Essays on Child Development

by

Morteza Saharkhiz

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved February 2018 by the
Graduate Supervisory Committee:

Daniel Silverman, Co-Chair
Matthew Wiswall, Co-Chair
Esteban Aucejo
Gregory Veramendi

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

This dissertation comprises three chapters. In chapter one, using a rich dataset for the United States, I estimate a series of models to document the birth order effects on cognitive outcomes, non-cognitive outcomes, and parental investments. I estimate a model that allows for heterogeneous birth order effects by unobservables to examine how birth order effects varies across households. I find that first-born children score 0.2 of a standard deviation higher on cognitive and non-cognitive outcomes than their later-born siblings. They also receive 10% more in parental time, which accounts for more than half of the differences in outcomes. I document that birth order effects vary between 0.1 and 0.4 of a standard deviation across households with the effects being smaller in households with certain characteristics such as a high income.

In chapter two, I build a model of intra-household resource allocation that endogenously generates the decreasing birth order effects in household income with the aim of using the model for counterfactual policy experiments. The model has a life-cycle framework in which a household with two children confronts a sequence of time constraints and a lifetime monetary constraint, and divides the available time and monetary resources between consumption and investment. The counterfactual experiment shows that an annual income transfer of 10,000 USD to low-income households decreases the birth order effects on cognitive and non-cognitive skills by one-sixth, which is five times bigger than the effect in high-income household.

In chapter three, with Francesco Agostinelli and Matthew Wiswall, we examine the relative importance of investments at home and at school during an important transition for many children, entering formal schooling at kindergarten. Moreover, our framework allows for complementarities between children's skills and investments from schools. We find that investments from schools are an important determinant of children's skills at the end of kindergarten, whereas parental investments, although strongly correlated with end-of-kindergarten outcomes, have smaller effects. In addition, we document a negative complementarity between children's skills at kindergarten entry and investments from schools, implying that low-skill children benefit the most from an increase in the quality of schools.

To my parents.

ACKNOWLEDGMENTS

I am thankful to my advisor Matthew Wiswall for his continuous encouragement, guidance, and support. This dissertation would not have been possible without him. I have also greatly benefited from many discussions with Esteban Aucejo, Daniel Silverman, and Gregory Veramendi. I am thankful for their comments. I also thank Francesco Agostinelli, Edward C. Prescott, Edward Schlee, Sajad Vahedi, and Gustavo Ventura. I am responsible for all errors.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Chapter 1

THE EFFECT OF BIRTH ORDER ON CHILDHOOD DEVELOPMENT

1.1 Introduction

Decades of research in economics and other disciplines have documented differences in many socio-economic indicators by order of birth. [1] In this chapter, using a rich dataset for the United States, I estimate a series of models in order to document the birth order effects on cognitive outcomes, non-cognitive outcomes, and parental investments during childhood development. I also estimate models that allow for heterogeneous birth order effects by unobservables (household fixed effects) to examine how the birth order effects vary across households.

I find that first-born children score 0.2 of a standard deviation higher on cognitive and non-cognitive outcomes than their later-born siblings. They also receive 0.1 of standard deviation more in parental investments as measured by parental time. Moreover, naive regressions show that the birth order effects on parental time account for more than half of the birth order effects on outcomes. I document that birth order effects are heterogeneous and vary between 0.1 and 0.4 of a standard deviation across households with the effects being smaller in households with certain characteristics such as a high income. The literature assumes homogeneous birth order effects across households with the exception of Buckles and Munnich (2012) which find hetero-

[1] Some examples include cognitive outcomes (Lehmann *et al.*, 2016; Monfardini and See, 2012; Heiland, 2009), non-cognitive outcomes (Black *et al.*, 2017), educational attainment (Bu, 2014; Kantarevic and Mechoulan, 2006; Black *et al.*, 2005), adult earnings (Kantarevic and Mechoulan, 2006), risky behavior such as smoking, drinking, or copulating (Argys *et al.*, 2006; Rodgers *et al.*, 1992), delinquency outcomes (Breining *et al.*, 2017), physical health, mental health, and happiness (Black *et al.*, 2016), all-cause and cause-specific mortality risk (Barclay and Kolk, 2015), and suicide risk (Rostila *et al.*, 2014).

geneity in the effects by birth spacing. Documenting the heterogeneity in the birth order effects along various dimensions is important and it helps us to understand the underlying mechanisms that generate the effects.

Subsequent sections of this chapter are organized as follows. Section 1.2 briefly summarizes the literature. Section 1.3 describes the data. Section 1.4 documents the evidence on the birth order effects. Section 1.5 concludes the chapter.

1.2 Previous Research

There are three main hypotheses that have been put forward to explain the observed differences between siblings: differences in initial endowments among children, differences in monetary and time resources allocated to children, and differences in household environment and parenting practices among children.

The first group of literature focuses on maternal behavior during prenatal and neonatal development periods. This literature argues that differences in prenatal and neonatal investments cause the differences in initial endowments which, in turn, cause the differences in outcomes. There is evidence that mothers are more likely to engage in risky prenatal behavior, such as alcohol or cigarette consumption, during later pregnancies in comparison with the earlier pregnancies (Lehmann *et al.*, 2016; Kleinman and Madans, 1985). There is also evidence that mothers are less likely to breastfeed later-born children in comparison with earlier-born children (Buckles and Kolka, 2014).

The second group of literature is concerned with resource dilution (Becker and Tomes, 1976; Becker and Lewis, 1973). This literature claims that differences in allocated time and monetary resources across siblings causes the differences in outcomes. For example, there is evidence that later-born children have fewer investment goods at home, such as books (Lehmann *et al.*, 2016; Booth and Kee, 2009; Blake, 1981).

There is also evidence that parents spend less time with the later-born children in comparison with the earlier-born children (Pavan, 2015; Monfardini and See, 2012; Price, 2008).

The third group of literature concentrates on the household environment and parenting practices. The family mental confluence model posits that a child with a higher birth order is born in a household with a lower mental age in comparison to a child with a lower birth order and, therefore, they have a lower rate of intellectual growth than their older siblings (Zajonc, 2001, 1976; Zajonc and Markus, 1975). There is also literature on parenting practices which postulates that parents are more strict with the earlier-born children, which results in better outcomes, in order to establish a reputation of toughness to deter bad behavior among later-born children. For example, parents have less rigid rules about watching television and monitoring homework for later-born children (Hotz and Pantano, 2015; Hao and Jin, 2008).

1.3 Data

The data is from the Panel Study of Income Dynamics (PSID) and its Child Development Supplement (CDS). The PSID is a nationally representative longitudinal study of US families which began in 1968. It was conducted annually through 1997 and biennially ever since. The PSID collects data on the "head" and the "wife" of the household and has information on employment, income, marriage, and fertility.

The CDS was introduced in 1997 to study the development of children. It randomly selected up to two children aged 12 or younger per PSID Family Units with two follow-ups in 2002 and 2007. The CDS includes about 3600 children from 2400 households. It contains data on children's cognitive and non-cognitive skills, parental investments in children's skills, and children's investments in their own skills.

A close dataset to the PSID-CDS is the National Longitudinal Study of Youth

(NLSY). The birth order differences in outcomes that I find using the PSID-CDS are similar to those documented using the NLSY (Pavan, 2015). The CDS is preferable as it has better investment measures. The NLSY measures parental investments by asking parents for the frequency of a selected activities while the CDS also includes time diaries of children activities. Time diary is a more objective measure of parental investments than activity frequency as they measure what parents actually do as opposed to what parents claim they do; therefore, time diary is an important component of this study.

