The Effect of Parental Rural-to-Urban Migration on Children's Cognitive Skill Formation

Bolun Li*

(**Click here for the most recent version**)

December, 2020

Abstract

Large-scale rural-to-urban economic migration in developing countries leaves millions of rural-origin children growing up separated from their migrant parents. Due to the limited parent-child interaction, parental migration poses developmental challenges for left-behind children. This paper develops a structural model of household migration to evaluate the effects of parental migration decisions on the dynamics of children's cognitive skill formation from birth until the end of the developmental stage. I estimate the model using data from the Indonesian Family Life Survey via Simulated Maximum Likelihood. I find that children's cognitive skill formation is sensitive to the duration of parental migration. Using the estimated model, I find that there is a 0.3 standard deviations increase in left-behind children's skills at the end of the developmental stage had their parents not left. I also simulate a series of counterfactual migration policies. I show that migration policies that incentivize family migration with their children to urban destinations are effective in fostering children's cognitive development.

Keywords: migration, left-behind children, cognitive skills, human capital
JEL Classification: O15, J13, J24

*Email: bolun.allen.li@gmail.com
I am deeply indebted to my advisors Ken Wolpin, Flávio Cunha and Rossella Calvi for their constant guidance and support. I owe special thanks to Diego Amador and Andrea Salvati for many helpful discussions and suggestions. I also thank Isaiah Andrews, Manuela Angelucci, Chao Fu, Mehreen Gul, John Kennan, Ajinkya Keskar, Elaine Liu, Ekim Muyan, Mark Rosenzweig, Fan Wang, and faculty members and graduate students at Rice University for their comments and feedback. I am grateful for financial support from the Texas Policy Lab. All remaining errors are my own.

1 Introduction

In many developing countries, an increasing number of children grow up with one or no parents due to parental labor migration (United Nations Children's Fund, 2007). For instance, it is estimated that there are roughly 5 million left-behind children in Indonesia, 6 million in the Philippines, and more than 30 million in China (Bryant, 2005; Xu & Xie, 2015). These numbers imply that approximately 8 % of Indonesian children, 19 % of Filipino children, and 12 % of Chinese children have experienced parental absence early in their lives due to parental migration. The majority of migrant parents rely on extended family members or the one parent who remains behind to take care of the left-behind children. In rare cases, the left-behind children are looked after at boarding schools (Meng & Yamauchi, 2017). Due to the limited parent-child interaction, parental migration poses developmental and emotional challenges for left-behind children. Given the importance of child development in predicting later life outcomes (Currie & Thomas, 2001; Currie & Vogl, 2013), the presence of left-behind children in developing countries has received widespread attention from international organizations and national governments (United Nations, 2015).

The purpose of this paper is to understand how parental rural-to-urban migration affects the dynamics of children's cognitive skill development. The formation of early human capital depends on the entire history of parental investments from birth until the end of the developmental stage (Todd & Wolpin, 2003, 2007; Cunha & Heckman, 2007; Cunha et al., 2010; Heckman & Mosso, 2014). These investments in children, including time and material inputs, in turn, depend on the migration decisions of parents because time inputs rely on the presence of parents, and material inputs are a function of household income. To fully capture the cumulative exposure to parental migration, I develop and estimate a dynamic model of cognitive skill formation nested within household migration decisions. I use this structural framework to 1) quantify the effect of parental migration on children, and 2) evaluate the impacts of counterfactual migration policies intended to foster children's cognitive skill formation.

The analysis begins by describing a dynamic model of skill formation and household migration from the birth of the firstborn child until the end of the developmental stage when the firstborn child turns 14. Each year, rural parents choose whether to stay in a rural location with their child, migrate alone and leave their child behind in their home village, or move together with their child to an urban destination. Because parents care about consumption and the cognitive skill of their child, they face the following trade-off when making migration decisions: if parents migrate alone and leave their children behind in rural areas, the amount of parent-child interaction is reduced, but the parents are likely to earn higher incomes in urban locations; on the other hand, if rural parents migrate with their children to urban locations, they encounter migration costs, despite the incremental parental presence and improved household economic conditions. The arrival of additional children, which evolves according to an exogenous stochastic process, affects household migration decisions and the skill formation of the oldest child. Given the trade-off, a budget constraint, and a technology constraint on their child's cognitive skill, parents make sequentially optimal migration decision each year to maximize their discounted expected lifetime utility, taking into account the current and future returns to their childs human capital, which is endogenously accumulated through cumulative migration experiences. Finally, the unobserved heterogeneities in preferences, in income formation, and in children's cognitive skills allow parents to adjust their migration decisions and thus compensate in the production technology of their child.

Having solved the dynamic migration model, I estimate the structural parameters using micro-level data from the Indonesian Family Life Survey via simulated maximum likelihood. The Indonesian Family Life Survey collects retrospective and longitudinal information on parental migration decisions, which provides a transparent mapping from the data to the dynamic migration model. Intuitively, a full history of migration experiences is necessary to estimate the marginal productivity of inputs to the production technology of skill formation. In addition, the Indonesian data provide a systematic measure of children's cognitive skills using the Raven's Progressive Matrices Test, a commonly used tool to examine fluid intelligence that is related to a number of important skills such as learning and comprehension (Raven, 2000; Unsworth et al., 2014). The main challenge to the identification of the model is that children's skills are endogenously formed through migration experiences, which are correlated with the skills via unobserved heterogeneity. To address this issue, I exploit two sources of variation as exclusion restrictions. The first is the point-to-point distance between one's home village location and provincial capital cities, which captures the potential cost of moving and the need to move. Because distance affects parents' moving decisions regardless of whether they bring their child, it identifies the effects of migration from the effect of non-migration on children's cognitive skills.[1] The second is the ratio of the number of schools divided by population in one's home village to its counterpart in provincial capitals, which captures the notion of school capacity constraints. Because a low ratio provides incentives for parents to move with their children, the school ratio identifies the effect of moving with parents to an urban location from the effect of being left behind in a rural location on children's skills.

The estimated model reveals that children's cognitive skill formation is sensitive to the duration of parental migration. In comparison to both parents staying with their child in a rural location, leaving a child behind for one year during the developmental stages of childhood, i.e., from birth to age 14, reduces the child's cognitive skill by 0.02 standard deviations per year; whereas moving with parents to an urban location is associated with a 0.03 standard deviations improvement in cognitive skill per year. To quantify the adverse effect of leaving children behind, I implement a counterfactual scenario in which it is not feasible for parents to migrate alone. The counterfactual analysis shows that left-behind children would have been better off in terms of their cognitive skill development if their families had remained together. Specifically, there is a positive shift in the distribution of cognitive skills, of approximately 0.30 standard deviations, suggesting that parent-child interaction plays a critical role in the promotion of cognitive development, as discussed in the literature (Heckman & Mosso, 2014). I further measure the improvement in skills on the basis of unobserved household heterogeneity. Left-behind children with unobserved high skill endowment benefit the most when leaving children behind is prohibited and enforced perfectly, in part because their parents are more likely to migrate alone. On

[1] Distance has been widely used as instruments to estimate the return to schooling and education (Card, 1995, 2001).

the other hand, parents of low-skilled children compensate in the form of increased time investments by remaining with their child either in rural locations.

Although quantifying the effect of restricting parents from leaving children behind is interesting as a thought experiment, restricting rural-to-urban movements is not a feasible policy in practice. Therefore, I assess the impacts of a variety of implementable migration policies in developing countries, including relaxing migration constraints and providing migration subsidies. The welfare effects of encouraging rural-to-urban movement, such as consumption patterns and labor productivity, have been widely analyzed in the literature (Bryan et al., 2014; Bryan & Morten, 2019; Lagakos et al., 2018). Unlike previous research, this paper focuses on the welfare analysis of fostering children's cognitive development.

I start by analyzing the effects of cash transfer programs. An unconditional cash transfer program, which subsidizes households regardless of their migration decisions, despite being expensive, has little impact on children. In contrast, a migration subsidy, a popular method to encourage rural-to-urban movement in developing countries (Bryan et al., 2014; Lagakos et al., 2018), has a sizable positive effect on children's skill formation. A migration subsidy of $150 per year if parents migrate with their child, approximately 14% of average annual household income, increases cognitive skills by 0.14 standard deviations. However, the subsidy-induced improvement in skills does not offset the adverse effect of being left behind: a higher subsidy of $200 per year only reduces the total fraction of left-behind children from 12% to 10%. Therefore, I also consider a migration tax to discourage parents from leaving their children behind and hence, to improve the skill development of these children through incremental parent-child interaction. Taxing parents $150 per year sharply reduces the fraction of left-behind children from 12% to 6%, leading to a moderate increase in skills by 0.07 standard deviations. Lastly, motivated by the policy debate over whether to remove the household registration system in China, I estimate an increase of 0.28 standard deviations in average cognitive skills if the migration cost for parents to migrate together with their child is reduced by 25%.[2] This experiment provides an esti-

[2]The household registration system in China was introduced to limit rural-to-urban movement. The restrictions prevent rural workers from staying in urban areas for long and from bringing their families with them.

mate for policy-makers on what to expect in terms of welfare gain among children as the Chinese government recently issued a policy proposal that China will relax its household registration limits in small and medium-sized cities (National Development and Reform Commission of China, 2019).

The principal contribution of this paper is to understand the cognitive development of left-behind children and its relation to parental migration over the entire childhood cycle. A substantial body of research in economics and migration studies has documented the impact of parental migration on children's well-being in a variety of national contexts, focusing on the dimensions of educational attainment (Antman, 2012; McKenzie & Rapoport, 2011), time allocation (Antman, 2011), cognitive achievement (Lu, 2014; Zhang et al., 2014; Xu & Xie, 2015; Bai et al., 2018), and psychological wellbeing (Graham & Jordan, 2011). The mixed findings from the previous research are due not only to the use of a wide range of specifications and identification strategies but more so the common limitation of not accounting for the duration of parental migration, which may lead to biased and inconsistent estimates. However, one exception by Meng & Yamauchi (2017) estimates the effect of lifetime exposure to parental migration on children's educational outcomes using weather shocks and distances as instruments. I build on previous research by providing a coherent understanding of how parental rural-to-urban migration is related to the determinants and dynamics of cognitive skill formation. This goal is achieved by estimating the cognitive production technology jointly with household migration decisions over the entire developmental stages of childhood in a standard dynamic discrete choice model setting (Wolpin, 1984; Rust, 1987; Kennan & Walker, 2011; Gemici, 2011). This paper is closely related to Liu et al. (2010), which estimates a model of maternal employment and location decisions and their impacts on child development using data from the National Longitudinal Study of Youth. This paper differs from Liu et al. (2010) in that it focuses on left-behind children as a result of parental migration, which is not the case in the United States, and in that households are assumed to be forward-looking in a dynamic setting rather than myopic as in Liu et al. (2010) due to computational intractability resulting from a large state space in their model.

This paper is also related to the literature on the cognitive skill formation of children and associated policy implications in developing countries (Glewwe, 2002). Estimating the skill formation of children has been an active area of research. Many papers have relied on data from the United States (Todd & Wolpin, 2003, 2007; Cunha & Heckman, 2007; Bernal, 2008; Cunha et al., 2010; Del Boca et al., 2013; Agostinelli, 2018; Griffen, 2019) with a few exceptions (Attanasio, Cattan, et al., 2015; Attanasio, Meghir, & Nix, 2015; Attanasio et al., 2017). By exploiting data from Indonesia, one of the world's most populous developing countries, this paper estimates the welfare effects of migration policies on children's cognitive skills through counterfactual simulation, which would not be possible if one were to estimate the cognitive production technology function alone. Evaluating migration policies ex ante is important because it provides information on how much the programs would cost and thereby makes it possible to avoid implementing costly programs that are ineffective (Todd & Wolpin, 2008), such as the unconditional cash transfer program discussed in this paper. In a companion paper, Li (2019) proposes a nonparametric matching estimator to directly evaluate the impacts of a different set of migration policies due to the unique institutional constraint in China. Li (2019) finds that a non-migration subsidy raises the probability of graduation by 8.6 percentage points for left-behind children from low-income households in China. This paper extends Li (2019) by estimating a fully dynamic structural model to better understand the mechanism of migration and children's outcomes.

The remainder of this paper is organized as follows. Section 2 introduces the dynamic migration model. Section 3 describes the Indonesian data used for structural estimation. Section 4 discusses the model identification and the simulated maximum likelihood estimation method. Section 5 shows parameter estimates and model fits. Section 6 presents results from counterfactual policy simulations. Section 7 concludes the paper.

2 Economic Model

This section presents a dynamic model of a rural household's migration decisions and illustrates how these migration choices affect the cognitive skill development of their child. A rural household is defined as one with its first child born in a rural location. Parents from a rural household make sequential annual migration decisions at each year t starting from the birth of the firstborn child, $t = 1$, until the child turns 14 years old, $T = 15$. The model is restricted to a married couple in a unitary household setting. I assume that fertility, and therefore the arrival of additional children is exogenous but evolves stochastically based on other state variables. The model considers the firstborn child because doing so allows me to track the arrival of additional children and how the stock of children affects the migration decision as well as the cognitive skill formation of the firstborn child.[3]

2.1 Choice

In each period, rural parents make three mutually exclusive migration decisions: stay in the village with their child, migrate alone to an urban location and leave their child behind in the home village, or migrate to an urban location together with their child. The migration decision considered here can be thought of as a combination of a change in location and change in household composition. The location decision is made between rural areas and urban regions, and household composition is determined by whether parents are present with their child.[4] Formally, the choice set is defined as $J \equiv \{j_t \in$

[3]The additional restriction comes from the data because they contain information on up to 2 randomly selected children per household.

[4]I exclude the situations in which a child migrates with one parent to an urban area and when a child migrates to the urban area alone because these cases account for less than 4% of the entire sample. There are also cases in which parents migrate to a different village. To keep the choice set simple, I collapse these decisions with those when parents migrate to the city for work because both decisions are considered economic migration. I do not differentiate seasonal migration from investment migration, as seen in Kleemans (2015). In addition, I combine the cases when one parent migrates and when both parents migrate into one because the majority of households that migrate without their child consists of a migrant father, a non-migrant mother and a left-behind child. This type of household composition accounts for more than 75% of the left-behind children. Households with migrant mothers are rare, 3%, and the remaining 22% left-behind children have both parents migrated. In addition, only 15%, measured cross-sectionally, of the parents leave their child behind after combing these two cases. Differentiating between them means fewer people in each category. Insufficient variations in choices might cause imprecise estimates. A failure to combine cases as described above would result in as many as 18 available migration alternatives.

$\{1,2,3\}\}$, where

$$j_t = \begin{cases} 1, & \text{if both parents stay with the child in a rural location} \\ 2, & \text{if at least one parent migrates and child is left behind in a rural location} \\ 3, & \text{if both parents move with the child to a urban location.} \end{cases}$$

Because parents can transition between the migration alternatives described above, the model allows for return migration and circular migration between home villages and urban destinations. For instance, if parents leave their child behind in period t and return to their home location in period $t+1$, they simply transition from migrating alone to staying in the village with their child.

