_{i0}, Z_i) + \hat{\beta}_0 \frac{\phi(\hat{V}_{\delta i})}{1-\Phi(\hat{V}_{\delta i})}, & \text{if } k_i^* = 1. \end{cases}$$

The last two rows in Table 7 present the effect of the migration restriction under *Assumption SU*. The impact on word test scores is negative yet insignificant, whereas the impact on math test scores is positive and significant, as in the case of selection on observables. The associated middle school graduation probability is increased trivially by 1.2% and not economically meaningful.

5.2.2 Effect of the Non-migration Subsidy

The impact of the non-migration subsidy on migration behavior is computed in the same way as Equation (12) if we additionally assume η is independent of ε and ν. The additional assumption does not rule out selection on unobservables because the composite error δ is correlated with latent income η through η itself. Table 8 analyzes the difference in predicted impacts on migration rates between two sets of assumptions. The change in migration rate is overestimated in absolute value when selection on unobservables is ignored.

I estimate the impact of non-migration subsidy on child quality using the same matching strategy as seen in Equation (13) after correcting selection on unobservables. Table 10 presents estimation results. The impact of subsidy on the word score dominates the impact on the math score at all subsidy levels as before. However, the marginal impact of each dollar spent is small at low levels of the subsidy for the case of selection on unobservables and increases sharply once the subsidy becomes high. For instance, the graduation rate increases by 12.0% at 12000 RMB (1800 USD) when compared to the case of selection on observables.

Panel 2 in Figure 3 demonstrates the policy impact when the subsidy targets low-income families under *Assumption SU*. The policy impact is substantial. When the subsidy approaches 8000 RMB (1200 USD), the standard deviation changes from the mean in both test scores for the low-income group are twice as much as the ones for the entire sample. When the subsidy reaches 14000 RMB (2000 USD), it improves children's word scores by 10% standard deviation and math scores by 6% standard deviation, which translates to a substantial increase, 8.6%, in graduation probability for rural children from the bottom half of the income distribution. When compared to the case of selection on observables, the predicted impact is also greater in terms of standard deviation change as well as graduation rate (8.6% vs. 6.9%), especially when the subsidy is high. These findings provide additional evidence to the existing literature that cash transfer programs that target low-income families is more effective in improving education attainment than unconditional ones in developing countries (Bourguignon, Ferreira & Leite, 2003; Schultz, 2004).

As the case of selection on observables, Panel 3 & 4 in Figure 4 show the impact by age. Younger adolescents aged 10-12 benefit more from the subsidy than older ones. Particularly, a subsidy of 14000 RMB (2000 USD) targeting children aged 10-12 improves their word and math test scores by 8% and 6% standard deviation, respectively. When associated with graduation probability, this subsidy translates to a 7.6% increase for the younger group and a 5.3% increase for the older group. This finding coincides with the case of selection on observables and suggests that subsidy is more effective when it targets younger adolescents. Panel 3 & 4 in Figure 5 depict the impact of subsidy is greater on boys than on girls. Panel 3 & 4 Figure 6 show the impact on word score is greater for households with one child and trend is not clear for math score.

6 Conclusion

Parental rural-to-urban migration has affected a large number of children in China. The negative impact documented in previous literature has raised considerable concerns for policy makers. This paper contributes to the literature by (1) proposing several migration policies that are aimed at improving left-behind children's cognitive achievements; and (2) quantifying the potential impacts of policy changes prior to their implementation. To achieve these goals, I develop a model of household migration that nests a child's cognitive skill formation process. I exploit the variation in household income and directly estimate the policy impacts nonparametrically using data from the China Family Panel Studies (CFPS). Ex ante evaluation of the impacts of the migration policies on cognitive achievements and associated middle school graduation probability suggests that a subsidy is most effective when it targets low-income families and younger children.

The findings in this paper to a certain extent provide policy makers what to expect if the policy recommendation, State Council Opinions on Efforts to Strengthen Care and Protection of Rural Left-Behind Children (Ministry of Human Resources and Social Security of the People's Republic of China, 2016), is fully implemented. On the one hand, the migration restriction, although fully stops circulation of migrant workers, only improves left-behind children's cognitive achievements and associated socioeconomic outcome marginally. On the other hand, the non-migration subsidy, equivalent to tax cut if migrant workers move back as proposed in the policy recommendation, is relatively costly. One way to reduce program costs is to target low-income households or younger children, which cuts program costs in half. However, the subsidy lifts graduation probability by 8.6% for rural children from the bottom of the income distribution. Given the fact that the returns to education in rural China has increased drastically over the past several decades (De Brauw & Rozelle, 2008), the high cost could be compensated by economic development in rural China.