1.3.1 Sample

The sample includes all households in the CDS who have two children as of 2015. I focus on households with two children to rule out family size effects, which is an important confounding factor with birth order, and to simplify the modeling of birth order effects, which is the subject of Chapter 2. Households that have a twin are excluded from the sample because birth order is not well defined for twins. In addition, households that have adopted a child are also removed from the sample. The evidence on birth order effects is replicated using the sample of households with three or fewer children and is included in Appendix A.6. Generally, the qualitative results are similar.

Table 1.1 presents the descriptive statistics for the sample in the first wave of the CDS. The sample is composed of 1258 children from 862 households; [2] although, it varies by some outcomes because of non-response to certain questions. The children are on average 6.6 years old. Fifty percent are male. About fifty-seven percent are

[2]There are three types of households in the sample: first, households that had two children younger than 12 in 1997 and did not give birth to another child (both children are observed); second, households that had one child younger than 12 and another child older than 12 in 1997 and did not give birth to additional children (only the second child is observed); third, households that had one child younger than 12 in 1997 and gave birth to a child afterward (only the first child is observed).

White non-Hispanic, thirty-three percent are Black non-Hispanic, and five percent are Hispanic. In addition, sixty-nine percent of the children live with their biological parents. Mothers are on average 33.4 years old and have had their first pregnancy at 24 years of age. They have an average of 14 years of education and work an average of 25 hours per week. The average household income is 50,000 in 1996 USD. Note that Whites are underrepresented and Blacks are overrepresented in the sample as PSID has an over sample of low-income households.

1.3.2 Cognitive and Non-Cognitive Scales

Panels A and B of Table 1.2 present the descriptive statistics for cognitive and non-cognitive scores in the first wave of the CDS, respectively. For all three waves of the CDS, three subtests of the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R) were administered to children aged three years or older: the Letter-Word Identification (LWI), which assesses symbolic learning and reading identification skills, the Passage Comprehension (PC), [3] which assesses vocabulary and reading comprehension, and the Applied Problems (AP), which assesses mathematics reasoning and knowledge. The CDS also has two non-cognitive scales: the Behavior Problem Scale (BPS), which measures the incidence and severity of child behavior problems, and the Positive Behavior Scale (PBS), which measures the positive aspects. The BPS is borrowed from NLSY and has the advantage that it has been well used by economists (Cunha *et al.*, 2010). Both of the scales were administered to parents of children aged three years or older in all waves of data collection. One limitation of these scales, and other cognitive and non-cognitive scales in general, is that they have a bounded range. This may result in the ceiling effects, which occurs when a child cannot score more than the maximum available points. Appendix A.1 demonstrates

[3] PC is administered to children aged six years or older.

that the findings are robust to the ceiling effects.

1.3.3 Parental Investments

Panel C of Table 1.2 summarizes measures of parental investments in the first wave of the CDS. The CDS contains several instruments that measure parental investments including a Time Diary, which is a chronological report of a child's activities over a 24-hour period that records the activity, duration, location, active participants, and passive participants. Two time diaries are available for each child, one for a random weekday and another for a random weekend day.

For this study, I focus on the total active time that children spend with their parents (biological or step) measured as the weighted sum of the weekday and the weekend durations with weights of five and two, respectively. I further break down each measure of parental time by sibling participation. Specifically, if a sibling is an active or a passive participant of an activity, then the time is measured as parental shared time, otherwise, it is measured as parental alone time. All the birth order effects on parental active time qualitatively hold for parental passive time and are included in Appendix A.3. The birth order effects on alternative measures of parental investments in the CDS are included in Appendix A.4.

1.4 Evidence on Birth Order Effects

In this section, I estimate a series of models in order to document the effect of birth order on cognitive outcomes, non-cognitive outcomes, and parental investments. I also examine how much of the birth order effects on outcomes is explained by the birth order effects in parental investments. I conclude this section by estimating a model that allow for heterogeneity by unobservables (household fixed effects) in order to examine how the birth order effects vary across households.

1.4.1 *Differences in Outcomes*

The baseline identification strategy exploits the within household variation. Let $i \in \{1, 2\}$ denote the order of birth, h denote the household, and consider the following equation:

$$S_{i,h} = \gamma_0 q_h + \gamma_1 1\{i = 2\} + X_{i,h}\gamma_2 + \epsilon_{i,h} \tag{1.1}$$

where $S_{i,h}$ is a child's outcome, q_h is the household fixed effects, $1\{i = 2\}$ is an indicator for the second-born children, $X_{i,h}$ is a vector that includes child specific variables such as gender, age, age squared, and an intercept, and $\epsilon_{i,h}$ is the error term. [4] Household fixed effects, q_h, stands for all household characteristics that do not vary across children. Specifically, it includes important confounding factors with birth order such as household size, mother's age at birth, and birth spacing of the household (age difference between siblings). Without any loss of generality, I normalize the mean of q_h to zero and its standard deviation to one. [5] The coefficient of interest is γ_1 which is the effect of being born second on $S_{i,h}$ relative to the first child.

The within transformation for children in household h is the difference between the first and second child outcomes, that is:

$$S_{1,h} - S_{2,h} = \gamma_1 + (X_{1,h} - X_{2,h})\gamma_2 + \epsilon_{1,h} - \epsilon_{2,h} \tag{1.2}$$

This removes the household fixed effects, q_h. Under the assumption that $E[\epsilon_{1,h} - \epsilon_{2,h}|X_{1,h} - X_{2,h}] = 0$, γ's are identified and can be estimated using OLS. Since the

[4] I have also estimated models with age dummy variables, and there is little difference in the results. I do it this way to be consistent with models that allow for heterogeneity in the birth order effects by the unobservable, q_h, in which the computation time increases dramatically with additional variables.

[5] In typical fixed effect analyses, $\gamma_0 = 1$ and the standard deviation of the fixed effect is allowed to be free. I have chosen a different normalization here for ease of interpretation.

identification strategy uses within household variation, it is robust to unobserved household characteristics that do not vary by birth order. For example, the birth order may be correlated with household lifetime monetary resources, which is an unobservable to the researcher. Then, the naive OLS estimate of γ_1 is biased while FE estimate is unbiased.

Equation 1.2 can also be written as:

$$S_{1,h} - S_{2,h} = \Delta_h + (X_{1,h} - X_{2,h})\gamma_2 \tag{1.3}$$

where $\Delta_h = \gamma_1 + (\epsilon_{1,h} - \epsilon_{2,h})$ is the difference between outcomes of the first and second child in household h adjusted for age and gender. Then, γ_1 is the expectation of Δ_h in the population, $\gamma_1 = E(\Delta_h)$, where Δ_h is the heterogeneous birth order effect and γ_1 is the homogenous birth order effects.

Table 1.3 displays estimates of γ_1 for various cognitive and non-cognitive outcomes. [6] I have pooled data from all three waves of the CDS to maximize the number of observations. I have also rescaled each outcome by its respective standard deviation at age seven, the approximate average age in the first wave of the CDS, to facilitate interpretation of the units. [7]

Panel A demonstrates that second-born children score lower than their first-born siblings on the three cognitive outcomes. The effects are 0.26 of a standard deviation on the Letter-Word Identification test, 0.19 of a standard deviation on the Paragraph Comprehension test, and 0.10 of a standard deviation on the Applied Problems test; although, the latter is imprecisely estimated. The effects represent three, four, and two months of development, respectively. Pavan (2015) finds similar magnitudes, 0.15

[6] Appendix A.2 merges the PSID-CDS with the Transition into Adulthood Supplement (TAS) to examine the birth order effects on Adult outcomes, such as college attendance, years of education, and wages.

[7] There is some variation in the standard deviation of the outcomes over the age, but the variation is small. For most outcomes the standard deviation at age seven is also about the average of standard deviations over the age.

to 0.18 of a standard deviation, on cognitive outcomes available in the NLSY.