2.2 Utility

A household derives utility in period t from consumption C_t, the oldest child's cognitive skill Q_t, the migration decision j_t, observed state variables, unobserved heterogeneity, and a preference shock ε_t. The utility function in period t is given by:

$$\begin{aligned} U_t = {} & C_t + \alpha_{2c}\mathbb{1}\{j_t = 2\}C_t + \alpha_{3c}\mathbb{1}\{j_t = 3\}C_t \\ & + Q_t + \alpha_{2q}\mathbb{1}\{j_t = 2\}Q_t + \alpha_{3q}\mathbb{1}\{j_t = 3\}Q_t + \alpha_{cq}C_tQ_t \\ & + \alpha_{21}\mathbb{1}\{j_t = 2\}\mathbb{1}\{j_{t-1} \neq 2\} \\ & + \alpha_{31}\mathbb{1}\{j_t = 3\}\mathbb{1}\{j_{t-1} \neq 3\} \\ & + \mathbb{1}\{j_t = 2\}\big(\alpha_{22}\, age + \alpha_{23}\, age^2 + \alpha_{24}\, relative + \alpha_{25}\, school\ ratio\big) \\ & + \mathbb{1}\{j_t = 3\}\big(\alpha_{32}\, age + \alpha_{33}\, age^2 + \alpha_{34}\, relative + \alpha_{35}\, school\ ratio\big) \\ & + \mathbb{1}\{j_t = 2\}\textstyle\sum_{k\in K}\alpha_{2k}\mathbb{1}\{\text{type} = k\} \\ & + \mathbb{1}\{j_t = 3\}\textstyle\sum_{k\in K}\alpha_{3k}\mathbb{1}\{\text{type} = k\} \\ & + \mathbb{1}\{j_t = 1\}\varepsilon_{1t} + \mathbb{1}\{j_t = 2\}\varepsilon_{2t} + \mathbb{1}\{j_t = 3\}\varepsilon_{3t}. \end{aligned} \tag{1}$$

The utility function is linear in consumption and a child's cognitive skill.[5] The interaction term between consumption and child quality allows the marginal utility of consumption to depend on a child's cognitive skill. Consumption interacts with migration decisions to capture the notion of location-dependent amenities. A child's cognitive skill interacts with migration decisions to allow parents to care for their child differently due to changes in household composition. The parameters α_{21} and α_{31} denote the transition cost associated with switching between locations. I allow the migration preference to vary across time through parameters α_{22} and α_{32} because parents might prefer to move alone when their child is young or to move together as their child grows older. The parameters α_{24} and α_{34} capture the marginal utility of whether other relatives such as grandparents are living with the household because the existence of relatives provides parents additional reasons to leave their child behind. The variable *school ratio* is the ratio of the number of schools weighted by population in one's home village to the counterpart in the provincial capital city. A lower ratio indicates that there are not enough schools in one's home village, which provides incentives for parents to bring their child if they decide to migrate.

In addition to observed state variables, a household's utility depends on unobserved heterogeneous characteristics that are introduced through household type k. For instance, some parents are more capable of earning income than others, whereas some children are endowed with higher skills than others. I allow four types of household, i.e., $K = \{k_1, k_2, k_3, k_4\}$. The household types also interact with migration decisions. Parents with a sick child may decide to compensate their child by spending more time with the child, in which case the parents do not leave their child behind. To close the model, preference shock $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t})$ is assumed to be serially uncorrelated and follow a joint distribution f_ϵ.[6]

[5]The restrictive assumption that the child's skill enters the utility function linearly and is additively separable is made to simplify a household's utility maximization problem because now the optimization is parallel in the presence of additional children since fertility is exogenous in the model.

[6]The exact specification of the utility function is determined in part to fit the data.

Consumption C_t satisfies the following budget constraint:

$$C_t = \Big(\mathbb{1}\{j_t = 1\} Y_{1t} + \mathbb{1}\{j_t = 2\} Y_{2t} + \mathbb{1}\{j_t = 3\} Y_{3t} \Big) \\ - \Big(\Delta_1 \mathbb{1}\{j_t = 2, 3\} D + \Delta_2 \mathbb{1}\{j_t = 2\} N_t + \Delta_3 \mathbb{1}\{j_t = 3\} N_t \Big), \tag{2}$$

where Y_{jt} denotes household income at period t, D is the distance between a household's home village and the capital city of the province where the household resides, and N_t is the number of children at period t. Unlike the distance to the actual destination, the distance D is exogenous to the model and is included in the initial condition that a household faces. The exogeneity in distance relies on the assumption that residential location prior to the birth of the oldest child is not made with respect to subsequent child development. The parameter Δ_1 is the potential monetary cost if migration takes place because the farther the village is from the provincial capital city, the higher the cost of migration might be due to long transportation time. I allow the migration cost to differ on the basis of household because a family with many children might incur additional costs of migration.[7]

2.3 Income

A household receives a stochastic location-dependent income. Income formation is assumed to be a function of parental education *educ*, time-invariant household type k, and time-varying income shocks η_{jt}. The parameters are indexed by migration choice j to capture income differentials between rural and urban areas. For instance, the returns to parental education and unobserved productivity levels are likely to be different among locations. Formally, the income process is given by:

$$\ln Y_{jt} = \beta_{j1}\, educ_f + \beta_{j2}\, educ_m + \textstyle\sum_{k \in K} \beta_{jk} \mathbb{1}\{\text{type} = k\} + \eta_{jt}, \tag{3}$$

[7] There is no additional cost of fertility as this cost is not seperately identified from parameters Δ_2 and Δ_3.

where the income shock $\eta_t \equiv (\eta_{1t}, \eta_{2t}, \eta_{3t})$ is assumed to follow a joint distribution f_η, which is assumed to be independent of the distribution of preference shocks.

2.4 Child's Cognitive Skill Formation Process

A household gains utility from its oldest child's cognitive ability, which is observed by parents. The cognitive skill production technology depends on a household's migration experiences H, observed characteristics such as parental eduction $educ$, child age age and number of children N, and unobserved components captured by household type k. The cumulative migration experience is defined as $H_{jt} = \sum_{\tau=1}^{t-1} \mathbb{1}\{j_\tau = j\}$ for each migration choice $j_\tau \in J$. The skill formation function adopts a conventional representation of the standard Mincer human capital earnings function. Formally, the production technology is as follows:

$$\begin{aligned} Q_t = {} & \delta_1\, age + \delta_2\, age^2 + \delta_3\, gender + \delta_4\, educ_f + \delta_5\, educ_m + \delta_6 N_t \\ & + \delta_7 H_{2t} + \delta_8 H_{3t} + \delta_9 H_{2t}^2 + \delta_{10} H_{3t}^2 + \textstyle\sum_{k \in K} \delta_k \mathbb{1}\{\text{type} = k\} + \omega_t, \end{aligned} \tag{4}$$

where ω_t is the stochastic component. Equation (4) has several features and restrictions. First, parental migration status serves as a proxy for parental investments because time investments depend on the presence of parents, and material investments depend on income, which in turn depends on migration choice. The effect of migration is a net effect of time investments and material inputs.[8] Second, cumulative migration experience matters instead of their timing. The quadratic terms in migration experience allow the marginal productivities of migration to vary based on the cumulative history of migration. For instance, an additional year of being left behind might have a less adverse impact on a child

[8] Some papers have used household income as a proxy for material inputs under the assumption that parents spend a fixed proportion of their income on children. The production technology here has not included household income because the empirical literature suggests that transitory fluctuations in parental income do not have a substantial effect on promoting children's skill formation, especially when parental education is controlled for (Carneiro & Heckman, 2002; Bernal, 2008; Heckman & Mosso, 2014). However, material inputs, captured by household income, might have a different and potentially bigger impact on children's brain and hence, cognitive development in developing countries than developed countries due to the concern of insufficient nutrition. Including cumulative income history requires tracking the complete sequences of migration alternatives, i.e., income shocks in each period up to period t. Estimating the model with this additional feature becomes computationally challenging, but it is a possible and feasible extension of the current paper.

whose parents moved away a long time ago than a child whose parents recently moved. [9] Third, the inclusion of the stock of children N_t captures the resource allocation among children in a parsimonious way because fertility is exogenous to the model. Fourth, the unobserved heterogeneity in the production technology is assumed to have a constant effect on skill development. The unobserved heterogeneity can arise from differences in parenting skills as well as children's cognitive skill endowments. However, the model does not distinguish which channels the unobserved heterogeneity is attributed to.

2.5 State Space, Initial Condition & Law of Motion

The solution of the dynamic optimization problem requires numerically solving for the value function at each point in the state space. The state space Ω_t at period t consists of all determinants of the household's decision known to the household at time t. Therefore, the state space is defined as $\Omega_t = \{educ_f, educ_f, age, gender, relative, school\ ratio, D, N_t, j_{t-1}, H_{jt}, k, \eta_t, \omega_t, \varepsilon_t\}$. The model starts at the time when the oldest child is born. At this moment, households differ in terms of their initial conditions, including parental education, the existence of relatives, distance to provincial capitals and school ratios, all of which are taken as given and assumed to be fixed over the course of child development.[10] Households also differ in their initial cumulative migration experience H_{j0}, which is assumed to be zero because prenatal migration spells are unlikely to affect a child's cognitive development. The endogenous time-varying migration experiences evolve according to:

$$H_{jt} = H_{jt-1} + \mathbb{1}\{j_{t-1} = j\}, \quad j \in J. \tag{5}$$

[9] Although investments during early childhood are more likely to have differential impacts than later ones, introducing time cutoffs poses challenges to model identification because the econometrician needs two exclusion restrictions per age cutoff to separately identify the effect of moving at different age points. For instance, one would need time-varying exogenous variations to separately identify the effect of leaving children behind when they are young versus the effect leaving children behind when they are old, and these effects need to be separately identified from the effects of moving with their parents as well as the effects of not moving at all.

[10] The initial conditions are assumed to be fixed over time for computational reasons, which is standard in discrete choice dynamic programming models. Otherwise, when solving the dynamic programming program, one needs to compute the expectation with respect to these state variables according to their joint distribution at each point in time in addition to shocks and transition probabilities. As a validation exercise, I check how some of these state variables change over time. For example, roughly 89% of the grandparents who live in the household remain alive during the model period.

The exogenous time-varying state variable, stock of children, evolves stochastically based on other state variables such as lagged stock of children. Let n_t denote fertility such that $n_t = 1$ if giving birth and $n_t = 0$ otherwise, I specify the fertility transition as a function of stock of children in the previous period N_{t-1} according to:

$$
\begin{aligned}
N_t &= N_{t-1} + n_t \\
\Pr(n_t \mid N_{t-1}) &= \begin{cases} \frac{\exp(\gamma_0+\gamma_1 N_{t-1})}{1+\exp(\gamma_0+\gamma_1 N_{t-1})} & \text{if } n_t = 1 \\ \frac{1}{1+\exp(\gamma_0+\gamma_1 N_{t-1})} & \text{if } n_t = 0. \end{cases}
\end{aligned} \tag{6}
$$

Equation (6) implicitly assumes the observed stochastic state variable is independent of past preference shocks conditional on past observations of the stochastic state variable.[11]

2.6 Dynamic Problem

Given the household income and cognitive skill of the child, a household chooses sequential optimal migration alternatives to maximize its expected discounted utility over T periods. Let ρ denote the discount factor, the household problem is:

$$
\max_{\{j_t \in J\}_{t=0}^{T}} E\left[\sum_{t=0}^{T} \rho^t U_{jt} \mid \Omega_t\right] \tag{7}
$$

subject to the budget constraint, the income process, the child's cognitive skill formation, the transition probability distribution of state variables and the joint distribution of all the unobserved components.[12] Within each period, the timing of events is as follows. Fertility, i.e., the stock of children, is realized first. Preference, income, and skill shocks are realized next. Parents make migration decisions thereafter. When making migration decisions, parents take into account future returns to the human capital of their child,

[11]This is similar to the conditional independence assumption seen in Rust (1987). The joint distribution of observed and unobserved stochastic state variables given their past can be factored into two densities, with one being the density of preference shock independent of past preference shocks given observed state variables in the same period and the other being the density of current observed state variables independent of current and past preference shocks given the last period's observed state variables.

[12]I assume the discount factor $\rho = 0.95$ in the structural estimation below. The discount factor is identified through the difference in the future component of the expected value functions. See a detailed discussion in Keane et al. (2011).

which accumulates endogenously through migration experiences. By Bellman's principle of optimality, the value function $V(\Omega_t)$ can be obtained using the recursive expression:

$$\begin{aligned} V(\Omega_t) &= \max_{j \in J} \Big\{ U_{jt}(\Omega_t) + \rho E[V(S_{t+1}) \mid \Omega_t, j_t] \Big\} \quad \text{for } t < T, \\ &= \max_{j \in J} \Big\{ U_{jT}(\Omega_T) + \alpha_{jqT} \ln Q_{T+1} \Big\} \quad \text{for } t = T, \end{aligned} \tag{8}$$

where the expectation is taken with respect to the joint distribution of future shocks and the transition probability of state variables. I model the child development process as lasting for $T = 15$ periods. Once a child reaches adulthood, the nature of the household's decision problem changes, and I do not model decisions beyond that point. Because the true dynamics of my model depend solely on child's cognitive skill formation, I assume that the terminal value function at time T is the flow utility in the final period plus the "carry over" child quality from the developmental stage with α_{jqT} being a free parameter to estimate.[13]

3 Data and Empirical Evidence

The dynamic nature of parental migration and children's cognitive skill development entails specific data requirements. The econometrician needs to observe a full history of household migration and income beginning at the oldest child's conception until the final stage of child development at age 14. To account for unobserved heterogeneity, the econometrician also needs to observe the cognitive skill measure of a child multiple times during childhood. Although data of this kind are limited in developing countries, the Indonesia Family Life Survey, hereafter IFLS, comes close. The IFLS covers a sample of 7,224 households from across 13 provinces in Indonesia (see Figure A.1). The resulting sample represents 83% of the Indonesian population. The IFLS tracks individuals and households in 5 waves from 1993 to 2014, and major efforts to reinterview all respondents result in a remarkable success rate of over 90%. Retrospective information on migration and in-

[13]This assumption is commonly made in the literature, see, for instance, Bernal (2008) and Del Boca et al. (2013).

come histories is collected for adults, and cognitive measures are collected for children. I use a subsample of 795 rural households for empirical analysis. The sample selection procedure and additional information about the dataset are discussed in Appendix A.1.