I conclude this paper by considering several extensions of this work. First, the intensive margin of the subsidy is not studied in this paper. Quantification of the intensive margin requires explicitly modeling investment decisions in a child skill formation process. When material and time inputs are involved explicitly, the nonparametric matching technique used in this paper does not carry over. Second, it would be useful to gather additional data once the policy is fully implemented in the future. Comparison between ex ante evaluation and ex post effect adds discussion on model validation as discussed in (Todd & Wolpin, 2006, 2008). Therefore, future work is needed.

References

Agostinelli, F., & Sorrenti, G. (2018). Money vs. time: family income, maternal labor supply, and child development.

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.

Bai, Y., Zhang, L., Liu, C., Shi, Y., Mo, D., & Rozelle, S. (2017). Effect of parental migration on the academic performance of left behind children in North Western China. *Journal of Development Studies*, 1-17.

Bernal, R. (2008). The Effect of Maternal Employment and Child Care on Childrens Cognitive Development. *International Economic Review* 49 (4), 11731209.

Bernal, R., Keane, M. P. (2011). Child Care Choices and Childrens Cognitive Achieve- ment: The Case of Single Mothers. *Journal of Labor Economics* 29 (3), 459512.

Bourguignon, F., Ferreira, F. H., & Leite, P. G. (2003). Conditional cash transfers, schooling, and child labor: Micro-simulating Brazil's Bolsa Escola program. *World Bank Economic Review*, 17(2), 229-254.

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

Chen, X., Huang, Q., Rozelle, S., Shi, Y., & Zhang, L. (2009). Effect of migration on childrens educational performance in rural China. *Comparative Economic Studies*, 51, 323343.

Chen, Y., & Feng, S. (2013). Access to public schools and the education of migrant children in China. *China Economic Review*, 26, 75-88.

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 738-782.

Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3), 883-931.

De Brauw, A., & Rozelle, S. (2008). Reconciling the returns to education in offfarm wage employment in rural China. *Review of Development Economics*, 12(1), 57-71.

Duncan, M., & Mao, S. (2011) Closure of migrant children schools in China sparks anguish. *Reuters*. Retrieved from https://www.reuters.com/article/us-china-migrant-schools/closure-of-migrant-children-schools-in-china-sparks-anguish-idUSTRE77H10I20110818

Edes, A., & Inskeep, S. (2017) In Chinese Cities, Migrants' Work Is Welcome. Their Children Are Not. *National Public Radio*. Retrieved from https://www.npr.org/sections/parallels/2017/11/16/564380188/

Heckman, J. J. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *Quarterly Journal of Economics*, 115(1), 45-97.

Heckman, J. J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, 109(4), 673-748.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64(4), 605-654.

Hernández, J. C., & Zhao, I (2017, December 24) One Target in Beijings Migrant Crackdown: Schoolchildren. *New York Times*. Retrieved from https://www.nytimes.com/2017/12/24/world/asia/china-schools-migrants.html

Ichimura, H., & Taber, C. R. (2000). Direct estimation of policy impacts.

Ichimura, H., & Taber, C. (2002). Semiparametric reduced-form estimation of tuition subsidies. *American Economic Review*, 92(2), 286-292.

Keane, M. P., & Wolpin, K. I. (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42(4), 1051-1103.

Lee, L., & Park, A. (2010). Parental migration and child development in China.

Liu, L. (2015). The effects of education restrictions on rural urban migration in China.

Marschak, J. (1953). Economic measurements for policy and prediction. In *Studies in Econometric Method*, ED. W.C. Hood and T.C. Koopmans, pp. 1-16. New Haven: Yale University Press.

McKenzie, D., & Rapoport, H. (2011). Can migration reduce educational attainment? Evidence from Mexico. *Journal of Population Economics*, 24(4), 1331-1358.

Meng, X., & Yamauchi, C. (2017). Children of Migrants: The cumulative impact of parental migration on childrens education and health outcomes in China. *Demography*, 54(5), 1677-1714.