Panel B demonstrates that second-born children also score lower than their first-born siblings on non-cognitive outcomes, although some of the effects are imprecisely estimated. For example, second-born children score 0.14 and 0.13 of a standard deviation lower on the Dependent and the Peer Problems, respectively. Black *et al.* (2017) documents that second-born boys score 0.11 of a standard deviation lower than their first-born brothers on non-cognitive outcomes using the population of eighteen-year-old Swedish men.

1.4.2 Differences in Parental Investments

Figure 1.1 plots maternal and paternal time by birth order, type (alone or shared), and age. There is a considerable difference in the type of time that first- and second-born children receive over their developments. First-born children receive more maternal and paternal alone time while they receive less maternal and paternal shared time relative to their second-born siblings. This is not surprising given that first-born children enjoy undivided attention of parents until second-born children are born. The differences are also greater early in development, which is a critical period of skill formation in children (Heckman and Masterov, 2007). I quantify the magnitudes using the following model. Let $i \in \{1, 2\}$ denote the order of birth and h denote the household:

$$P_{i,h} = \gamma_1 1\{i = 2\} + \gamma_2 1\{i = 2\} A_{i,h} + \gamma_3 1\{i = 2\} A_{i,h}^2 + X_{i,h}\gamma_4 + \epsilon_{i,h} \tag{1.4}$$

where $P_{i,h}$ is a measure of parental time, $1\{i = 2\}$ is an indicator for the second-born children, $A_{i,h}$ is the child's age, $X_{i,h}$ is a vector that includes gender, age, age squared, mother's age at birth, mother's years of education, and an intercept, and $\epsilon_{i,h}$ is the error term. The birth order effect equals to $\gamma_1 + \gamma_2 A_{i,h} + \gamma_3 A_{i,h}^2$. γ_1 is the effect of being born second on $P_{i,h}$ relative to the first child at birth, while $\gamma_2 A_{i,h} + \gamma_3 A_{i,h}^2$ governs how the effect varies over development. [8]

Table 1.4 displays the results. Columns 1 and 3 demonstrate that second-born children receive 21.2 and 14.8 hours per week less maternal and paternal alone time at birth relative to first-born children while, on average, the effects decrease by 1.0 and 0.8 of an hour per week for every year of development. The average marginal effects of being born second on maternal and paternal alone time are 0.4 and 1.1 of an hour less per week over development, respectively. In contrast, columns 2 and 4 show that second-born children receive 19.6 and 8.2 hours per week more maternal and paternal shared time while, on average, the effects decrease by 0.9 and 0.4 hours per week for every year of development. However, the average of marginal effect is negative; that is, second-born children receive an average of 0.7 and 0.9 of an hour less of maternal and paternal shared time per week over development.

Overall, second-born children receive an average of 1.1 hours per week less maternal time and another 2.0 hours per week less paternal time over development relative to their first-born siblings. [9] On average, mothers and fathers spend a total of 30 hours per week with a child over her development. Therefore, second-born children receive 10% less parental time over development relative to their first-born siblings. Price (2008) using American Time Use Survey documents that parents spend 2.3 hour per week less "quality" time with second-born than the first-born children. [10]

---

[8] Notice that I use OLS instead of FE to identify the model. This is necessary because parental time is a flow variable (as opposed to skill that is a stock variable) and the design of the CDS only allows FE to identify the effect of being born second on parental time for periods that both children live in the household. However, an important part of the comparison between the two children are the periods that the first child is the only child in the household with the corresponding periods in the second child's development.

[9] There are also birth order differences on how children allocate their time among different activities. Appendix A.5 documents that second-born children watch more television relative to their first-born siblings.

[10] Quality parental time is a subset of parental time.

1.4.3 Decomposing Birth Order Effects

In this section, I estimate the portion of the birth order effects on cognitive and non-cognitive outcomes that is explained by the differences in the lagged outcomes and parental investments between children. Let $i \in \{1, 2\}$ denote the order of birth, h denote the household, t denote time, and consider the following "value added" equation:

$$S_{i,h,t+5} = \gamma_1 1\{i = 2\} + X_{i,h,t}\gamma_2 + \gamma_3 S_{i,h,t} + \gamma_4 S'_{i,h,t} + P_{i,h,t}\gamma_5 + \epsilon_{i,h,t} \tag{1.5}$$

where $S_{i,h,t}$ is a cognitive (non-cognitive) outcome, $S'_{i,h,t}$ is a non-cognitive (cognitive) outcome, $P_{i,h,t}$ is a vector of parental investments that includes maternal alone time and maternal shared time, $1\{i = 2\}$ is an indicator for the second-born children, $X_{i,h,t}$ is a vector that includes gender, age, age squared, mother's age at birth, mother's years of education and an intercept, and $\epsilon_{i,h,t}$ is the error term. [11] The coefficient of interest is γ_1, which is the effect of being born second on $S_{i,h,t+5}$ relative to the first child after adjusting for lagged scores and parental investments. Note that two consecutive data collections in the CDS are five years apart, which means that parental investments in periods $t + 1$ to $t + 4$ are omitted variables in equation (1.5).

In estimating equation (1.5), I control for the inherent measurement error in the standardized tests. In particular, for each cognitive (non-cognitive) outcome I use another cognitive (non-cognitive) outcome as an instrument, assuming that the two cognitive (non-cognitive) outcomes are independent measurements of the same latent cognitive (non-cognitive) skills, plus an additively separable measurement error that is uncorrelated across the two outcomes. Similarly, I use paternal time to control for the measurement error in maternal time under the same assumption.

Table 1.5 displays the estimates of γ_1. Columns 1 to 3 show the estimates for the

[11] Notice that I use OLS instead of FE to estimate (1.5). See footnote 8.

three cognitive outcomes: Letter-Word Identification, Passage Comprehension, and Applied Problems. Panel A reports the baseline birth order effects, i.e., not controlling for the lagged outcomes or parental investments. Panel B displays the birth order effects controlling for the lagged outcomes and parental investments. The estimates demonstrate that the differences in lagged cognitive and non-cognitive outcomes and parental investments between siblings account for at least one-half of the birth order effects in Letter-Word Identification and Applied Problems tests and all of the difference in Passage Comprehension test. Pavan (2015) using the NLSY concludes that differences in parental investments approximately explain half of the birth order effects in cognitive outcomes.

Columns 4 and 5 display the estimates for the two non-cognitive outcomes: the Dependent and the Peer Problems scales. Comparison of the estimates in Panel A with the estimates in Panel B demonstrates that the differences in lagged cognitive and non-cognitive outcomes and parental investments between siblings account for one-half of the birth order effects in each of the outcomes. Black *et al.* (2017) using the population of eighteen-year-old Swedish men documents that two-fifth of the birth order effects in non-cognitive outcomes are accounted by the differences in cognitive outcomes. The portion of the birth order effects explained by the lagged outcomes and parental investments is noteworthy given that parental investments in periods $t+1$ to $t+4$ are unobserved and therefore omitted from equation (1.5).

1.4.4 Heterogeneity in Birth Order Effects

In this section, I allow for heterogeneity in the birth order effects. It is important to know whether or not the birth order effects vary across households in any dimension. Documenting heterogeneity helps to formulate conjectures about the underlying causes of the birth order effects.