3.1 Descriptive Statistics

Table 1 presents sample descriptive statistics. The average schooling is between grade school and middle school for both parents. The average annual income of rural households is \$1,069.[14] The average annual income of a rural migrant household is 40% higher than a rural nonmigrant household. Roughly half of the households have relatives living with them at the birth of their firstborn child. The villages are generally distant from the capital cities of their provinces, with an average distance of 60.35 miles. The majority of the households do not migrate, which is consistent with findings in other developing countries (Bryan et al., 2014). Specifically, 57% have never moved, 11% have moved together to urban areas at least once, and 36% have at least one parent who has moved at least once. In terms of cumulative migration experience, the average number of years is 5 years for both moving with and without a child, which is conditional on a household having moved at least once. Figure A.4 depicts that for households that migrate, the majority of the parents are likely to leave their firstborn child behind, particularly when this child is young; few households move together with their child to urban areas, and they tend to do so when this child is older. Table 5 shows one-period transition rates for migration from choice in $t-1$ to the choice in t. The transition matrix provides strong evidence on the persistence of each choice, suggesting potential transition costs from switching locations.

The IFLS provides cognitive measures of youth aged 7-14 years. The cognitive test consists of 12 questions drawn from the Raven's Colored Progressive Matrices test and 5 mathematical problems.[15] The Raven's Colored Progressive Matrices test assessment is

[14] US dollar value of the year 2000.

[15] In IFLS, there are no final or composite scores for the cognitive test items. Due to time constraints, the IFLS team reduced the number of questions but still allowed for sufficient scope in difficulty, so questions were selected to run a range from simple to harder. The IFLS team conducted pretesting to determine the questions selected, but there are no published results for that pretesting. Because there is no formal guideline, I could not assign a weight to each question in terms of difficulty levels. For

commonly used as a measure of general intelligence, especially fluid intelligence which is related to comprehension skill and learning ability (Raven, 2000; Unsworth et al., 2014). Fluid intelligence refers to the ability to reason and analyze, for example solving math problems or puzzles.[16] The Raven's Colored Progressive Matrices test consists of pattern-matching exercises in which the respondent is asked to identify the missing piece that best matches the shown patterns (see an example in Figure A.2). Because the test does not depend heavily on verbal skills, it is considered relatively culture-free. Figure 1 depicts the raw test score distribution, with the raw score being the total number of correctly answered questions. On average, children answer 10 questions correctly. Older children answer more questions correctly than younger children. However, the ceiling effect is strong among older children as the score distribution is skewed to the right. Therefore, the naive raw measures are sufficient to distinguish low-ability children from moderate-ability children but not sufficient to distinguish high-ability children from moderate-ability children. For instance, for two children who both answer 11 questions correctly, one child might answer simpler questions correctly whereas the other child answers more difficult questions correctly. The raw score measure is not able to identify the differentials in their skills because each item is weighted equally.

To obtain a consistent measure of latent cognitive skills, I map the test responses of sampled children to their latent cognitive skills using a Two-parameter Logistic model from Item Response Theory developed in psychology (Lord, 2012). The basic idea of the Item Response Theory models is that the probability of correctly answering a question from a test is a function of test characteristics such as the difficulty levels of each question as well as a test taker's latent skills. A detailed discussion of implementing and estimating Item Response Theory models is provided in Appendix A.2. I recover the expected latent skill for each child using the empirical likelihood and the skill density via empirical Bayesian updating after estimating the Two-parameter Logistic model. Figure 2 graphs

additional information about the dataset, visit https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS/datanotes.html

[16]The traditional view on fluid intelligence is that it cannot be learned as opposed to crystallized intelligence. However, recent studies have shown that fluid intelligence can be improved through training and social interactions (Jaeggi et al., 2008). One study using a randomized control trial indicates that parent-child interaction is important in advancing fluid intelligence in young children(Tachibana et al., 2012)

the latent cognitive skills of the sampled children. As with raw scores, children's cognitive skills grow with age. However, the ceiling effect slowly disappears because the transformation stretches the higher end of the skill distribution. In fact, the latent cognitive skills demonstrate sufficient variations and disparities. The skill distributions are bell-shaped due to the underlying assumption in Item Response Theory models that latent skills are normally distributed.

3.2 Evidence Linking Migration and Child Development

To document the correlation between children's cognitive test scores and parental migration, I regress the estimated latent cognitive skills on migration. To measure the contemporaneous effect, I use the lagged migration status, which is the standard specification in the literature. To measure cumulative exposure, I use the total number of periods for each possible migration alternative divided by the total number of periods so that the coefficients provide a similar interpretation as the contemporaneous case. Table 2 reports correlation coefficients controlling for household demographics. I compute robust standard errors clustered at the village level. In comparison to the case where both parents stay with their child in rural areas, leaving a child behind is correlated with a 0.16 standard deviations reduction in cognitive skills in the contemporaneous specification; whereas moving with a child to urban destinations is correlated with an improvement in skills by 0.28 standard deviations. The correlation coefficients change once cumulative migration duration is taken into consideration, suggesting that ignoring cumulative migration experience might not fully capture the association between parental migration and skill development. Recall that 36% of parents leave their child behind, despite that parental absence is negatively related to children's cognitive development. On the other hand, very few parents, 12% of the sample, move together with their child to urban destinations even though this decision leads to potential improvements in children's skills. Therefore, it is essential to understand what factors drive these observations and households' decisions, and what kinds of policies can improve children's cognitive skills.

4 Estimation

4.1 Identification

The main equation of interest in the model is the production technology of cognitive skill formation. The challenge to identification comes from the fact that cognitive skill accumulation is correlated with migration decisions through unobserved heterogeneity. To identify the differential effects of migration, it requires two sets of exclusion restrictions because the effect of moving, regardless of whether parents moves with a child or not, needs to be separately identified from the effect of staying on the childs cognitive skill. Additionally, the effect of moving with a child needs to be separately identified from the effect of moving without a child, i.e., the effect of being left behind. I exploit the point-to-point distance from one's home village location to the provincial capital to study the effect of moving relative to the effect of staying because distance captures potential migration costs but unlikely to affects children's human capital. The intuition is similar to instrumenting variations in the distance to the nearest college in estimating the returns to schooling (Card, 1995, 2001). I do not use the distance to the actual destination because that distance is endogenous in the model as a result of endogenous migration decisions. To tease out the effect of moving without a child from the effect of moving with a child on cognitive skills, I leverage the ratio of the number of schools divided by population in one's home village to the counterpart in provincial capital cities. This instrument captures capacity constraints in school enrollment between the home location and provincial cities. A low ratio indicates an insufficient number of schools in one's home village and hence provides incentives for parents to bring their child when they migrate.

The validity of the instruments depends on 1) whether they are relevant and 2) whether they are exogenous. The first condition requires that instruments are closely related to migration decisions. Figure 3 graphs the variation of the two instruments to demonstrate the identifying sources of the variation. I formally test the relevance of the two sets of instruments since weak instruments can produce biased and inconsistent estimates. Given that the model has multiple instruments and endogenous variables, I report the joint tests for under identification and weak instruments in panel A in Table 3. The Kleibergen-Paap

LM statistic for under identification test rejects the null that the coefficients of the instruments from the 1st stage are jointly zero (Kleibergen & Paap, 2006). The Kleibergen-Paap F statistic for the weak instrument test equal to 10.57, suggesting that the instruments are relevant, as the approximate relative bias is between 10% and 15%.[17] The second condition requires that these instruments are uncorrelated with children's skills conditional on other state variables. Unfortunately, this condition is untestable. This condition is violated if, for instance, the closer to the provincial capital a village is, the better the economic conditions of the village, which might lead improvement in children's skills. The correlations between distance and local wealth indicators reported in Panel B in Table 3 provide some suggestive evidence that the distance to provincial capital satisfies the requirement. In addition, I compute the correlation between the number of schools and the subjective measure of school quality by the IFLS team. The insignificant correlation coefficient rules out the concern that school quality might impact children's skills through the number of schools.

As is standard in dynamic discrete choice models, utility parameters are identified through the variation in choices and state variables across individuals and across time. The variation in income level identifies the consumption parameters α_{jc} due to the identifying restriction that income shocks are independent of preference shocks.[18] Functional form and normalizations are made throughout the model to aid identification. The utility parameters associated with the choice of not migrating are normalized to zero because only the differences in utilities are identified. In addition, unobserved types capture permanent heterogeneity across parents and children. If there are two households with identical initial conditions and the same observed characteristics but consistently make different choices, then they are very likely to have different unobserved types. Similar to the

[17] The Cragg-Donald Wald F statistic is computed under the homoskedasticity assumption (Cragg & Donald, 1993), which is unlikely to be the case given the unobserved heterogeneities and other complications from the structural model. Stock & Yogo (2005) provides critical values such that the maximum relative bias with respect to the OLS estimates is no more than a certain percentage. In this case, Stock and Yogo's critical values for the F-stat of the excluded instruments are 13.43 (10%) and 8.18 (15%).

[18] It is possible that income shocks are correlated with preference shocks. In this case, additional exclusion restrictions are needed. For instance, the exogenous variation in labor market demand shocks or weather shocks can serve as instruments to identify the consumption parameter α_{jc} in the utility since they affect income independently.

identification argument of the permanent heterogeneity in panel data, type parameters in the production technology of skills are identified as long as we observe the same child multiple instances over time. Household types are essential to capture the persistence in the transition matrix and improve the overall fit of the model.

4.2 Simulated Maximum Likelihood

Having solved the dynamic optimization problem numerically via backward recursion, I estimate the model by simulated maximum likelihood. The solution to the dynamic programming problem serves as the input when estimating the structural parameters of the model given data on choices, earnings and cognitive skills. Let the indicator function d_{jt} equal 1 if alternative $j \in J$ is chosen at time t and θ denote the vector of model parameters, the contribution to the likelihood of each household conditional on unobserved type k is given by:

$$L_{it}(\theta) = \sum_{j \in J} d_{jt} \Pr(d_{jt} = 1, Y_{jt}, Q_t \mid \Omega_t, k; \theta), \tag{9}$$

where $\Pr(d_{jt} = 1, Y_{jt}, Q_t \mid \Omega_t, k; \theta)$ is the joint density of choice, income, skills, and transition probability. The unconditional likelihood contribution $L_i(\theta)$ for household i is a weighted average over all possible types p and k, weighting by the type proportions μ_p and π_k, which are structural parameters to estimate:

$$L_i(\theta) = \sum_{k \in K} \pi_k \prod_{t=1}^{15} L_{it}(\theta). \tag{10}$$

Taking logs and summing over all observations yields the sample log-likelihood:

$$LL(\theta) = \sum_{i=1}^{N} \log L_i(\theta). \tag{11}$$

Estimation here is an iterative process that involves solving the dynamic program and maximizing the likelihood until it converges. In practice, I make additional assumptions to accelerate the estimation process. Preference shocks are assumed to follow a type I extreme value distribution, income shocks are assumed to follow a joint normal distribution, and the stochastic component in a child's cognitive skill formation process is assumed to be measurement error rather than a real productivity shock. These assumptions do not undermine the identification argument but are made to improve computation. I estimate the standard errors using the outer product of the numerical gradients of the log-likelihood, i.e., the BHHH estimator, since the likelihood function does not have an analytical form.

In constructing the likelihood function, I have thus far assumed that we fully observe household income and children's skills at each point in time. However, the income information suffers a missing data problem because many households do not report their income throughout IFLS1-5, and because income histories are not collected in IFLS4 and IFLS5. Consequently, 48% of the total 11,925 household-period observations do not have income information available. For these households with missing income, I integrate out income whenever unobserved in computing the choice probability. I approximate the integral with Monte Carlo simulation. In addition, because the cognitive tests are administered at the time of each interview, children's test responses and their estimated cognitive skills are only known at that age. In this case, numerical integration with respect to skill is not necessary given the assumption that the only stochastic component in a child's cognitive skill is measurement error. Appendix A.3 presents the details of the simulated maximum likelihood estimation method.

5 Results

5.1 Parameter Estimates

Table A.1 and Table A.2 display parameter estimates and associated standard errors. To provide a clear interpretation of the utility parameters, I compute the cost of migrations in terms of consumption units.[19] The costs of migration include the monetary component from the budget constraint and the psychic part from the utility function. The annual

[19] In the estimation, I rescale consumption and measure it in units of $1,000 year 2000 dollars.

migration cost is \$3,255 on average if parents move with their children and \$2,869 on average if parents move without their children. The cost of joint household migration is 14% higher than the cost of parents moving alone, in part explaining why parents are likely to leave their children behind when they decide to move. The parameters of the cognitive skill function are important in identifying the impact of different counterfactuals on children. Specifically, the productivity of parental inputs, captured by cumulative migration alternatives, indicates that leaving children behind for one year during early childhood reduces skill by 0.02 standard deviations, while migrating with children to urban areas improves cognitive skill by 0.03 standard deviations per year. The oldest child's skill decreases with the number of children, implying that parents might allocate fewer resources in the oldest child with the arrival of additional children to the family.

The unobserved type distribution reveals considerable heterogeneity among households. Roughly 10% of the households are type 1, 43% are type 2, 36% are type 3, and the remaining 11% are type 4. The unobserved heterogeneity in income profile reveals strong correlation with the unobserved component in children's skills: if a household is productive at earning income, it is also productive at child rearing, and vice versa. For instance, the income-earning abilities (child-rearing skills) of type 1 households are 70% (11%) higher than that of type 4 households. Figure 4 plots household migration decisions by household type. In comparison to other types of households, type 3 & 4 households, which are less productive at raising their children, are more likely to remain in rural areas, suggesting that parents choose to compensate their children by increasing parent-child interaction. In contrast, households that are the most productive, type 1 households, at raising children are more likely to leave their children behind. One plausible explanation is that children from this type of household are endowed with higher skills and hence are likely to be independent.

As discussed in Section 1, the literature that estimates the effect of parental migration on children's outcomes has mostly focused on the contemporaneous effect. To demonstrate how different the estimated impacts could be with different definitions of migration, I estimate the cognitive skill production function via OLS using simulated data conditional on household types to correct the endogenous selection.[20] The contemporaneous and cumulative measures are the same as defined previously in Section 3.2. Table 6 presents the results. Household characteristics are stable across specifications. The magnitude of the coefficients for migration experience is 4-5 times that of those from the contemporaneous case, suggesting that papers that use contemporaneous measures for parental migration are likely to underestimate the true effect of parental absence. The findings are consistent with the study by Meng & Yamauchi (2017).[21]

5.2 Model Fit

This section shows that the model is able to fit the data reasonably well in several dimensions. In terms of migration choice, Figure A.4 graphs the migration choice distribution by the age of the oldest child between the model and the data. Table 4 formally presents within-sample χ^2 goodness-of-fit test statistics.[22] Figure A.4 together with Table 4 suggest that the choice distributions generated by the model are quite close to those in the data. Table 5 compares the period-by-period transition matrix. Although the transition rates in the model and the data are similar in magnitude, the χ^2 goodness-of-fit test statistics show that the model is not able to capture the degree of persistence observed in the data. Regarding the effects of migration on children's skills, Figure A.6 plots the mean and standard deviations of skills by discretized cumulative migration status. The estimated model is able to replicate the differential patterns of skill formation on the basis of different migration experiences. Figure A.5 presents evidence on the model fit of cognitive skills by child age. The left panel, comparing the predicted and actual mean of skills, does not match well at some age points due to the small sample size. Figure A.7 and Figure A.8 show that other key moments such as simulated income by household migration choices and parental education levels match data moments quite closely.