Ministry of Human Resources and Social Security of the People's Republic of China. (2016) *State Council Opinions on Efforts to Strengthen Care and Protection of Rural Left-Behind Children*

Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, 931-954.

Schultz, T. P. (2004). School subsidies for the poor: evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, 74(1), 199-250.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485).

Todd, P. E., & Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic review*, 96(5), 1384-1417.

Todd, P. E., & Wolpin, K. I. (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*, 1(1), 91-136.

Todd, P. E., & Wolpin, K. I. (2008). Ex ante evaluation of social programs. *Annales d'Economie et de Statistique*, 263-291.

Wen, M., & Lin, D. (2012). Child development in rural China: Children left behind by their migrant parents and children of non-migrant families. *Child Development*, 83(1), 120-136.

Wolpin, K. I. (2013). *The limits of inference without theory*. MIT Press.

Xu, H., & Xie, Y. (2015). The causal effects of rural-to-urban migration on childrens well-being in China. *European Sociological Review*, 31(4), 502-519.

Xie, Y., Zhang, X., Tu, P., & Ren, Q. (2008). *China Family Panel Studies (CFPS)*. Ann Arbor, MI: [distributor], 2018-01-25. https://doi.org/10.3886/ICPSR36524.v2

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

Zhang, H., Behrman, J. R., Fan, C. S., Wei, X., & Zhang, J. (2014). Does parental absence reduce cognitive achievements? Evidence from rural China. *Journal of Development Economics*, 111, 181-195.

Zhou, M., Murphy, R., & Tao, R. (2014). Effects of parents' migration on the education of children left behind in rural China. *Population and Development Review*, 40(2), 273-292.

Table 1: Summary Statistics

	Non-migrant	Left-behind	
	Mean (St. Dev.)	Mean (St. Dev.)	Diff. in Mean (St. Err.)
Father Education	2.53 (0.95)	2.66 (0.86)	-0.13 (0.023)
Mother Education	2.29 (1.08)	2.33 (0.98)	-0.042 (0.026)
Father Age	36.54 (6.73)	34.85 (6.40)	1.69 (0.17)
Mother Age	34.75 (6.69)	33.17 (6.24)	1.59 (0.17)
Father Migration Duration	-	11.27 (27.26)	-
Mother Migration Duration	-	10.55 (20.99)	-
Grandparent	0.64 (0.48)	0.78 (0.41)	-0.143 (0.011)
Bus Distance	2.35 (5.17)	1.74 (2.27)	0.61 (0.085)
Number of Children	1.79 (0.85)	1.91 (0.99)	-0.11 (0.025)
Household Income	32565 (34942)	43378 (39867)	-10813 (1012)

[a] Education: 1-Illiterate/Semi-literate, 2-Primary school, 3-Junior high school, 4-Senior high school, 5-2 or 3 year college, 6-Bachelor's degree, 7-Master's degree, 8-Doctoral degree.
[b] Migration duration is measured in month.
[c] The variable Grandparent indicates whether there is at least one grandparent living with the child.
[d] The variable Bus Distance is the distance to the closest bus stop, measured in km.

Table 2: Cognitive Test Scores by Age

	Word Test Score			Math Test Score		
Age	Non-migrant Mean (St. Dev.)	Left-behind Mean (St. Dev.)	Diff. in Mean (St. Err.)	Non-migrant Mean (St. Dev.)	Left-behind Mean (St. Dev.)	Diff. in Mean (St. Err.)
10	15.42	14.53	0.89	7.13	6.54	0.58
	(6.86)	(6.37)	(0.89)	(2.97)	(2.79)	(0.29)
11	18.21	17.81	0.40	7.79	8.19	-0.39
	(6.82)	(6.72)	(0.89)	(2.97)	(2.94)	(0.39)
12	20.26	20.59	-0.33	9.46	9.75	-0.30
	(7.14)	(5.54)	(0.81)	(3.91)	(3.29)	(0.47)
13	22.10	22.06	0.04	11.07	10.42	0.64
	(6.62)	(6.75)	(0.94)	(3.78)	(3.99)	(0.55)
14	24.31	24.47	-0.16	12.58	11.32	1.25
	(6.58)	(5.98)	(0.85)	(4.18)	(4.51)	(0.62)
15	24.98	25.00	-0.02	13.68	13.24	0.44
	(6.81)	(5.60)	(0.73)	(4.34)	(4.92)	(0.60)

[a] Word test score is based on a 34-item Chinese word test.
[b] Math test score is based on a 24-item mathematical test.