I allow for heterogeneity by augmenting the baseline specification in equation (1.1). In particular, I include the interaction of the household fixed effects and the second-born children indicator in the model:

$$S_{i,h} = \gamma_0 q_h + \gamma_1 1\{i = 2\} + \gamma_2 q_h 1\{B_{i,h} = 2\} + X_{i,h}\gamma_3 + \epsilon_{i,h} \tag{1.6}$$

where $S_{i,h}$ is the child's outcome, q_h is the household fixed effect, $1\{i = 2\}$ is an indicator for the second-born children, $X_{i,h}$ is a vector that includes child specific variables such as gender, age, age squared, and an intercept, and $\epsilon_{i,h}$ is the error term. The birth order effect equals to $\gamma_1 + \gamma_2 q_h$. γ_1 is the average effect of being born second on $S_{i,h}$ relative to the first child, or the average birth order effect. γ_2 is the effect of one standard deviation increase in the household fixed effect on the birth order effect relative to the average effect. If $\gamma_2 = 0$, then there is no heterogeneity and the effect is the same across households. However, if $\gamma_2 > 0$ (< 0), then the birth order effect is heterogeneous and is smaller in households with a higher (lower) q_h. Note that equation (1.6) only determines whether or not birth order effects are heterogeneous. To characterize the attributes of households that have smaller birth order effects, it is necessary to go one step further and regress the estimated household fixed effects on observable characteristics of the households. This last step suffers from the standard omitted variable bias.

It is possible to model heterogeneity by observables. Modeling heterogeneity by unobservables is a superior method in the sense that it is robust to the omitted variable bias and attenuation bias. However, it is not trivial to estimate specification (1.6). The model is non-linear in the household fixed effects, q_h, and therefore, the within-household transformation cannot identify the model. But, one can recover the coefficients and the distribution of the fixed effects by dividing the estimation into two steps. In the first step, given a guess for the household fixed effects, estimate the coefficients using ordinary least squares. In the second step, given the coefficients, estimate the household fixed effects and iterate until two consecutive estimates of the coefficients are sufficiently close. Under the assumption that $E[\epsilon_{i,h}|q_h] = 0$, the iterative procedure is a consistent estimator of the coefficients of interest γ_1 and γ_2 (Arcidiacono *et al.*, 2012).

Table 1.6 presents the estimates. The average birth order effect on cognitive (Letter-Word Identification test) and non-cognitive (Dependent scale) outcomes is 0.26 and 0.14 of a standard deviation, respectively. In addition, the table displays that the coefficient of the interaction term is positive which means that the birth order effect is smaller in households with a higher fixed effects, q_h. One standard deviation increase in the household fixed effects decreases the birth order effect on cognitive and non-cognitive outcomes by 0.08 and 0.17 of a standard deviation, respectively. The average and the standard deviation of distribution of the marginal effects are both displayed at the bottom of the table. Figure 1.2 plots the distribution of the birth order effects against the household fixed effects percentiles. The birth order effects on cognitive and non-cognitive outcomes at the tenth-percentile of the distribution are 0.36 and 0.38 of a standard deviation while the corresponding effects at the ninetieth-percentile of the distribution are 0.18 and 0.01 of a standard deviation.

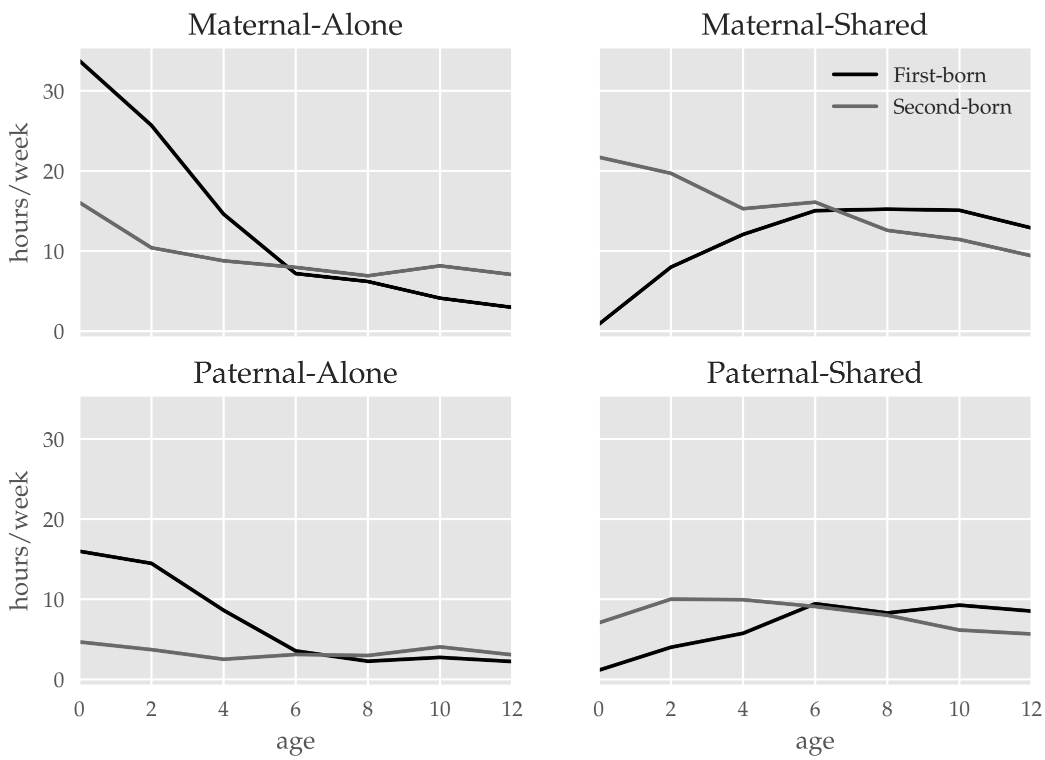
Table 1.7 displays the regressions of estimated household's cognitive and non-cognitive fixed effects on household's observable characteristics. The table demonstrates that presence of both biological parents, an increase in mother's years of education, and an increase in household's income is associated with a smaller birth order effects. In contrast, an increase in household's number of boys or an increase in household's birth spacing is associated with a larger birth order effects. Buckles and Munnich (2012) using the NLSY document that an increase in birth spacing increases the differences in outcomes of the siblings. Note that these regressions characterize the attributes of households with smaller/larger birth order effects, but they suffer from the standard omitted variable bias. For example, a household with a higher income may have parents with higher ability. But ability is not observed and if it is correlated with income, then the estimates will be biased and inconsistent. Therefore, the magnitudes should not be interpreted as the causal effects.

1.5 Conclusions

In this chapter, I document that first-born children score 0.2 of a standard deviation higher on both cognitive and non-cognitive outcomes than their second-born siblings. Furthermore, first-born children receive 0.1 of a standard deviation more in parental investments. The birth order effects on parental time account for more than half of the birth order effects on outcomes. In addition, I document that birth order effects are heterogeneous and vary between 0.1 and 0.4 of a standard deviation with the effects being smaller in households with certain characteristics such as a high income.