[20] Admittedly, all regressions considered here are misspecified given the true data generating process by the structural model. The estimates here are at best approximations of the true model.

[21] Meng & Yamauchi (2017) do not estimate the effect of joint household migration to urban locations on children's skills.

[22] The χ^2 statistics computed here are not adjusted for parameter estimates from the model.

6 Counterfactual and Policy Experiments

In this section, I use the estimated dynamic migration model to explore predicted changes in children's cognitive outcomes and migration patterns through counterfactual scenarios and policy experiments. To do so, I simulate household behavior under a counterfactual scenario where simulated households have the same initial conditions and face the same environment as the baseline. I then compare cognitive skills and choice distributions between the counterfactual scenario and the baseline case. This paper considers two sets of counterfactual experiments. I start by discussing whether left-behind children would be better off if their families had remained together. I then quantify the predicted impacts of a set of implementable or existing migration policies on children in developing countries.

6.1 Are Left-behind Children Worse Off?

To address this question, I implement a counterfactual scenario in which the cost of leaving children behind is raised to a level such that it is never feasible to do so. Under this counterfactual, parental time investments increase because they are forced to be with their children. However, the direction of material inputs remains unclear because parents are able to choose between rural areas with lower income and urban locations with higher income. The overall impact of the two channels is not clear ex ante. Figure 5 provides evidence that children would be better off in terms of cognitive development if their families had remained together. Panel A compares the entire population of children between the baseline and the counterfactual scenarios. There is a gain of 0.14 standard deviations in skills on average if leaving children behind is banned and enforced perfectly. The overall positive impact implies that the time investments dominate material inputs. This is consistent with findings in the literature that parent-child interaction plays a critical role in promoting child learning and cognitive development (Heckman & Mosso, 2014). Panel B shows a greater improvement in cognitive skills, 0.29 standard deviations, if the sample is restricted to left-behind children from the baseline. I explore this improvement by counterfactual migration choices. Of the parents that are observed to leave their children behind in the baseline scenario, 94% choose to remain in rural areas with their children under this counterfactual, resulting in 0.12 standard deviations increase in skills. The remaining 6% decide to migrate with their children to an urban location, leading to a substantial gain of 1.63 standard deviations in the cognitive skills of their children. This suggests that policies of incentivizing parents to move with their children to urban areas offer the most potential gain in advancing children's cognitive development.

To further understand the potential effect of leaving children behind, Figure 6 explores the counterfactual skill distribution by household type. Restricting parents from leaving their children behind improves cognitive skills for children from all types of households. Moreover, children from type 1 households (the most productive type) have the highest gain in skills, 0.19 standard devisions, because they are more likely to be left behind by their parents than children of other types and hence are impacted the most by this counterfactual. Children from type 4 households (the least productive type) also face a sizable gain of 0.16 standard devisions increase in their skills, reinforcing the important roles of parent-child interaction in shaping cognitive development.

Leaving children behind has shown considerable adverse effects on children's skills. However, completely prohibiting parents from leaving their children behind is not feasible in practice. Policy-makers need to consider programs that discourage parents from leaving their children behind to achieve the same goal as migration restriction. As two extreme cases, reducing the number of left-behind children to zero would require subsidizing a household $1,894 per year to remain in the rural areas or $1,315 per year as a family to move to urban regions. Given that these programs are incredibly costly, recall that annual household income is $1,069, I turn to analyze the effects of implementable migration policies intended to foster children's cognitive development in the next section.

6.2 Migration Policies

The policy analysis is centered on estimating the benefits of encouraging rural-to-urban movements, which is in line with the literature on encouraging internal migration in developing countries (Kleemans, 2015; Bryan & Morten, 2019). I start with quantifying the effects of subsidizing households, both unconditionally and conditionally. The unconditional cash transfer is an annual cash transfer to a family regardless of its migration decision. I find that this program has a marginal impact on children's skills. Specifically, it improves cognitive skills by less than 0.03 standard deviations at all subsidy levels. This finding is attributed to the trivial changes in the choice distribution of migration under this counterfactual. Given that unconditional cash transfer programs are expensive and ineffective, I analyze conditional cash transfer programs next. The conditional cash transfer involves a subsidy to a household if parents migrate together with their child to urban destinations. Intuitively, household rural-to-urban migration favors children's cognitive skill development as a result of ensured parental presence and improved economics conditions. Figure 7 illustrates this case. The solid blue line graphs the impact of a constant annual subsidy for all households if parents migrate together with their child. The results indicate that migration subsidies lead to a sizable improvement in children's cognitive achievements. An annual subsidy of $150, approximately 14% of household income, raises children's skills by 0.14 standard deviations on average. To further reduce program costs, recall that left-behind children most from type 1 and type 4 households benefit the most (see Figure 6), I consider a subsidy schedule designed to target these types. Although targeting based on unobserved types is not practical, identifying households with specific observed characteristics is feasible from a policy-maker's perspective. Therefore, I estimate the posterior distribution of household types given choices and initial conditions using the likelihood, the integrand of the likelihood, and the estimated type proportion via an empirical Bayesian method. I then regress the predicted type probabilities of each household on their initial conditions to determine the correlation between type probabilities and household characteristics. I define the target group as households with less educated mothers or female children. The red dashed line in Figure 7 depicts the impact of this program. Targeting is effective because the targeted sample accounts for less than 50% of the population, while the improvement in cognitive skills is approximately 65% of the increase from the previous schedule that subsidize every household conditional on moving with their child.

Table 7 provides explanations for the cognitive skill improvement from migration subsidies. The proportion of parents who migrate with their child increases with the amount of subsidy. However, the subsidy-induced movers primarily come from the nonmigrant households in the baseline rather than the parents of left-behind children. Despite the considerable improvement in skills, the number of left-behind children is only reduced by approximately 2% at the highest level of annual subsidy at $200. To reduce the number of left-behind children, the last policy experiment I consider is to disincentivize parents from leaving their children behind. This policy involves a migration tax that requires parents to pay a fine if they migrate without their children. I simulate the effects for the whole sample and the target groups for the reasons discussed above. As shown in the last column in Table 8, parents of left-behind children are responsive to a migration tax. There is a sharp reduction from 12% to 7% in the proportion of households with parents that leave their children behind and a 0.4% increase in the proportion of households that migrate together to urban locations. As a result of the decline in the number of left-behind children, Figure 8 shows that skills increase moderately by 0.07 standard deviations on average given an annual tax of $150.

The last policy I consider is to relax migration constraints. In recent years, the Chinese government has begun to reform its household registration system, which was introduced to limit rural-to-urban movement. A recent policy proposal states that China is undertaking key measures to relax its household registration system restrictions in small and medium-sized cities to solve migrant workers' residency challenges (National Development and Reform Commission of China, 2019). Because the household registration system imposes institutional barriers and financial constraints on migrant workers and their families, I estimate the effect of this policy proposal by reducing migration cost by 25% through changes in utility parameters and cost parameters in the budget constraint if a household migrates together with their child to an urban location. Recall that the estimated cost of family migration with their child is $3,155; the reduction is equivalent to $789 in monetary value. The simulation shows that there is an average of 0.28 standard deviations increase in children's cognitive skills, which is accompanied by 14% inflow of rural parents and their children to urban destinations.

7 Conclusion

The large-scale parental rural-to-urban internal migration in developing countries has affected millions of rural-origin children who have been left behind by their migrant parents. This paper studies the effect of parental migration on the dynamics of children's cognitive skill formation in Indonesia. To do so, I estimate a dynamic model of household migration embedding a production technology of cognitive skill formation using data from the Indonesia Family Life Survey. The estimation results indicate that being left-behind during childhood has adverse effects on cognitive skill formation. In fact, a counterfactual analysis shows that left behind children would have been better off if their entire household had remained together. In addition, migrating with parents to urban destinations improves children's skills. However, the cost of migration hinders household mobility and child development. Motivated by these findings, the counterfactual exercises show that policies of encouraging household migration are effective in improving children's cognitive skills.

I conclude by considering possible extensions of this paper. School inputs and peer effects, in addition to parental investments, are essential in shaping children's skills, especially during adolescence (Burke & Sass, 2013; Fu & Mehta, 2018; Agostinelli, 2018). Modeling these channels is challenging because it requires modeling migration as a joint search of household income and children's schooling. Estimating this extended model requires comprehensive data on school choices and classroom composition in addition to the stringent data requirement on migration history and cognitive measures. Future work could aim to deepen our understanding of how school investments and social interaction, as a result of migration, affect children's cognitive skill formation in developing countries. Moreover, as noted in in recent literature (Heckman & Mosso, 2014), skills are multi-dimensional, including cognitive, non-cognitive, social-emotional, and behavioral skills. Multiple skills affect performance in life across a variety of dimensions. An interesting avenue of research is to quantify the effect of parental migration on the multiple domains of children's skill formation.

References

Agostinelli, F. (2018). Investing in children's skills: An equilibrium analysis of social interactions and parental investments.

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. (2012). Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics*, *25*(4), 1187–1214.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., & Rubio-Codina, M. (2015). Estimating the production function for human capital: Results from a randomized control trial in Colombia.

Attanasio, O., Meghir, C., & Nix, E. (2015). Human capital development and parental investment in India.

Attanasio, O., Meghir, C., Nix, E., & Salvati, F. (2017). Human capital growth and poverty: Evidence from Ethiopia and Peru. *Review of Economic Dynamics*, *25*, 234–259.

Bai, Y., Zhang, L., Liu, C., Shi, Y., Mo, D., & Rozelle, S. (2018). Effect of parental migration on the academic performance of left behind children in North Western China. *Journal of Development Studies*, *54*(7), 1154–1170.

Bernal, R. (2008). The effect of maternal employment and child care on children's cognitive development. *International Economic Review*, *49*(4), 1173–1209.

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. *Econometrica*, *82*(5), 1671–1748.

Bryan, G., & Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from Indonesia. *Journal of Political Economy*, Forthcoming.

Bryant, J. (2005). Children of international migrants in Indonesia, Thailand and the Philippines.

Burke, M. A., & Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, *31*(1), 51–82.

Card, D. (1995). Using geographic variation in college proximity to estimate the return to schooling. In *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkam* (pp. 201–222). University of Toronto Press.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, *69*(5), 1127–1160.

Carneiro, P., & Heckman, J. J. (2002). The evidence on credit constraints in post-secondary schooling. *Economic Journal*, *112*(482), 705–734.

Cragg, J. G., & Donald, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, *9*(2), 222–240.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31–47.

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

Currie, J., & Thomas, D. (2001). Early test scores, school quality and ses: Longrun effects on wage and employment outcomes. In *Worker Wellbeing in a Changing Labor Market* (pp. 103–132). Emerald Group Publishing Limited.

Currie, J., & Vogl, T. (2013). Early-life health and adult circumstance in developing countries. *Annual Reveiw of Economics*, *5*(1), 1–36.

Del Boca, D., Flinn, C., & Wiswall, M. (2013). Household choices and child development. *Review of Economic Studies*, *81*(1), 137–185.

Fu, C., & Mehta, N. (2018). Ability tracking, school and parental effort, and student achievement: A structural model and estimation. *Journal of Labor Economics*, *36*(4), 923–979.

Gemici, A. (2011). Family migration and labor market outcomes.

Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature*, *40*(2), 436–482.

Graham, E., & 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.

Griffen, A. S. (2019). Evaluating the effects of childcare policies on children's cognitive development and maternal labor supply. *Journal of Human Resources*, *54*(3), 604–655.

Heckman, J. J., & Mosso, S. (2014). The economics of human development and social mobility. *Annual Reveiw of Economics*, *6*(1), 689–733.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829–6833.

Keane, M. P., Todd, P. E., & Wolpin, K. I. (2011). The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications. In *Handbook of labor economics* (Vol. 4, pp. 331–461). Elsevier.

Kennan, J., & Walker, J. R. (2011). The effect of expected income on individual migration decisions. *Econometrica*, *79*(1), 211–251.

Kleemans, M. (2015). Migration choice under risk and liquidity constraints.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126.

Lagakos, D., Mobarak, A. M., & Waugh, M. E. (2018). The welfare effects of encouraging rural-urban migration.

Li, B. (2019). Ex ante evaluation of the effects of migration policy on children's cognitive achievements.

Liu, H., Mroz, T. A., & Van der Klaauw, W. (2010). Maternal employment, migration, and child development. *Journal of Econometrics*, *156*(1), 212–228.

Lord, F. M. (2012). Applications of Item Response Theory to Practical Testing Problems.

Lu, Y. (2014). Parental migration and education of left-behind children: A comparison of two settings. *Journal of Marriage and Family*, *76*(5), 1082–1098.

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 children's education and health outcomes in China. *Demography*, *54*(5), 1677–1714.

National Development and Reform Commission of China. (2019). 2019 Urbanization plan.

Raven, J. (2000). The raven's progressive matrices: change and stability over culture and time. *Cognitive psychology*, *41*(1), 1–48.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica*, 999–1033.

Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. *Identification and Inference for Econometric Models*, 80-108.

Tachibana, Y., Fukushima, A., Saito, H., Yoneyama, S., Ushida, K., Yoneyama, S., & Kawashima, R. (2012). A new mother-child play activity program to decrease parenting stress and improve child cognitive abilities: A cluster randomized controlled trial. *PloS one*, *7*(7), e38238.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, *113*(485), F3–F33.

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.

United Nations. (2015). Transforming our world: the 2030 agenda for sustainable development.

United Nations Children's Fund. (2007). The state of the world's children 2007.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive psychology*, *71*, 1–26.

Wolpin, K. I. (1984). An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political Economy*, *92*(5), 852–874.

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

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.

Table 1: Descriptive Statistics

	Mean	St. Dev.	Num. Obs.
Household characteristics			
Household income ($)	1069.82	2352.29	6,160
Father eduction	2.62	1.04	11,925
Mother eduction	2.46	0.95	11,925
Number of children	1.64	0.85	11,925
Relative	0.44	0.50	11,925
Distance (miles)	60.35	46.23	11,925
Migration fraction			
Never move	56.60%	-	450
Move at least once with child	12.33%	-	98
Move at least once without child	35.34%	-	281
Cumulative migration (years)			
Stay with child	12.84	3.52	795
Move with child	1.59	3.07	795
Move with child conditional	4.66	3.56	98
Move without child	0.58	1.97	795
Move without child conditional	4.49	3.69	281
Cognitive skill measures (raw scores)			
Raven's score	8.24	2.82	910
Math score	2.81	1.32	910

[a] Variable parental education is a discretized variable. Education = 1 if unschooled, = 2 if grade school, = 3 if middle school, = 4 if high school, = 5 if $\geq$ College
[b] Variable household income is measured in US dollars in the year 2000
[c] Variable relative indicates whether other relative lives with the household at the birth of the oldest child
[d] Variable distance is the distance between the capital of the province and the village where the child is born
[e] Cumulative migration is the total number of years for each migration scenario. The conditional cumulative migration variable is the total number of years for each case given the choice has been chosen once.