Table 3: Comparison of Cognitive Test Scores

Panel A: Test Scores Comparison

	Word Test Score Diff. in mean (St. Err.)	Math Test Score Diff. in mean (St. Err.)
Naive Estimator	-0.07 (0.39)	-0.32 (0.24)
Propensity Score Matching Estimator	-0.39 (0.54)	-0.59 (0.31)

Panel B: Probit Regression for Migration

	Coefficient	p-Value
log(Income)	0.17	0.00
Father Education	0.04	0.03
Mother Education	-0.52	0.00
Father Age	-0.01	0.03
Mother Age	-3.2e-03	0.56
Child Gender	0.05	0.12
Child Age	-0.01	0.36
Number of Siblings	0.08	0.00
Grandparent	0.30	0.00
Bus Distance	-0.03	0.00
Constant	-0.48	0.00

[a] Education: 1-Illiterate/Semi-literate, 2-Primary school, 3-Junior high school, 4-Senior high school, 5-2 or 3 year college, 6-Bachelor's degree, 7-Master's degree, 8-Doctoral degree.
[b] Gender equals 1 if male.
[c] The variable Grandparent indicates whether there is at least one grandparent living with the child.
[d] The variable Bus Distance is the distance to the closest bus station, measured in km.

Table 4: Estimates of Anchoring Cognitive Test Scores

	Coefficient
Word Test Score	0.062 (0.010)
Math Test Score	0.261 (0.016)
Intercept	-4.800 (0.234)

[a] Standard errors are reported in the parentheses.

Figure 1: Observed Income Distribution

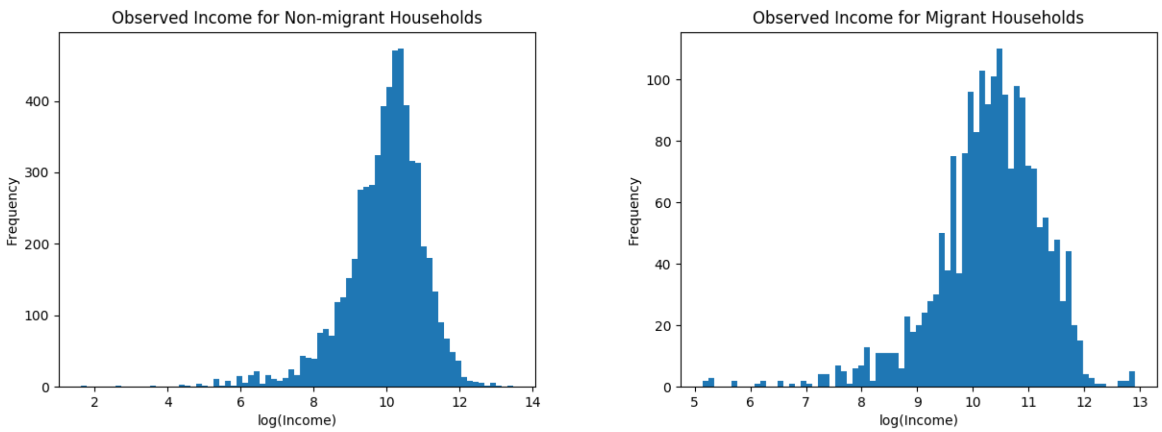

Table 5: Full Income Distribution (SO)

	Non-migrant Income	Migrant Income	
	Mean (St. Dev.)	Mean (St. Dev.)	Mean of diff. (St. Err.)
Non-migrant Household	32425 (34744)	**39341** **(10100)**	-6916 (465)
Migrant Household	**33937** **(4393)**	42642 (39361)	-8705 (890)
All Household	32802 (30191)	40164 (21556)	-7362 (414)

[a] The observed income is slightly different from the one provided in Table 1 because I drop observations that are not in common support S_Y in the process of estimating counterfactual income.