1.6 Figures

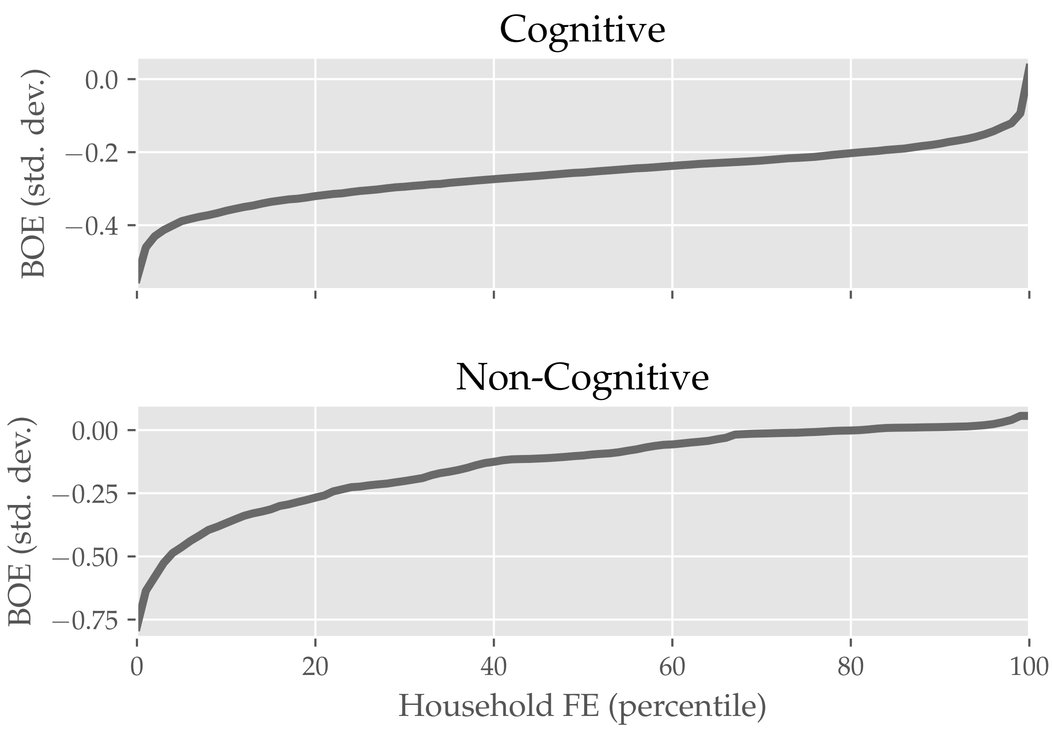
Figure 1.1: Active Parental Time by Birth Order and Age in 1997



Source: the PSID-CDS.

Notes: Alone time denotes absence of a sibling while shared time denotes presence of a sibling.

Figure 1.2: Heterogeneity in Birth Order Effects



Source: the PSID-CDS

Note: BOE denotes the birth order effect and is defined as the difference between the second and the first child scores.

Table 1.1: Descriptive Statistics in 1997

	Mean	Std. Dev.
A: Characteristics of Children		
Number of children	1258	
Age	6.65	3.73
Fraction male	0.50	
Fraction White, Non-Hispanic	0.57	
Fraction Black, Non-Hispanic	0.33	
Fraction Hispanic	0.05	
Fraction other Race/Ethnicity	0.04	
Fraction living with both biological parents	0.69	
Fraction living with only biological mother	0.28	
Fraction living with only biological father	0.02	
Fraction living with no biological parents	0.01	
B: Characteristics of Households		
Number of households	862	
Mother's age	33.40	6.82
Mother's age at first pregnancy	24.13	5.06
Mother's years of schooling	13.96	2.09
Mother's hours worked per week	24.79	17.06
Household income (10,000 1996USD)	5.00	4.42

Source: the PSID-CDS.

Table 1.2: Children Outcomes and Parental Investments in 1997

	Obs	Mean	Std. Dev.	Min	Max
A: Cognitive Outcomes					
Letter-Word Identification	777	28.28	16.65	0	57
Paragraph Comprehension	518	21.36	8.11	0	39
Applied Problems	773	25.05	11.87	0	53
B: Non-Cognitive Outcomes					
Dependent	976	2.59	0.43	1	3
Peer Problems	976	2.86	0.28	1	3
C: Parental Time					
Maternal Alone	1002	10.84	14.56	0	77
Maternal Shared	1002	13.54	12.52	0	70
Paternal Alone	1002	5.04	9.59	0	77
Paternal Shared	1002	7.51	9.13	0	51

Source: the PSID-CDS.

Notes: Cognitive outcomes are the raw scores. Non-cognitive outcomes are rated by parents on a scale of 1 to 3; 1=Often true, 2=Sometimes true, 3=Not true. Parental time is measured in hours/week.

Table 1.3: Effect of Being Born Second on Outcomes

Outcome	OLS	FE
A: Cognitive Outcomes		
Letter-Word Identification	-0.121***	-0.259***
	(0.037) [2251]	(0.074) [2251]
Paragraph Comprehension	-0.080**	-0.190**
	(0.038) [1960]	(0.077) [1960]
Applied Problems	-0.154***	-0.104
	(0.057) [2244]	(0.100) [2244]
F-statistic	9.42***	8.59***
B: Non-Cognitive Outcomes		
Dependent	-0.083**	-0.136*
	(0.041) [2586]	(0.081) [2586]
Peer Problems	-0.069	-0.128
	(0.046) [2586]	(0.101) [2586]
F-statistic	3.67**	2.61*

Source: the PSID-CDS.

Notes: Each entry is the estimate from a separate regression. All outcomes are rescaled by standard deviation at age seven. The scale of non-cognitive outcomes is reversed so that a higher score is a better outcome. Models control for child's gender, child's age and age squared, mother's years of education, and mother's age at birth. Number of children are in brackets. Household-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively. The null hypothesis for the F-test is that being born second has zero effect on all of the outcomes.

Table 1.4: Effect of Being Born Second on Active Parental Time

	Maternal		Paternal	
	Alone	Shared	Alone	Shared
1{birth order=2}	-21.20***	19.61***	-14.82***	8.22***
	(2.287)	(1.500)	(1.750)	(1.051)
1{birth order=2} $\times$ age	3.62***	-3.66***	2.37***	-1.61***
	(0.417)	(0.308)	(0.310)	(0.223)
1{birth order=2} $\times$ age^2	-0.13***	0.14***	-0.08***	0.06***
	(0.018)	(0.015)	(0.013)	(0.011)
Average Marginal Effects	-0.43	-0.69	-1.12	-0.93
N-Children	2435	2435	2435	2435
N-Households	1744	1744	1744	1744

Source: the PSID-CDS.

Notes: Time is measured in hours per week. Models control for child's gender, child's age and age squared, mother's years of education, and mother's age at birth. Household-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 1.5: Decomposing Birth Order Effects on Outcomes

	(1)	(2)	(3)	(4)	(5)
A: Baseline					
1{birth order=2}	-0.156***	-0.102*	-0.139**	-0.110**	-0.063
	(0.055)	(0.052)	(0.057)	(0.050)	(0.054)
B: Adding Parental Investments					
1{birth order=2}	-0.082	-0.005	-0.072	-0.033	0.010
	(0.056)	(0.048)	(0.057)	(0.046)	(0.049)
N-Children	1474	1442	1471	1610	1610
N-Households	697	692	696	744	744

Source: the PSID-CDS.

Notes: Each entry is the estimate from a separate regression. Cognitive outcomes are: 1=Letter-Word Identification, 2=Paragraph Comprehension, 3=Applied Problems. Non-cognitive outcomes are: 4=Dependent, 5=Peer Problems. The scale of non-cognitive outcomes is reversed so that a higher score is a better outcome. All outcomes are rescaled by standard deviation at age seven. Models control for child's gender, child's age and age squared, mother's years of education, and mother's age at birth. Household-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 1.6: Heterogeneity in Birth Order Effects

	Cognitive	Non-Cognitive
A: Estimates		
1{birth order=2}	-0.264***	-0.143*
	(0.068)	(0.087)
	[-0.391,-0.160]	[-0.290,-0.023]
$q \times$ 1{birth order=2}	0.076*	0.166***
	(0.043)	(0.056)
	[0.009,0.150]	[0.096,0.279]
B: Marginal Effects		
Average	-0.26	-0.14
Standard Deviation	0.08	0.17
N-Children	2251	2586
N-Households	1529	1747

Source: the PSID-CDS.