Table 2: Correlation Between Cognitive Skills and Migration

	Contemporaneous	Cumulative
Left-behind children (j = 2)	-0.147	-0.144
	(0.101)	(0.071)
Migrant children (j = 3)	0.256	0.467
	(0.106)	(0.181)
Covariates	Yes	Yes

[a] Clustered robust standard errors in parenthesis

Table 3: Instrumental Variable Test

Panel A: Tests for Weak Instruments	$j = 2$	$j = 3$
Distance	-0.001	0.001
	(0.000)	(0.000)
log-Ratio	-0.264	0.042
	(0.035)	(0.017)
Covariates	Yes	Yes
Under Identification Test		
Kleibergen-Paap rank LM statistic (p-value)	31.53 (0.00)	
Weak Instrument Test		
Cragg-Donald Wald F statistic	11.39	
Kleibergen-Paap Wald F statistic	10.57	
Stock-Yogo Critical values		
10% maximal relative biases	13.43	
15% maximal relative biases	8.18	
Panel B: Suggestive Evidence	**Correlation**	**St. Err.**
Distance		
Electricity availability	-0.025	0.016
Agricultural wage	-0.001	0.001
Housing price	-0.248	0.335
Number of School		
Subjective measure of school quality	-0.031	0.023

Table 4: χ^2 Goodness-of-Fit Tests of the Within-Sample Choice Distribution

Age	$j = 1$	$j = 2$	$j = 3$	Row
0	0.13	1.11	0.32	1.56
1	0.20	8.72*	1.72	10.64*
2	0.31	1.76	2.17	4.25
3	0.00	0.17	0.08	0.25
4	0.00	0.28	0.11	0.39
5	0.00	0.11	0.05	0.16
6	0.10	0.26	2.04	2.40
7	0.39	1.01	2.01	3.40
8	0.10	0.12	2.15	2.37
9	0.04	0.39	1.79	2.22
10	0.02	0.22	0.00	0.24
11	0.01	0.00	0.13	0.14
12	0.13	0.01	1.06	1.20
13	0.17	0.02	1.04	1.22
14	0.17	0.01	1.38	1.56

[a] These χ^2 statistics are not adjusted for the fact that the predicted distribution are based on estimated parameters.
[a] $\chi^2_{(1)}(0.05) = 3.84$ and $\chi^2_{(2)}(0.05) = 5.99$

Table 5: χ^2 Goodness-of-Fit Tests of the Migration Transition Matrix

	Choice (t)			
Choice ($t-1$)	$j = 1$	$j = 2$	$j = 3$	Row
$j = 1$				
Data	95.88%	0.91%	3.21%	-
Model	95.51%	0.99%	4.51%	-
χ^2	0.14	0.66	38.24*	39.04*
$j = 2$				
Data	19.51%	79.78%	0.71%	-
Model	28.42%	71.26%	0.32%	-
χ^2	35.22*	12.85*	5.99*	54.06*
$j = 3$				
Data	4.60%	1.31%	94.09%	-
Model	6.49%	0.17%	93.34%	-
χ^2	2.51	34.94*	0.03	37.48*

[a] These χ^2 statistics are not adjusted for the fact that the predicted distribution are based on estimated parameters.
[b] $\chi^2_{(1)}(0.05) = 3.84$ and $\chi^2_{(2)}(0.05) = 5.99$

Table 6: Cognitive Skill Formation Function Specification Comparison

	Contemporaneous	Cumulative
Child age	0.447	0.452
	(0.002)	(0.002)
Child age squared	−0.013	−0.013
	(0.000)	(0.000)
Child gender	0.002	0.001
	(0.004)	(0.004)
Father education	0.222	0.224
	(0.003)	(0.003)
Mother education	0.109	0.110
	(0.003)	(0.003)
Stock of children	−0.045	−0.045
	(0.004)	(0.004)
Contemporaneous $j = 2$	−0.033	-
	(0.007)	
Contemporaneous $j = 3$	0.106	-
	(0.011)	
Cumulative $j = 2$	-	−0.240
		(0.031)
Cumulative $j = 3$	-	0.434
		(0.017)
Household type 1	−3.799	−3.807
	(0.008)	(0.008)
Household type 2	−3.964	−3.976
	(0.011)	(0.011)
Household type 3	−4.018	−4.017
	(0.005)	(0.005)
Household type 4	−4.203	−4.215
	(0.001)	(0.001)

[a] Robust standard errors in parentheses.

Table 7: Effects of Cash Transfer Programs on Migration Rates

	Subsidy			Subsidy Target		
Amount	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
\$0	84.02%	11.55%	4.44%	84.02%	11.55%	4.44%
\$25	83.56%	11.49%	4.95%	83.75%	11.51%	4.74%
\$50	82.81%	11.37%	5.82%	83.26%	11.45%	5.29%
\$75	81.89%	11.22%	6.89%	82.69%	11.38%	5.93%
\$100	80.81%	11.08%	8.11%	82.08%	11.31%	6.61%
\$125	79.49%	10.90%	9.61%	81.25%	11.23%	7.52%
\$150	78.11%	10.71%	11.18%	80.39%	11.11%	8.50%
\$175	76.45%	10.43%	13.11%	79.36%	10.97%	9.67%
\$200	74.36%	10.17%	15.47%	78.05%	10.80%	11.14%

[a] j = 1 if both parents stay w/ child rural; j = 2 if at least one parent migrates w/o child; j = 3 if both parents migrate w/ child to urban

Table 8: Effects of Migration Tax on Migration Rates

	Tax			Tax Target		
Amount	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
\$0	84.02%	11.55%	4.44%	84.02%	11.55%	4.44%
\$25	85.21%	10.28%	4.51%	84.79%	10.74%	4.47%
\$50	86.31%	9.31%	4.55%	85.42%	10.09%	4.49%
\$75	86.98%	8.4%	4.62%	85.94%	9.54%	4.52%
\$100	87.68%	7.63%	4.69%	86.38%	9.07%	4.55%
\$125	88.29%	6.99%	4.72%	86.81%	8.63%	4.56%
\$150	88.81%	6.45%	4.74%	87.15%	8.28%	5.57%
\$175	89.24%	5.96%	4.80%	87.43%	7.97%	4.60%
\$200	89.65%	5.53%	4.82%	87.70%	7.68%	4.62%

[a] j = 1 if both parents stay w/ child rural; j = 2 if at least one parent migrates w/o child; j = 3 if both parents migrate w/ child to urban