Table 6: Full Income Distribution (SU)

	Non-migrant Income	Migrant Income	
	Mean (St. Dev.)	Mean (St. Dev.)	Mean of diff. (St. Err.)
Non-migrant Household	32469 (34811)	**42616** **(19565)**	-10147 (497)
Migrant Household	**31478** **(13571)**	43385 (39870)	-11907 (918)
All Household	32220 (30894)	42810 (26175)	-10590 (439)

[a] The observed income is different from the ones in Table 1 and Table 5 because I drop observations that are not in the common support and the support changes under different set of assumptions.

Table 7: Predicted Impact of the Migration Restriction on Child Outcome

	Diff. in Mean	St. Dev. Change
Word Test Score (SO)	0.046 (0.094)	0.7%
Math Test Score (SO)	0.099^{*} (0.053)	2.2%
Word Test Score (SU)	-0.11 (0.095)	-1.5%
Math Test Score (SU)	0.077^{*} (0.043)	1.7%

[a] Bootstrap standard errors based on 500 replications are reported in the parentheses.

Table 8: Predicted Impact of the Subsidy on the Migration Rate

Subsidy Level	3000 RMB	6000 RMB	9000 RMB	12000 RMB
Migration Rate (SO)	-0.001^{***} (0.000)	-0.019^{***} (0.004)	-0.051^{***} (0.006)	-0.083^{***} (0.007)
Migration Rate (SU)	-0.001^{***} (0.000)	-0.013^{***} (0.002)	-0.043^{***} (0.006)	-0.071^{***} (0.007)

[a] 1 USD = 6.77 RMB (Year 2010).
[b] Bootstrap standard errors based on 500 replications are reported in the parentheses.

Figure 2: Predicted Impact of the Subsidy on the Migration Rate

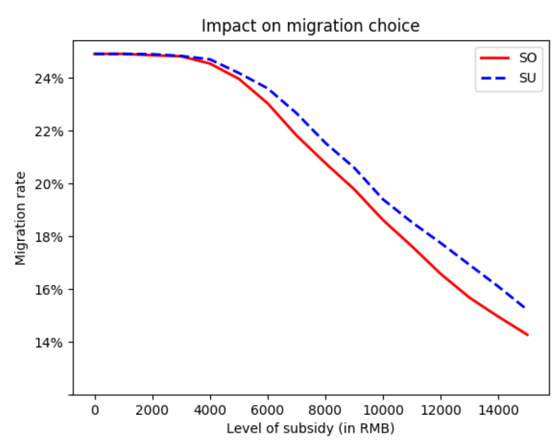

Table 9: Predicted Impact of the Subsidy on Child Outcome (SO)

Subsidy Level	3000 RMB	6000 RMB	9000 RMB	12000 RMB
Word Test Score	0.066	0.126^{*}	0.182^{***}	0.222^{***}
	(0.081)	(0.076)	(0.079)	(0.085)
Math Test Score	0.024	0.050	0.083^{*}	0.112^{**}
	(0.054)	(0.056)	(0.046)	(0.059)
Word Test Score St	0.8%	1.7%	2.5%	2.9%
Math Test Score St	0.6%	1.1%	1.9%	2.5%
Graduation Probability	1.0%	2.1%	3.3%	4.2%

[a] 1 USD = 6.77 RMB (Year 2010).
[b] Bootstrap standard errors based on 500 replications are reported in the parentheses.
[c] *St* indicates scores are transformed to the standard deviation change from the mean.

Table 10: Predicted Impact of the Subsidy on Child Outcomes (SU)

Subsidy Level	3000 RMB	6000 RMB	9000 RMB	12000 RMB
Word Test Score	0.024^{*}	0.062	0.144^{*}	0.272^{***}
	(0.015)	(0.065)	(0.089)	(0.121)
Math Test Score	0.002	0.015	0.051^{***}	0.114^{*}
	(0.021)	(0.037)	(0.015)	(0.063)
Word Test Score St	0.4%	0.9%	2.1%	3.8%
Math Test Score St	0.0%	0.5%	1.6%	3.1%
Graduation Probability	0.2%	0.7%	2.2%	4.7%

[a] 1 USD = 6.77 RMB (Year 2010).
[b] Bootstrap standard errors based on 500 replications are reported in the parentheses.
[c] *St* indicates scores are transformed to the standard deviation change from the mean.