Notes: Cognitive outcome is Letter-Word identification test. Non-Cognitive outcome is Dependent scale. The scale of the non-cognitive outcome is reversed so that a higher score is a better outcome. Outcomes are rescaled by standard deviation at age seven. q is the household fixed effect and is normalized to have mean of zero standard deviation of one. Regressions control for child's gender, child's age, and child's age squared. Bootstrapped standard errors are in parentheses. 90% bootstrapped confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 1.7: Determinants of Households Fixed Effects

	Cognitive	Non-Cognitive
1{number of boys=1}	-0.095	-0.127**
	(0.062)	(0.060)
1{number of boys=2}	-0.167**	-0.130*
	(0.072)	(0.069)
1{household is intact=1}	0.277***	0.311***
	(0.056)	(0.053)
spacing (years)	0.003	-0.018***
	(0.007)	(0.007)
mother's education (years)	0.123***	0.062***
	(0.013)	(0.012)
household's income (10,000 USD)	0.015***	0.007*
	(0.004)	(0.004)
N-Households	1529	1747

Source: the PSID-CDS.

Notes: Regressions are at the household level. Regressions include missing indicators. Standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Chapter 2

THE EFFECT OF CASH TRANSFERS ON BIRTH ORDER EFFECTS

2.1 Introduction

Chapter 1 demonstrates that there are birth order effects on cognitive and non-cognitive outcomes during childhood development. A natural question that follows this finding is that how does birth order affect a child's outcome? And perhaps, just as importantly, how can we alleviate the birth order effects? Although economists have long examined the implications of intra-household allocation of resources among partners, for example, consumption goods (Lise and Seitz, 2011; Browning *et al.*, 1994), or market and house work (Álvarez and Miles, 2003; Hersch and Stratton, 1994), they have only recently begun to analyze the implications of intra-household allocation of resources among children (Pavan, 2015; Del Boca *et al.*, 2013, 2016). This chapter contributes to the growing literature by developing a model of intra-household resource allocation (time and money) that endogenously generates the birth order effects. My modeling framework also allows estimation of counterfactual policy experiments in which the government uses targeted transfers to reduce intra-household inequality.

The model has a life-cycle framework in which a household with two children confronts a sequence of time constraints and a lifetime monetary constraint, and divides the available time and monetary resources between consumption and investment. The birth order effects arise in the model as a result of two theoretical and one empirical feature. The theoretical features include: time is a resource that cannot be saved or borrowed, and that, the first-born child has the undivided attention of the household early in development while the second-born child has the same opportunity late in development. What gives importance to the theoretical features is the empirical feature. Productivity of investments are higher early in development (Heckman and Masterov, 2007). Heterogeneity in the birth order effects arises as a result of heterogeneity in household's income, concavity of preferences, and concavity of the technologies of skill formation (diminishing marginal products of investments). Intuitively, households invest more monetary resources in their children as their income increases, but these additional investments are more productive for the second-born child relative to the first-born child because they have lower skills.

The model builds on Del Boca *et al.* (2013, 2016). I augment their model in two dimensions which are important in understanding the birth order effects. First, I allow for saving and borrowing which is a margin that enables the household to transfer money across time and between siblings. Second, I incorporate non-cognitive skills. A closely related study to my model is Pavan (2015) which uses a dynamic latent factor model to estimate the portion of birth order effects in cognitive outcomes that is explained by parental investments. However, the author does not explicitly model household choices. Therefore, the setup is unable to address the counterfactual policy experiments that I have in mind. My counterfactual experiments demonstrate that income transfers reduce the birth order effects, but they are most effective when they target low-income households. The model shows that an annual income transfer of 10,000 USD to low-income households decreases the birth order effects on cognitive and non-cognitive skills by one-sixth, which is five times bigger than the effect in high-income household.

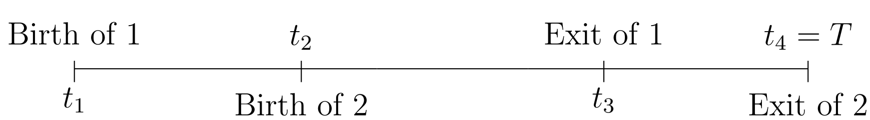
Subsequent sections of this chapter are organized as follows. Section 2.2 introduces a life-cycle model of endogenous birth order effects. Section 2.3 elaborates the econometric details. Section 2.4 discusses the results and counterfactual policy experiments. Section 2.5 concludes the chapter.

2.2 Model

In this section, I develop a life-cycle model of intra-household resource allocation that endogenously generates the heterogeneous birth order effects with the goal of using the model for counterfactual policy experiments. This model rationalizes the two empirical facts that I established in Chapter 1: there is a birth order effect in cognitive and non-cognitive outcomes which negatively affects second-born children, and that, these effects are smaller in households with certain characteristics such as a high income. The following subsections elaborates on the details of the model.

2.2.1 Environment

The model is parsimonious and focuses on intra-household allocation of resources. For simplicity, the model considers households that have two children. Households live for T periods. Their life begins at $t_1 = 1$ with the birth of the first child. At t_2, the second child joins the unit. At $t_3 = M$, development of the first child ends and they leave the unit. Finally, at $t_4 = t_2 + M - 1 = T$, development of the second child ends and households die. The timeline of the events is:



Households care about consumption, “leisure”, and skills of their children which are composed of cognitive and non-cognitive dimensions. They own age-specific technologies that combine the initial skills, time investments, and monetary investments into the next period’s skills. The households are endowed with a unit of time and an exogenous income. They allocate the time endowment among leisure, time investments in the first child alone, time investments in the second child alone, or time investments in both children concurrently. In addition, the households allocate the income endowment among consumption, monetary investments in the first child, or monetary investments in the second child. The households can also borrow or save at a competitive interest rate, which enables them to transfer monetary resources between periods. Households are homogeneous with respect to preferences, technologies, and time endowment, but they are heterogeneous in income endowment.

This model improves upon the literature by adding non-cognitive skills and allowing households to borrow or save (Pavan, 2015; Del Boca *et al.*, 2013). Non-cognitive skills are as important as cognitive skills in explaining adult outcomes (Heckman *et al.*, 2006; Heckman and Rubinstein, 2001). Borrowing or saving can be an important margin for households to transfer income across time and between children, which is important for the question at hand. The model takes fertility as an exogenous variable similar to the literature. Fertility decisions are potentially important for the results since households know about the birth order effects and they may choose to space their children differently in response to the effects (Buckles and Munnich, 2012; Rosenzweig, 1986).

2.2.2 Preferences

Households have a "parent" decision maker which has preferences over consumption, leisure and skills of children. Their utility flow at time t has a nested Cobb-Douglas form given by:

$$u(c_t, \ell_t, \theta^1_{t+1}, \theta^2_{t+1}) = \alpha_1 \log c_t + \alpha_2 \log \ell_t + \alpha_3 \log \theta^1_{t+1} + \alpha_4 \log \theta^2_{t+1} \tag{2.1}$$

where c is the parent's consumption, ℓ is the parent's leisure (non-child time), θ^1 is the skills of the first child, and θ^2 is the skills of the second child. I assume that $\alpha_j > 0$

for $j \in \{1, 2, 3, 4\}$, which along with additively separable logarithmic specification ensures that the utility function is strictly increasing in each of its arguments and jointly strictly concave in all of its arguments. Without any loss of generality, I normalize the scale of utility to 1, $\sum_{j=1}^{4} \alpha_j = 1$.

Skills are composed of cognitive and non-cognitive dimensions. I assume that households preferences over the two dimensions is characterized by the following Cobb-Douglas form. Let $i \in \{1, 2\}$ denote the order of birth, then:

$$\log \theta_{t+1}^{i} = \alpha_5 \log k_{t+1}^{i} + (1 - \alpha_5) \log n_{t+1}^{i} \qquad i = 1, 2 \tag{2.2}$$

where k is the cognitive skills, n is the non-cognitive skills, and $\alpha_5 \in (0, 1)$ is the preference for cognitive skills. Note that α_5 does not depend on i; that is, I assume that households value the skills identically for both children. This assumption is to reduce the number of parameters to estimate and can be relaxed at the cost of additional computational burden.