Figure 1: Test Score Distribution

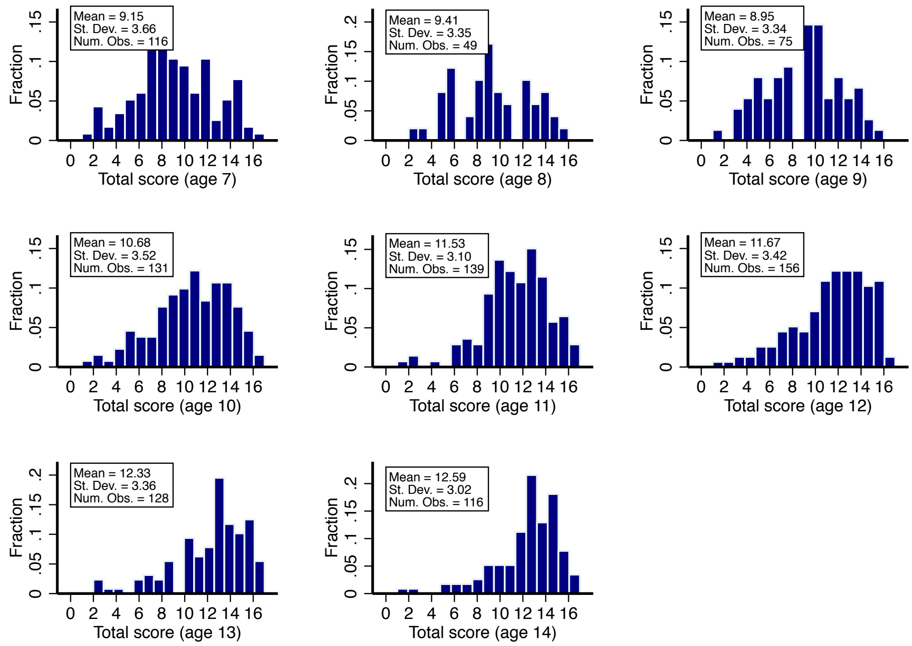


Figure 2: Latent Skill Distribution

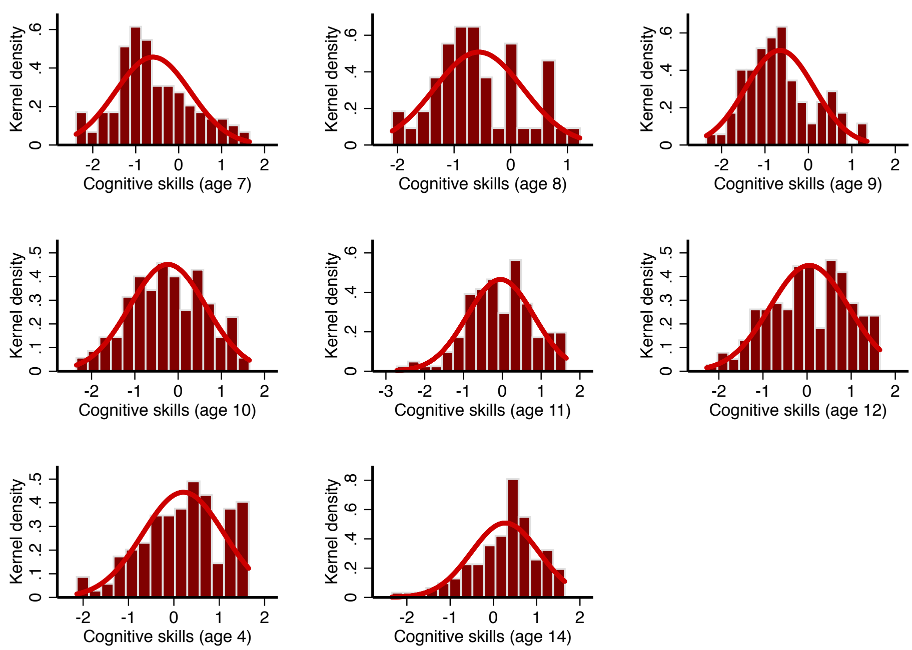


Figure 3: Exclusion Restriction Graphs

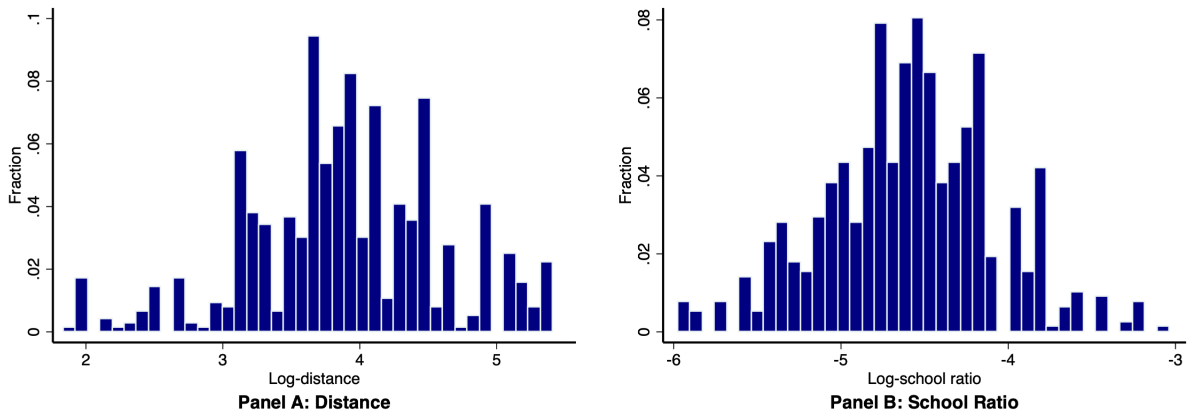


Figure 4: Migration Choice Distribution by Household Type

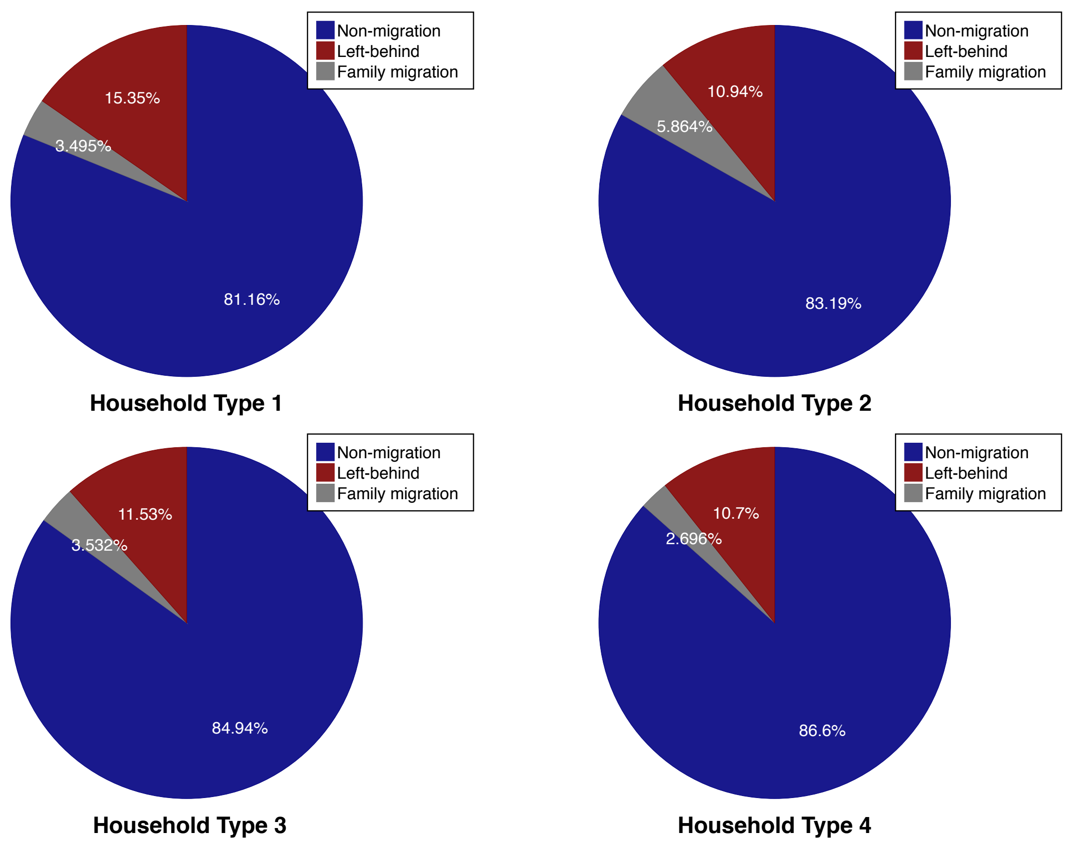


Figure 5: Counterfactual Cognitive Skill Distribution by Group

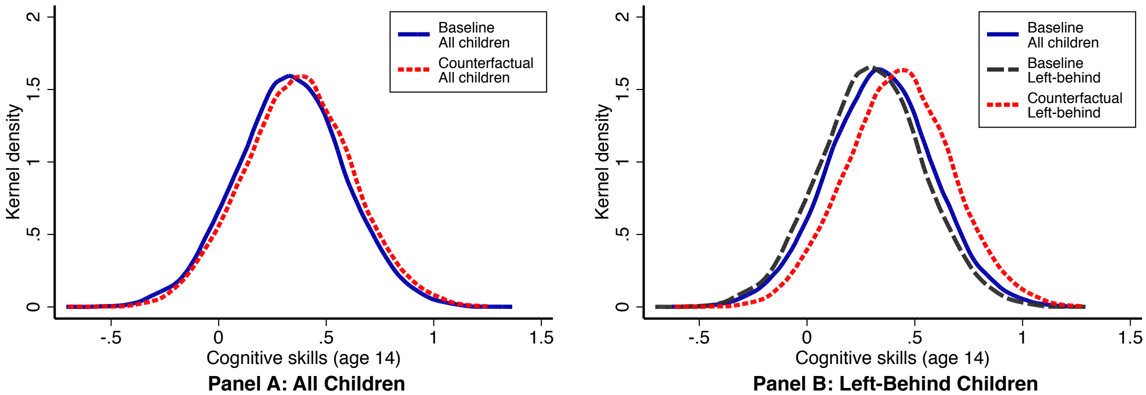


Figure 6: Counterfactual Cognitive Skill Distribution by Household Type

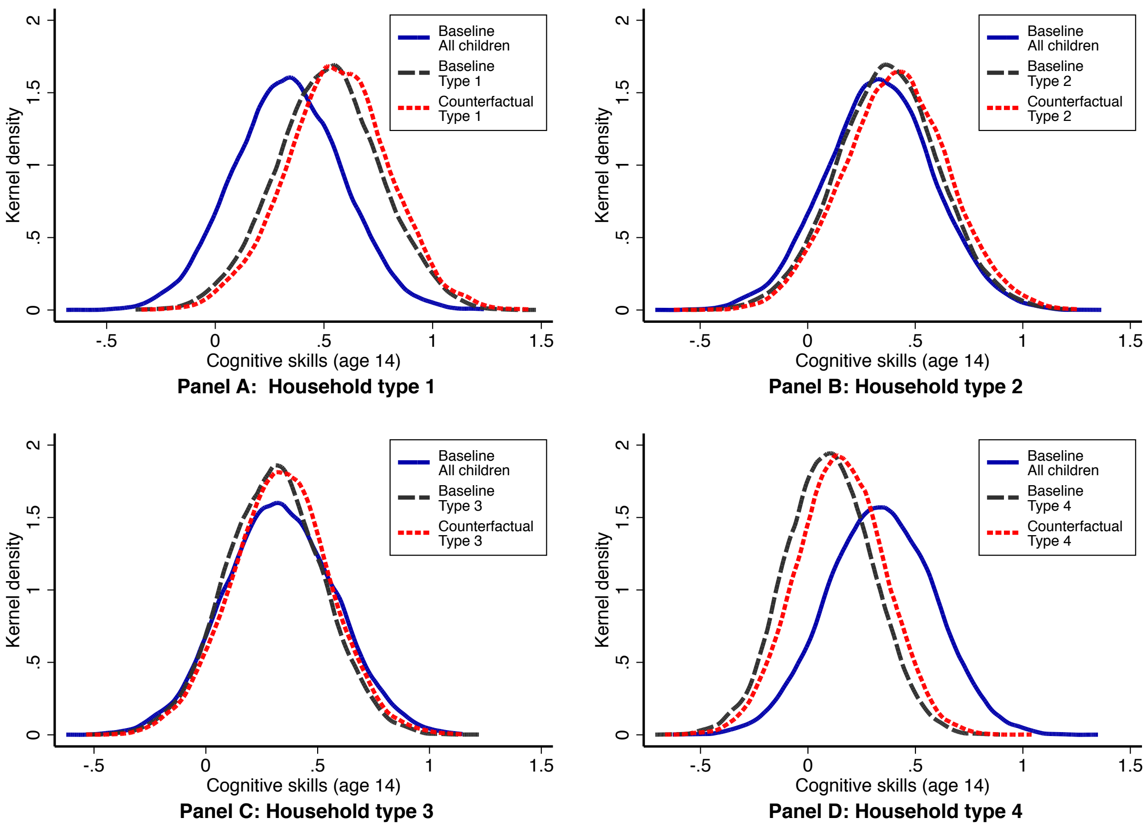


Figure 7: Effects of Cash Transfer Programs on Cognitive Skills

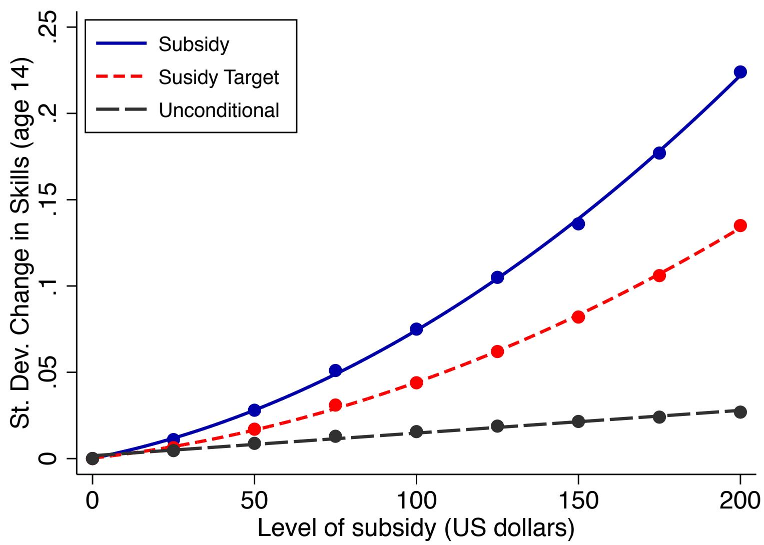
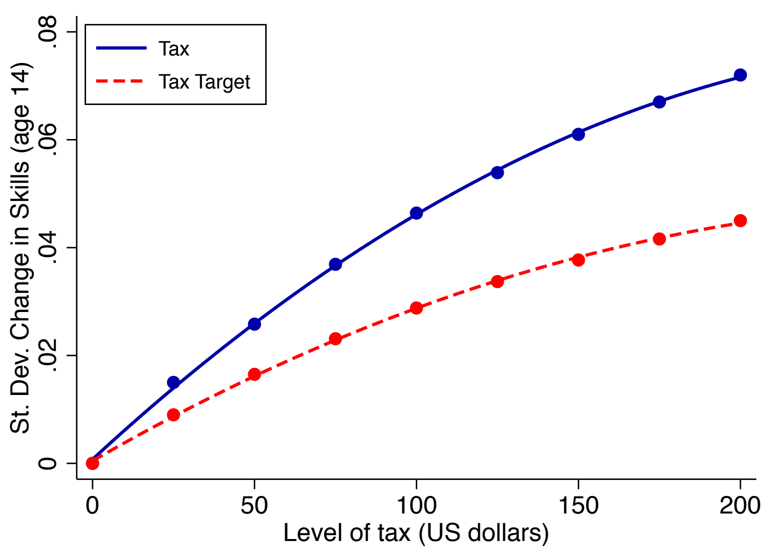


Figure 8: Effects of Migration Tax on Cognitive Skills



A Appendices

A.1 Data Appendix and Sample Formation

The Indonesia Family Life Survey (IFLS) provides data at the individual and family level on fertility, health, education, migration, and employment. Extensive community and facility data accompany the household data.[23] The IFLS tracks individuals and households in 5 waves: 1993, 1997, 2000, 2007 and 2014, known as IFLS1-5. Major efforts to reinterview all respondents result in a remarkable success rate of over 90%. The IFLS covers a sample of 7224 households spreading across 13 provinces in Indonesia. To form the sample, I collect information on household migration and income history, children's cognitive measures, and household characteristics. I construct a complete history of household migration from the 5 waves. In IFLS1, the head of the household, the spouse of the head, two randomly selected children are interviewed and followed in later waves. For adults 15 years and older, the survey tracks a migration history for every move that lasts 6 months and longer the person has made since 12 years old. For the parents who have already moved away at the time of the first interview, their migration histories are completed if they return home during later survey waves. Once an individual enters the sample, additional migration information is updated based on recall between survey waves. I gather information on household income in a similar way as migration history.[24] Next, I select rural households with parents who stay married throughout the developmental stages of childhood. A rural household is defined as one with its first child born in rural areas. From IFLS1, I sample 0-6 years old children because household income history is collected for up to the past 5 years. From IFLS2-3, I include additional new rural households with a firstborn child 0-3 years old. The constructed sample consists of 11 cohorts of 795 children and their parents. I do not include additional children from IFLS4-5 because children are not tracked long enough for this study. Given the selected sample, I compute the direct distance between the village where the oldest child was born to the provincial

[23] To retrieve the dataset, visit https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html.

[24] Income suffers missing data problems due to non-reporting and the survey structure. This creates difficulties in estimating the structural model. I provide a detailed discussion on how to deal with missing data in Section 4.

capital city of using the geographic information system. To compute the school ratio between locations, I obtain the number of schools in each provincial capital city from the Ministry of Education and Culture of Indonesia and the number of schools in one's home village using IFLS community survey. [25]

A.2 Item Response Theory and Related Models

Item Response Theory (IRT) models have been used extensively in psychology to study cognitive and personality traits. IRT is based on the idea that probability of correctly answering a questions is a function of persons latent traits and the characteristics of each question. Utilizing test questions and responses from test-takers, the IRT models relate test questions, also known as items, to the underlying traits of individuals. To fix ideas, consider the cognitive test in IFLS where questions are binary responses.[26] Let Y_{ij} represent the response to question j from individual i. Without loss of generality, I define $Y_{ij} = 1$ if individual i answer question j correctly. Consider a Two-parameter Logistic (2PL) model with parameters $\Gamma \equiv (\kappa, \lambda)$, the probability of person j with latent trait level latent ability ζ_j providing a correct response to the item i is given by:

$$\Pr(Y_{ij} = 1 \mid \Gamma, \zeta_j) = \frac{\exp\{\kappa_i(\zeta_j - \lambda_i)\}}{1 + \exp\{\kappa_i(\zeta_j - \lambda_i)\}}$$

where κ_i represents discrimination, and λ_i represents the difficulty of item i. The more difficulty a question is, the less likely an individual answers it correctly. The higher discrimination is, the better it can distinguish between low and high levels of the latent trait. In other words, in the neighborhood of a given difficult level, questions with higher discrimination mean that two students with distinct abilities would have different predicted probabilities of responding correctly. Individual latent ability ζ is assumed to have a standard Normal distribution. Then the individual j's contribution to the likelihood is:

$$\mathbb{L}_j(\Gamma) = \int_{\zeta_j} \prod_{i=1}^{I} \Pr(Y_{ij} \mid \Gamma, \zeta_j)^{\mathbb{1}\{Y_{ij}=1\}} [1 - \Pr(Y_{ij} \mid \Gamma, \zeta_j)]^{\mathbb{1}\{Y_{ij}=0\}} \phi(\zeta_j) d\zeta_j,$$

[25]For more information, please visit https://referensi.data.kemdikbud.go.id/index.php

[26]A complete set of questions can be found on the IFLS website https://www.rand.org/well-being/social-and-behavioral-policy/data/FLS/IFLS.html

where the integral is with respect to latent trait ζ_j and is approximated numerically.

I estimate the 2PL model with all children aged 7-15 in the IFLS sample. Estimation results in Table A.4 show there is a considerable variation in difficulty as well as discrimination levels. To interpret results, Panel A in Figure A.3 plots the Item Characteristic Curve (ICC). The difficulty parameter κ is represented by the location of an item on the ability scale. The discrimination parameter λ is related to the slope of the ICC for an given item. In Panel A, question 1 and question 10 have similar discrimination levels, but question 10 is more difficult than question 1. A person with an ability level $\zeta = -1$ has a higher chance of answering question 1 correctly. Now consider two people, one with ability just below $\zeta = 1$, and the other with ability just above $\zeta = 1$. According to the ICC for question 16, these people would have similar chances of answering it correctly. On the other hand, according to the ICC for question 12, the individual with the higher ability level would have a substantially higher probability of success on question 12. Given parameter estimates from 2PL model, I predict the latent trait using the empirical Bayesian method. It combines the prior information about the latent trait, i.e., a standard Normal, with the likelihood to obtain the conditional posterior distribution of the latent trait. Panel B in Figure A.3 plots the expected score for a given level of latent cognitive skill. As a robustness check, I also estimate a One-parameter Logistic (1PL) model. The 1PL model has an additional restriction that discrimination parameter $\kappa_i = \kappa$ for all items i.[27] Since the 1PL model is nested within the 2PL model, I perform a likelihood-ratio test to assess which model is preferred. The large likelihood-ratio statistic, i.e., LR $\chi^2(16) = 6221.98$, favors the 2PL model. Panel B also shows that the selected model provides a reasonable fit to the data.

[27] Three-parameter Logistic model (3PL) has a converging issue in the maximum likelihood estimation.

A.3 Simulated Maximum Likelihood Estimation

A.3.1 Model Solution

To illustrate how I compute the likelihood to the dynamic model, I first briefly discuss the solution method to the dynamic programing program. The value function in period t is:

$$V_{jt}(\Omega_t) = U_{jt}(\bar{\Omega}_t, \eta_t, \varepsilon_{jt}) + \rho E[V_{t+1}(\Omega_{t+1}) \mid \Omega_t, j_t = j]$$

where $\bar{\Omega}_t$ includes all state variables known to the econometrician and discrete types. To calculate the alternative specific value functions at t, we need to compute $E[V_{t+1}(\Omega_{t+1}) \mid \Omega_t, j_t]$. Due to the Type I Extreme Value assumption on preference shocks, Normal income shocks, and serial uncorrelation among these shocks, I derive the expectation over maximization of future payoff:

$$E[V_{t+1}(\Omega_{t+1}) \mid \Omega_t, j_t] = \sum_{n_t=0,1} p_{n_{t+1}} \left\{ \int_{\eta_{t+1}} \gamma + \log\big[\sum_{j=1}^{3} \exp\big(V_{jt}(\bar{\Omega}_{t+1}, \eta_{t+1})\big)\big] \, dF(\eta_{t+1}) \right\}$$

where the expectation with respect to preference shocks have a closed form solution but the expectation with respect to income is approximated with Monte Carlo simulation. I make 125 draws from a joint normal distribution. The number of draw provides reasonable approximations because the mean square errors are small when compared to the case using 1000 draws.