Figure 3: Predicted Subsidy Impact on Child Outcomes Low-income Families

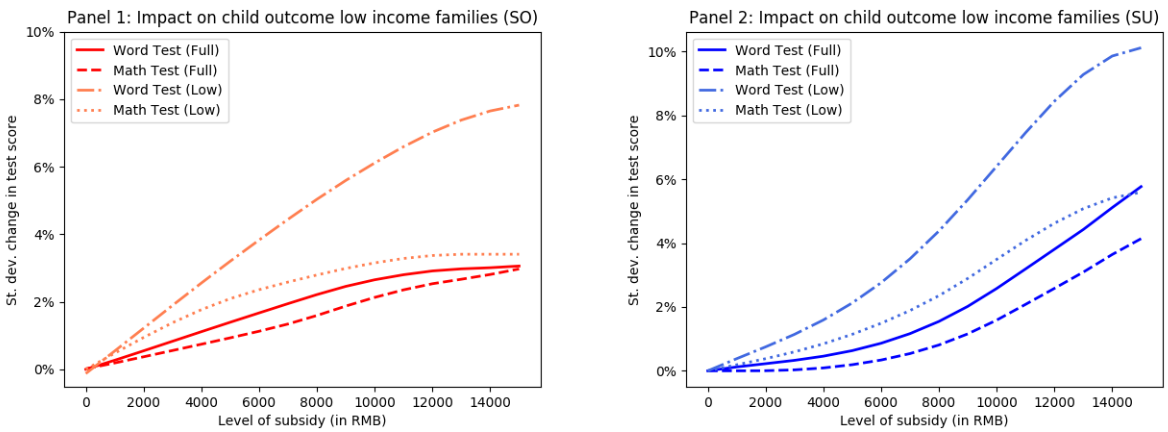

Figure 4: Predicted Impact on Child Outcomes by Age