2.2.3 Technologies

Households invest time and money in the cognitive and non-cognitive skills of children. Investments, along with the current level of skills, produce the next level of skills according to the technologies of skill formation. Let t_1 and t_2 denote the birth periods for the first and second child. Without any loss of generality, set $t_1 = 1$ and let $t' = t - t_2$. Cognitive skills evolve according to the following piece-wise technology:

$$k_{t+1}^{1} = \begin{cases} (k_t^1)^{\delta_{1,t}} \, (e_t^1)^{\delta_{2,t}} \, (a_t^1)^{\delta_{3,t}} & t = 1, \cdots, t_2 - 1 \\ \\ (k_t^1)^{\delta_{1,t}} \, (e_t^1)^{\delta_{2,t}} \, (a_t^1)^{\delta_{3,t}} \, (s_t)^{\delta_{4,t}} & t = t_2, \cdots, M \\ \\ k_{M+1}^{1} & t = M + 1, \cdots, T \end{cases} \tag{2.3}$$

$$k^2_{t+1} = \begin{cases} k^2_{t_2} & t = 1, \cdots, t_2 - 1 \\ (k^2_t)^{\delta_{1,t'}} \, (e^2_t)^{\delta_{2,t'}} \, (a^2_t)^{\delta_{3,t'}} \, (s_t)^{\delta_{4,t'}} & t = t_2, \cdots, M \\ (k^2_t)^{\delta_{1,t'}} \, (e^2_t)^{\delta_{2,t'}} \, (a^2_t)^{\delta_{3,t'}} & t = M+1, \cdots, T \end{cases} \tag{2.4}$$

where e is the monetary investment, a is the alone time investment (time that is spent with one child alone), and s is the shared time investment (time that is spent with both children concurrently). Note that alone time investment is an imperfect substitute for the shared time investment. This modeling choice follows the findings in Chapter 1. $\delta_{j,t}$, $j \in \{1, 2, 3, 4\}$, is the elasticity of cognitive skills with respect to input j, which depends on the child's age, t. That is, the same level of investment produces a different level of skill at different ages. This assumption is supported by evidence and is widely used in the literature (Heckman and Masterov, 2007).

Similarly, non-cognitive skills evolve according to the following technology:

$$n^1_{t+1} = \begin{cases} (n^1_t)^{\eta_{1,t}} \, (e^1_t)^{\eta_{2,t}} \, (a^1_t)^{\eta_{3,t}} & t = 1, \cdots, t_2 - 1 \\ (n^1_t)^{\eta_{1,t}} \, (e^1_t)^{\eta_{2,t}} \, (a^1_t)^{\eta_{3,t}} \, (s_t)^{\eta_{4,t}} & t = t_2, \cdots, M \\ n^1_{M+1} & t = M+1, \cdots, T \end{cases} \tag{2.5}$$

$$n^2_{t+1} = \begin{cases} n^2_{t_2} & t = 1, \cdots, t_2 - 1 \\ (n^2_t)^{\eta_{1,t'}} \, (e^2_t)^{\eta_{2,t'}} \, (a^2_t)^{\eta_{3,t'}} \, (s_t)^{\eta_{4,t'}} & t = t_2, \cdots, M \\ (n^2_t)^{\eta_{1,t'}} \, (e^2_t)^{\eta_{2,t'}} \, (a^2_t)^{\eta_{3,t'}} & t = M+1, \cdots, T \end{cases} \tag{2.6}$$

where $\eta_{j,t}$, $j \in \{1, 2, 3, 4\}$, is the elasticity of non-cognitive skills with respect to input j at age t.

2.2.4 Households' Problems

Households maximize the sum of discounted utility subject to a lifetime monetary budget constraint, sequence of time budget constraints, and cognitive and non-cognitive skill formation technologies. The household's problems in the sequence form is:

$$\max_{\substack{\{c_t\}_{t=1}^{T},\{e_t^1\}_{t=1}^{M},\{e_t^2\}_{t=t_2}^{T}\\ \{\ell_t\}_{t=1}^{T},\{a_t^1\}_{t=1}^{M},\{a_t^2\}_{t=t_2}^{T},\{s_t\}_{t=t_2}^{M}}} \sum_{t=1}^{T} \beta^{t-1} u(c_t, \ell_t, k_{t+1}^1, n_{t+1}^1, k_{t+1}^2, n_{t+1}^2)$$

$$s.t. \quad \sum_{t=1}^{T} R^{-(t-1)} \big(c_t + \mathbf{1}\{t \le M\} e_t^1 + \mathbf{1}\{t \ge t_2\} e_t^2\big) = \sum_{t=1}^{T} R^{-(t-1)} \mathrm{I}_t \tag{2.7}$$

$$\ell_t + \mathbf{1}\{t \le M\} a_t^1 + \mathbf{1}\{t \ge t_2\} a_t^2 + \mathbf{1}\{t_2 \le t \le M\} s_t = \tau \; \forall t$$

where β is the discount factor, R is the competitive gross interest rate, I is the household's income, τ is the periodic time endowment, and $\mathbf{1}\{.\}$ is the indicator function. Note that households have perfect information about their income profile and can fully borrow against their future income. The model abstracts from borrowing constraints or uncertainty about future income in order to make an important first step toward incorporating saving decisions. [1]

This model has a straightforward closed-form solution which is built on four assumptions: perfect information, frictionless borrowing, log preferences, and log technologies. Perfect information and frictionless borrowing imply that optimal decisions depend on the present value of the household's lifetime income, not on the household's income profile. These assumptions greatly reduce the computational burden of estimating the model. Additionally, log preferences and log technologies imply

[1] It is not possible to determine whether the model provides a lower or an upper bound for the effect of household income on allocation of resources between siblings. For example, all else equal, the allocation in the presence of borrowing constraints is a function of household's income profile which depending on its shape can make the allocation more or less equal relative to the allocation under no borrowing constraints.

that optimal decisions are interior and independent of children's cognitive and non-cognitive skills in all periods. These assumptions are for tractability and are sufficient to have a closed-form solution. Even though that the model is simple, it captures important empirical patterns, specifically, the age profile of cognitive and non-cognitive skills, time investments, and the birth order effects.

2.2.5 Optimal Decisions

Households allocate their present value of lifetime income among consumption, monetary investments in the first child, and monetary investments in the second child such that the allocation equates the marginal benefits across all three uses in all periods. The log preferences together with log technologies imply that in the optimal allocation, the ratio of any two uses is equal to the ratio of their lifetime marginal benefits. The optimal decision for $y \in \{c_t, e_t^1, e_t^2\}$ and all t is:

$$y = \frac{\varphi(y)}{\sum_{t=1}^{T}[\varphi(c_t) + \varphi(e_t^1) + \varphi(e_t^2)]} \sum_{t=1}^{T} R^{-(t-1)} \mathrm{I}_t \tag{2.8}$$

where $\varphi(y)$ is the lifetime marginal benefits of choice y. Note that the lifetime marginal benefits depend on current and future utility flows. In particular, households take into account that consumption yields only a current utility flow while monetary investments in children result in future utility flows as well.

$$\varphi(c_t) = \beta^{t-1} R^{t-1} \alpha_1 \tag{2.9}$$

$$\varphi(e_t^1) = \mathbf{1}\{t \leq M\} \beta^{t-1} R^{t-1} \big(\alpha_3 \alpha_5 \delta_{2,t}(1 + \phi_t^1) + \alpha_3(1 - \alpha_5)\eta_{2,t}(1 + \psi_t^1)\big) \tag{2.10}$$

$$\varphi(e_t^2) = \mathbf{1}\{t \geq t_2\} \beta^{t-1} R^{t-1} \big(\alpha_4 \alpha_5 \delta_{2,t'}(1 + \phi_t^2) + \alpha_4(1 - \alpha_5)\eta_{2,t'}(1 + \psi_t^2)\big) \tag{2.11}$$

where ϕ_t^i and ψ_t^i are the current value of future benefits of an additional investment in cognitive and non-cognitive skills of child i at time t, respectively. In the estimation of the model, I simplify the analysis further by assuming $\beta = R^{-1}$. This assumption cancels R out of the optimal decisions.