A.3.2 Likelihood Function & Standard Errors

To illustrate the idea of dealing with missing income, I consider the case where choice alternative $j_t = 1$. If income Y_{1t} is observed, the likelihood contribution is:

$$L_{it}(\theta) = \Pr(j_t = 1 \mid \Omega_t, k; \theta) \cdot g(Y_{1t} \mid \Omega_t, k; \theta) \cdot h(Q_t \mid \Omega_t, k; \theta)^{\mathbb{1}\{Q_t \text{ observed}\}}$$

where $\Pr(d_{jt} = 1 \mid \Omega_t, k; \theta)$ is choice probability, $g(Y_1 t \mid \Omega_t, k; \theta)$ and $h(Q_t \mid \Omega_t, k; \theta)$ are densities for income and skills, respectively. The conditional choice probability in period t is:

$$\Pr(j_t = 1 \mid \Omega_t, k; \theta) = \frac{\exp\{V_{1t}(\bar{\Omega}_t, \eta_{1t})\}}{\sum_{j=1}^{3} \exp\{V_{jt}(\bar{\Omega}_t, \eta_{jt})\}}$$

where $\eta_{1t} = \ln Y_{1t} - \ln \bar{Y}_{1t}$ is realized. I compute $V_{jt}(\bar{\Omega}_t, \eta_{jt})$ using model solution in Appendix A.3.1 and observed state variables in the data. However, I still need to integrate out the income for choice alternatives $j = 2$ and $j = 3$:

$$\Pr(j_t = 1 \mid \Omega_t, p, k; \theta) = \int_{\eta_{2t}, \eta_{3t}} \Pr(j_t = 1 \mid \Omega_t, \eta_{2t}, \eta_{3t}, k; \theta) \;\; dF(\eta_{2t}, \eta_{3t} | \eta_{1t} = \ln Y_{1t} - \ln \bar{Y}_{1t}).$$

To compute the integral, first, I draw income shocks η^m_{i2t} and η^m_{i3t} for each household i at each period t and construct simulated income $\tilde{Y}^m_{i2t}$ and $\tilde{Y}^m_{i3t}$ according to Equation (3). Next, I construct the simulated choice probability as,

$$\Pr(j_t = 1 \mid \Omega_t, k; \theta) = \frac{1}{M} \sum_{m=1}^{M} \frac{\exp\left\{V_{1t}(\bar{\Omega}_t, Y_{i1t})\right\}}{\exp\left\{V_{1t}(\bar{\Omega}_t, Y_{i1t})\right\} + \exp\left\{V_{jt}(\bar{\Omega}_t, \tilde{Y}^m_{i2t})\right\} + \exp\left\{V_{jt}(\bar{\Omega}_t, \tilde{Y}^m_{i3t})\right\}}.$$

where M is total number of draws. If income Y_{1t} is not observed, the likelihood contribution is:

$$L_{it}(\theta) = \Pr(j_t = 1 \mid \Omega_t, k; \theta) \cdot h(Q_t \mid \Omega_t, k; \theta)^{\mathbb{1}\{Q_t \text{ observed}\}}$$

To construct the choice probability, I integrate out all the unobserved income components:

$$\Pr(j_t = 1 \mid \Omega_t, Q_t, k; \theta) = \int_{\eta_{1t}, \eta_{2t}, \eta_{3t}} \Pr(j_t = 1 \mid \Omega_t, Y_{1t}, \eta_t, Q_t, k; \theta) \;\; dF(\eta_{1t}, \eta_{2t}, \eta_{3t}).$$

These integrals are again approximated numerically via simulation as the previous case.

I estimate the standard errors are using the gradients of the likelihood. The expression for the asymptotic standard error is:

$$\text{St. Err.}(\theta) = \left[\sum_{i=1}^{N} s_i(\hat{\theta}) s_i(\hat{\theta})'\right]^{-1/2}$$
$$s_i(\hat{\theta}) = \bigtriangledown_\theta \log L_{it}(\theta).$$

Because the likelihood $L_{it}(\theta)$ does not have an analytical form, I approximate $s_i(\hat{\theta})$ using numerical derivatives instead.

A.4 Additional Figures and Tables

Figure A.1: Map of Indonesia

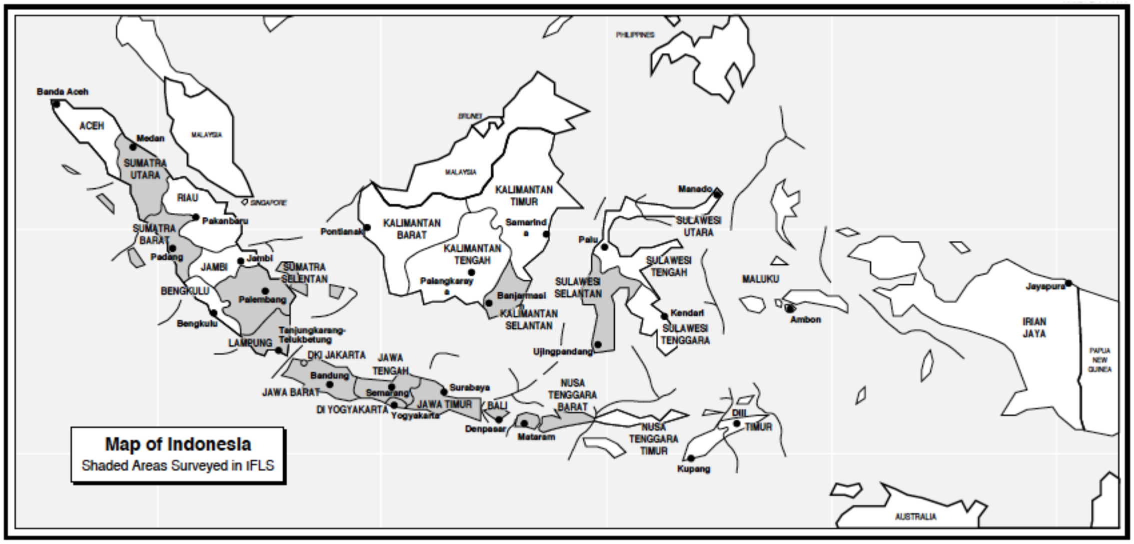


Figure A.2: Raven's Colored Progressive Matrices Example

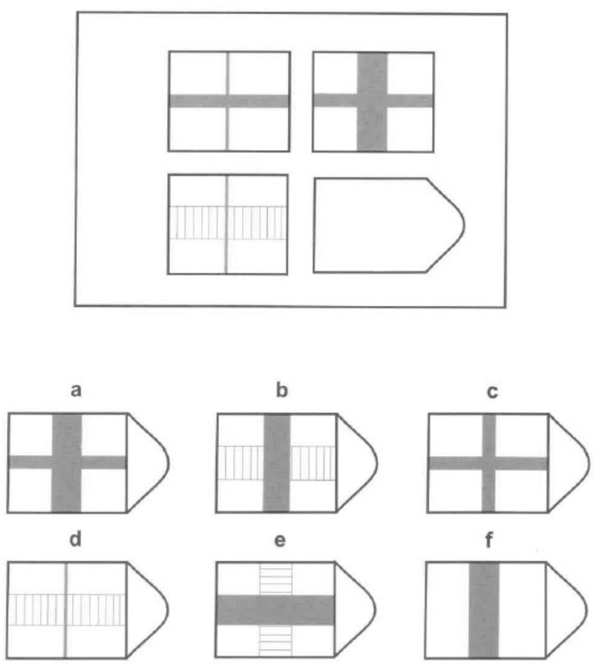
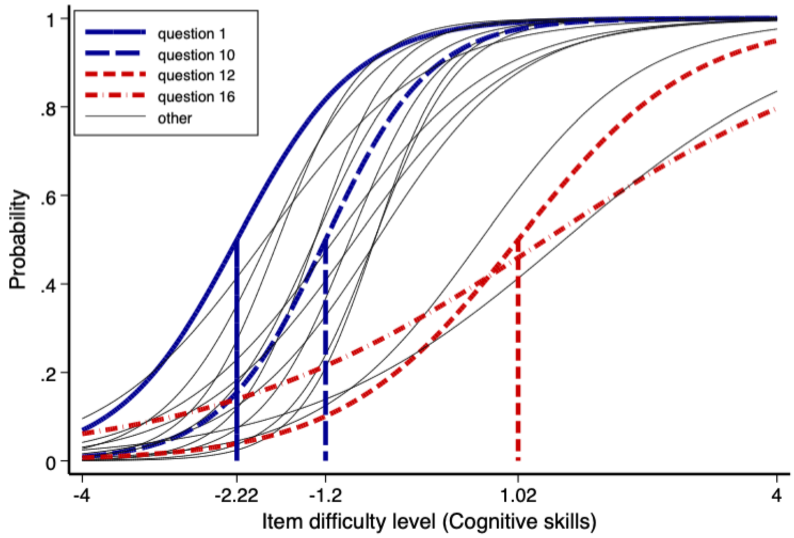
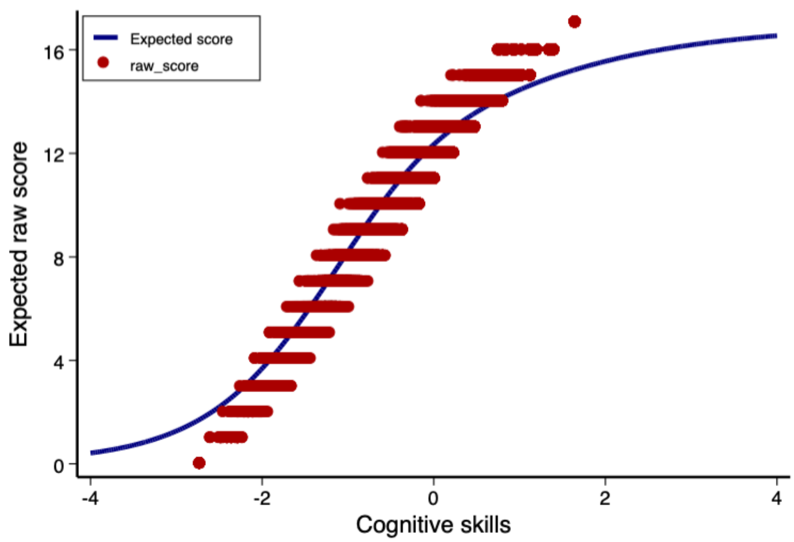