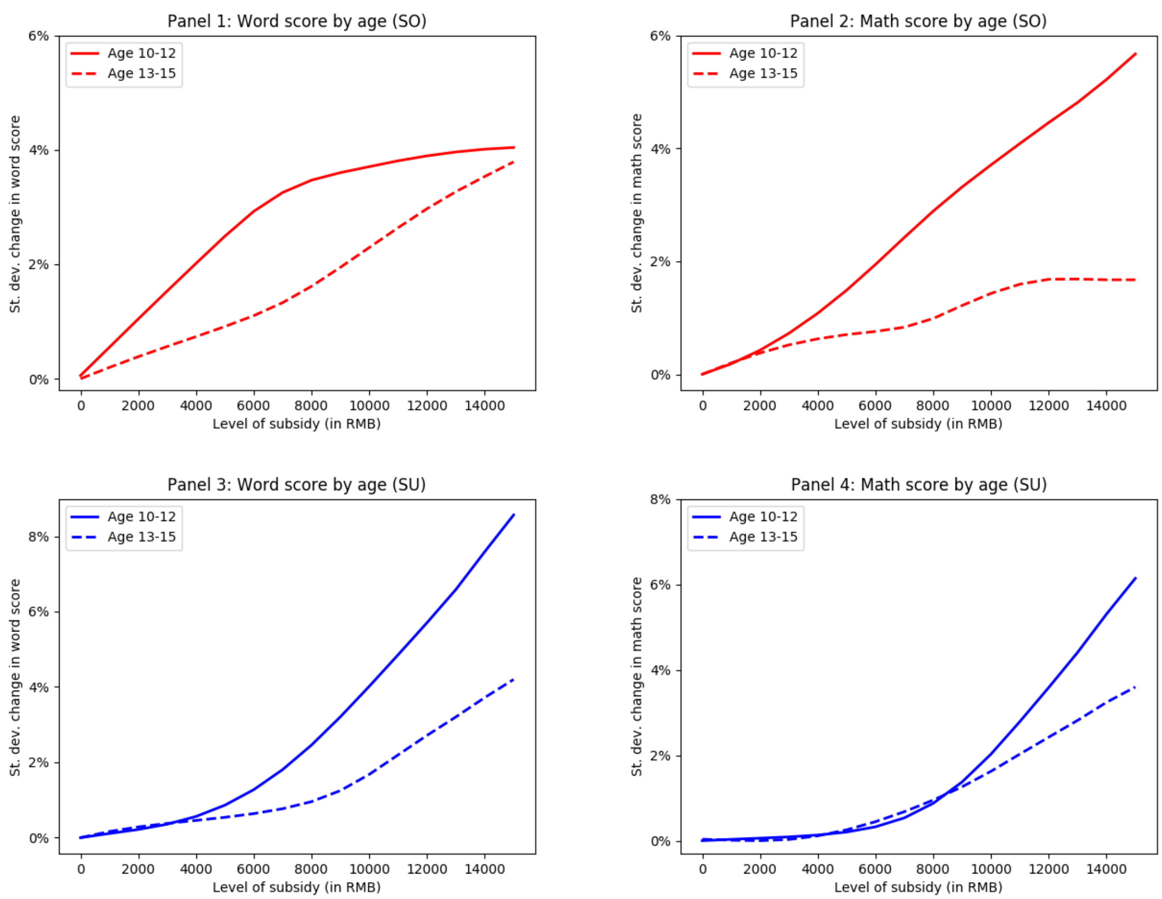

Figure 5: Predicted Impact on Child Outcomes by Gender

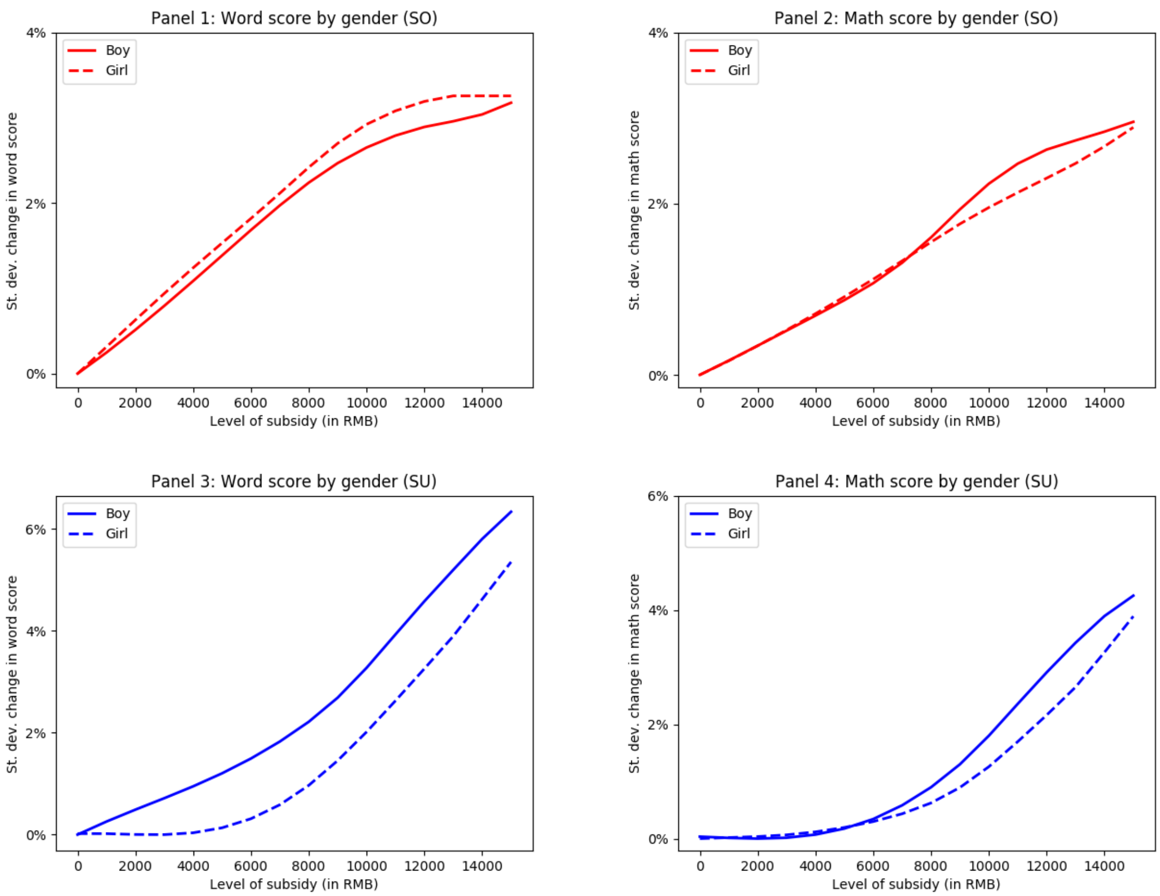

Figure 6: Predicted Impact on Child Outcomes by Number of Children