Figure A.3: Item Response Theory Model Graphs



Panel A: Item Characteristics Curve



Panel B: Test Characteristics Curve

Figure A.4: Model Fit to Migration Choice Distribution

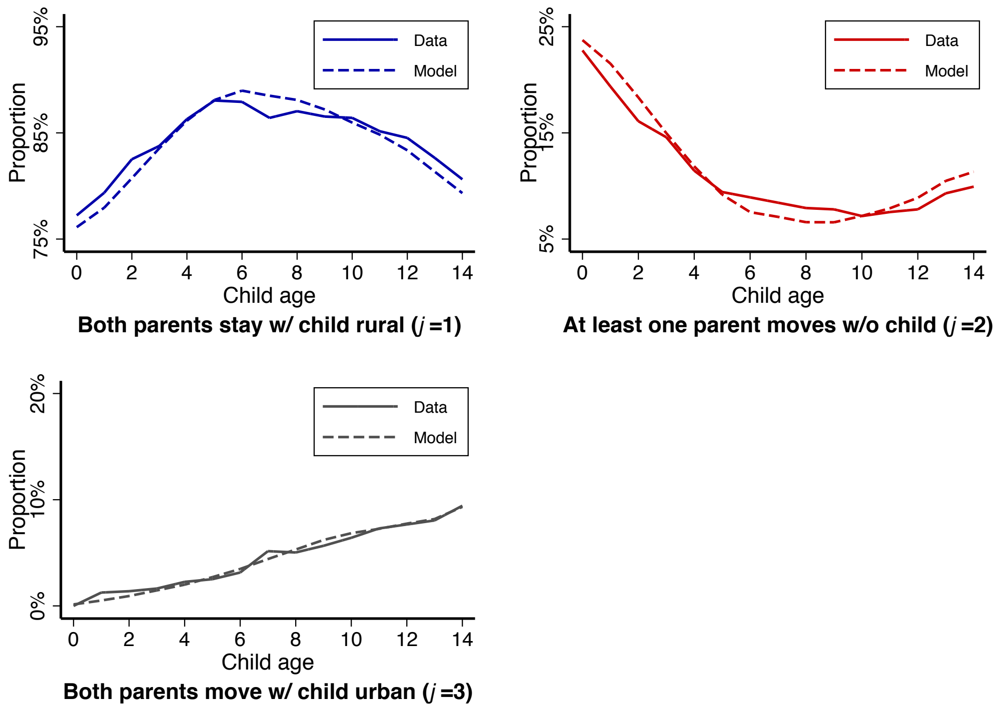


Figure A.5: Model Fit to Cognitive Skill Distribution by Child Age

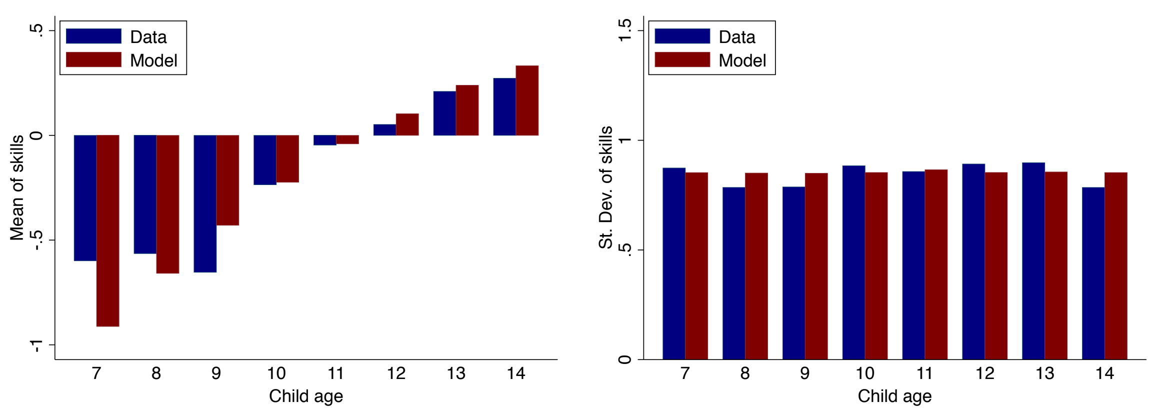


Figure A.6: Model Fit to Cognitive Skill Distribution by Cumulative Migration

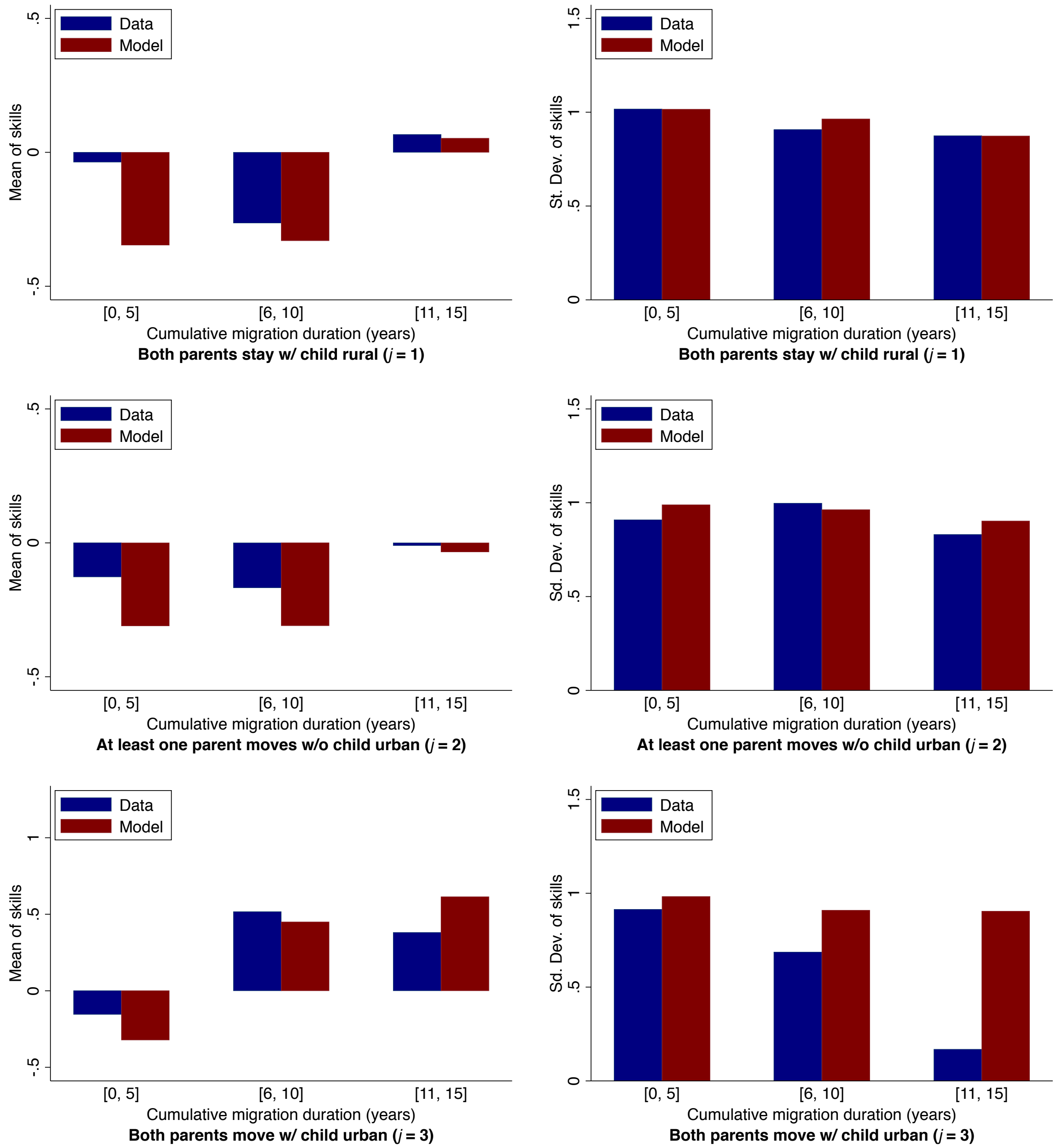


Figure A.7: Model Fit to Income Distribution by Migration Status

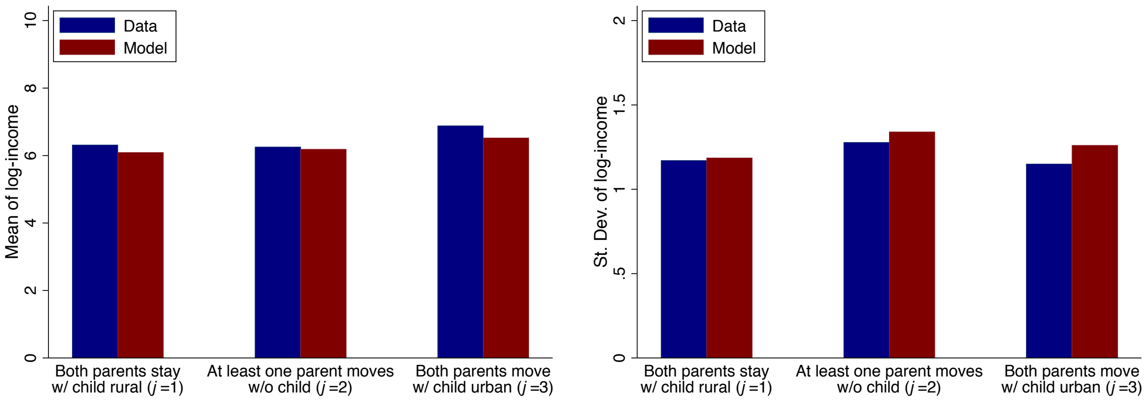


Figure A.8: Model Fit to Income Distribution by Parental Education

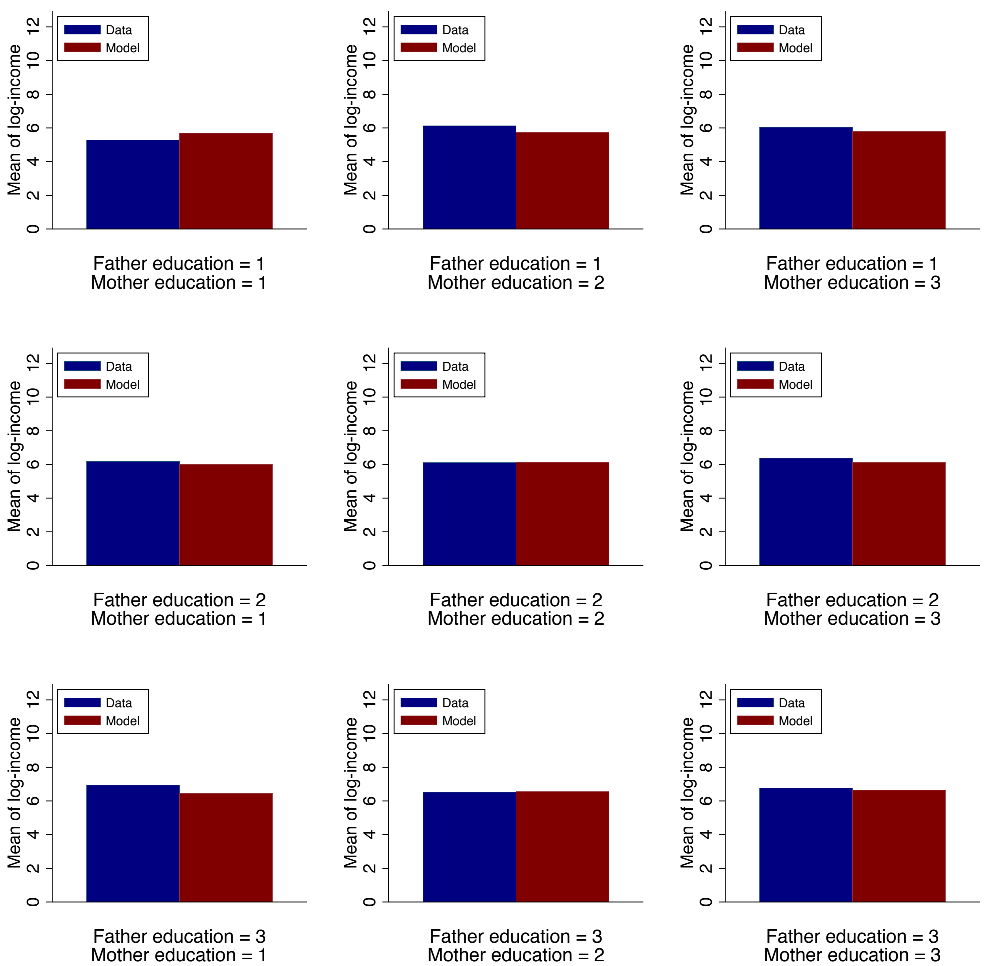


Table A.1: Parameter Estimates

	Parameter	Estimate	St. Err.
Utility Function			
Consumption & child quality interaction	α_{cq}	-0.745	0.303
$j = 2$			
Consumption	α_{2c}	0.237	0.273
Child production	α_{2q}	1.345	0.646
Child production (terminal)	α_{2qT}	3.191	0.661
Transition cost	α_{21}	-4.511	0.156
Migration cost by age	α_{22}	-0.313	0.095
Migration cost by age^2	α_{23}	0.015	0.026
Relative existence	α_{24}	0.019	0.003
School ratio	α_{25}	-0.068	0.007
Household type 1	α_{2k_1}	2.836	0.803
Household type 2	α_{2k_2}	2.406	0.137
Household type 3	α_{2k_3}	1.794	0.227
Household type 4	α_{2k_4}	1.780	0.098
$j = 3$			
Consumption	α_{3c}	0.848	0.355
Child quality	α_{3q}	0.838	0.808
Child quality (terminal)	α_{3qT}	5.169	0.969
Transition cost	α_{31}	-6.645	0.419
Migration gain by age	α_{32}	0.261	0.108
Migration gain by age^2	α_{33}	-0.011	0.046
Relative existence	α_{34}	0.079	0.061
School ratio	α_{35}	-0.007	0.748
Household type 1	α_{3k_1}	-1.419	1.792
Household type 2	α_{3k_2}	-1.505	0.332
Household type 3	α_{3k_3}	-1.746	0.541
Household type 4	α_{3k_4}	-1.625	0.254
Budget Constraint			
Monetary cost	Δ	0.008	0.002
Type Proportion			
Household type 1	μ_{k_1}	0.097	-
Household type 2	μ_{k_2}	0.429	0.036
Household type 3	μ_{k_3}	0.360	0.027
Household type 4	μ_{k_4}	0.114	0.034

[a] j = 1 if both parents stay w/ child rural; j = 2 if at least one parent migrates w/o child; j = 3 if both parents migrate w/ child to urban

Table A.2: Parameter Estimates

	Parameter	Estimate	St. Err.
Income			
$j = 1$			
Father education	β_{11}	0.384	0.048
Mother education	β_{12}	0.050	0.088
Household type 1	β_{k_1}	6.634	0.167
Household type 2	β_{k_2}	5.691	0.040
Household type 3	β_{k_3}	4.832	0.042
Household type 4	β_{k_4}	3.785	0.038
Income shock variance	$\sigma^2_{\eta_1}$	0.742	0.066
$j = 2$			
Father education	β_{21}	0.514	0.092
Mother education	β_{22}	0.066	0.197
Household type 1	β_{k_1}	6.319	0.183
Household type 2	β_{k_2}	5.391	0.104
Household type 3	β_{k_3}	4.557	0.084
Household type 4	β_{k_4}	3.189	0.105
Income shock variance	$\sigma^2_{\eta_2}$	0.995	0.049
$j = 3$			
Father education	β_{31}	0.626	0.154
Mother education	β_{32}	0.167	0.535
Household type 1	β_{k_1}	5.787	0.524
Household type 2	β_{k_2}	4.882	0.423
Household type 3	β_{k_3}	4.187	0.201
Household type 4	β_{k_4}	2.097	0.408
Income shock variance	$\sigma^2_{\eta_3}$	0.567	0.074

[a] j = 1 if both parents stay w/ child rural; j = 2 if at least one parent migrates w/o child; j = 3 if both parents migrate w/ child to urban

Table A.3: Parameter Estimates

	Parameter	Estimate	St. Err.
Cognitive Skill Formation Function			
Child age	δ_1	0.454	0.068
Child age squared	δ_2	-0.013	0.032
Child gender	δ_3	-0.004	0.050
Father education	δ_4	0.221	0.062
Mother education	δ_5	0.108	0.043
Stock of children	δ_6	-0.044	0.039
Cumulative $j = 2$	δ_7	-0.021	0.011
Cumulative $j = 3$	δ_8	0.039	0.004
Cumulative $j = 2$ squared	δ_9	0.000	0.000
Cumulative $j = 3$ squared	δ_{10}	-0.001	0.354
Household type 1	δ_{k_1}	-3.802	0.385
Household type 2	δ_{k_2}	-3.970	0.093
Household type 3	δ_{k_3}	-4.026	0.106
Household type 4	δ_{k_4}	-4.213	0.100
Measurement error variance	σ^2_ω	0.671	0.036
Fertility Transition			
Stock of children	γ_1	-1.672	0.095
Constant	γ_0	0.198	0.045

[a] $j = 1$ if both parents stay w/ child rural; $j = 2$ if at least one parent migrates w/o child; $j = 3$ if both parents migrate w/ child to urban

Table A.4: Two-parameter Logistic Model Estimation Results

	Difficulty Level κ		Discrimination λ	
Item	Estimate	St. Err.	Estimate	St. Err.
1	−2.220	0.039	1.462	0.037
2	−0.918	0.013	1.953	0.035
3	−0.631	0.011	2.322	0.041
4	−0.637	0.011	2.053	0.036
5	−0.606	0.011	1.211	0.022
6	0.589	0.016	1.087	0.021
7	−1.853	0.028	1.656	0.039
8	−1.743	0.023	2.047	0.048
9	−1.321	0.016	2.233	0.046
10	−1.198	0.017	1.660	0.031
11	−1.322	0.018	1.746	0.034
12	1.019	0.022	0.986	0.021
13	−1.897	0.025	1.069	0.025
14	−1.100	0.021	1.082	0.022
15	−0.859	0.018	1.097	0.021
16	1.334	0.046	0.511	0.016
17	1.550	0.041	0.662	0.017

[a] Item 1-12 are questions from Raven's Progressive Matrices' Test and item 13-17 are mathematical questions.
[b] Difficulty and discrimination level of each question increase as the magnitude of its coefficient.
[c] LR $\chi^2(16)$ = 6221.98 for likelihood-ratio test between One-parameter Logistic Model vs Two-parameter Logistic Model