$$\phi_t^1 = \mathbf{1}\{t \le M\} \sum_{i=t+1}^{T} \beta^{i-t} \prod_{j=t+1}^{i} \delta_{1,j} \tag{2.12}$$

$$\phi_t^2 = \mathbf{1}\{t_2 \le t < T\} \sum_{i=t+1}^{T} \beta^{i-t} \prod_{j=t+1}^{i} \delta_{1,j'} \tag{2.13}$$

$$\psi_t^1 = \mathbf{1}\{t \le M\} \sum_{i=t+1}^{T} \beta^{i-t} \prod_{j=t+1}^{i} \eta_{1,j} \tag{2.14}$$

$$\psi_t^2 = \mathbf{1}\{t_2 \le t < T\} \sum_{i=t+1}^{T} \beta^{i-t} \prod_{j=t+1}^{i} \eta_{1,j'} \tag{2.15}$$

Similarly, in each period, households allocate their time endowment among leisure, time investment in the first child alone, time investment in the second child alone, and time investment in both children concurrently, such that the allocation equates lifetime marginal benefits across all four uses. The log preferences together with log technologies imply that in the optimal allocation, the ratio of any two uses is equal to the ratio of their lifetime marginal benefits. The optimal decision for $y \in \{\ell_t, a_t^1, a_t^2, s_t\}$ and all t is:

$$y = \frac{\varphi(y)}{\varphi(\ell_t) + \varphi(a_t^1) + \varphi(a_t^2) + \varphi(s_t)} \tau \tag{2.16}$$

where $\varphi(y)$ is the lifetime marginal benefits of choice y:

$$\varphi(\ell_t) = \alpha_2 \tag{2.17}$$

$$\varphi(a_t^1) = \mathbf{1}\{t \le M\} \big(\alpha_3 \alpha_5 \delta_{3,t} (1 + \phi_t^1) + \alpha_3 (1 - \alpha_5) \eta_{3,t} (1 + \psi_t^1)\big) \tag{2.18}$$

$$\varphi(a_t^2) = \mathbf{1}\{t \ge t_2\} \big(\alpha_4 \alpha_5 \delta_{3,t'} (1 + \phi_t^2) + \alpha_4 (1 - \alpha_5) \eta_{3,t'} (1 + \psi_t^2)\big) \tag{2.19}$$

$$\varphi(s_t^1) = \mathbf{1}\{t_2 \leq t \leq M\}\big(\alpha_3\alpha_5\delta_{3,t}(1+\phi_t^1) + \alpha_3(1-\alpha_5)\eta_{3,t}(1+\psi_t^1)\big) \tag{2.20}$$

$$\varphi(s_t^2) = \mathbf{1}\{t_2 \leq t \leq M\}\big(\alpha_4\alpha_5\delta_{3,t'}(1+\phi_t^2) + \alpha_4(1-\alpha_5)\eta_{3,t'}(1+\psi_t^2)\big) \tag{2.21}$$

$$\varphi(s_t) = \varphi(s_t^1) + \varphi(s_t^2) \tag{2.22}$$

Note that time (monetary) investment, among other things, depends on the elasticity of skills with respect to time (monetary) investment. For example, if the elasticity of skills with respect to time (monetary) investment is high early in development and low late in development, then, all else equal, the household optimally invests more during the early years of development relative to the late years.

2.2.6 Discussion

It is constructive to describe the components of the model that differentially affect the skills of the two children. Let K^i and N^i denote the cognitive and non-cognitive skills of child with birth order of $i \in \{1, 2\}$ at the end of development. Then, the birth order effects on cognitive and non-cognitive skills are:

$$K^2 - K^1 = \Delta K\Big(\{\alpha_j\}_{j\in\{1,2,3,4,5\}}, \{\{\delta_{j,t}, \eta_{j,t}\}_{t=1}^T\}_{j\in\{1,2,3,4\}}, \sum R^{-(t-1)}\mathrm{I}_t\Big) \tag{2.23}$$

$$N^2 - N^1 = \Delta N\Big(\{\alpha_j\}_{j\in\{1,2,3,4,5\}}, \{\{\delta_{j,t}, \eta_{j,t}\}_{t=1}^T\}_{j\in\{1,2,3,4\}}, \sum R^{-(t-1)}\mathrm{I}_t\Big) \tag{2.24}$$

The first component is preferences. The household may have unequal preferences for the first and second child's skills. Then, all else equal, differences in preference weights will result in differences in investment and therefore, differences in skills between children.

The second component is elasticities. The cognitive and non-cognitive technologies are asymmetric for the two children. Skills of the first child evolve according to one-child technologies until the second child arrives. At this point, skills of both children evolve according to two-child technologies until the first child's development ends. Then, skills of the second child evolve according to one-child technologies. The elasticities of one-child technologies may be different from the elasticities of two-child technologies. Then, all else equal, identical investments would produce differences in skills.

The third component is time constraints. Time constraints can cause differences in investments over the development years which would, all else equal, cause differential skills in the children. This is because time is a resource that cannot be saved or borrowed. The first child enjoys the undivided attention of the household early in development while the second child has the same opportunity late in development.

The last component is income endowments. Income is the only source of heterogeneity in the model, which can either alleviate or exacerbate the birth order effects depending on the preferences and the technologies. Given that the household can save or borrow without friction, then, all else equal, the concavity of the preferences and the technologies imply that a higher household income results in smaller differences in skills between children.

2.3 Estimation

In this section, I discuss details of the assumptions regarding the identification and the estimation of the model. The estimator is based on the Method of Moments. Supplementary Monte Carlo experiments are included in Appendix B.3, which check the validity of the estimator.

2.3.1 Measurements

Available information in the PSID-CDS is sufficient to identify the model given the structure that is imposed upon it. However, the data is not perfect. In particular, household income is observed annually until 1997 and biennially thereafter. Cognitive skills, non-cognitive skills, and time investments are observed every five years. Monetary investments are not observed at all. Below, I describe the measurement of skills, time investments, and household income.

The cognitive and non-cognitive skills are set to the Letter-Word identification score and the Dependent score, respectively. Alone time investment is set to the union of maternal alone time and paternal alone time. Similarly, shared time investment is set to the union of maternal shared time and paternal shared time. Finally, current value of household's lifetime income in the initial period, i.e. first child's birth year, is computed based on the following algorithm. First, deflate the observed income using CPI. Then, compute the current value of the real income in the initial period using a gross interest rate of $R = 1/\beta$. Next, take the average as a measure of current value of the household's annual income and multiply it by the number of periods in the household's lifetime to obtain the current value of household's lifetime income at the initial period. This procedure is equivalent to estimating the household's permanent income using a family fixed effects model.

2.3.2 Estimator

The estimation strategy follows Del Boca *et al.* (2013, 2016). I set the number of development periods, M, to 17, and the discount factor, β, to 0.95. There are several functions to estimate: the utility function, the technologies of cognitive and non-cognitive skills when one child lives in the household, and the technologies of cognitive and non-cognitive skills when two children live in the household.

The preference parameters, $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\}$, are constrained to be greater than zero and smaller than one. [2] I use the following exponential transformations to

[2]The weight on cognitive skills, α_5, is identified under the assumption that the technology of cog-

impose the constraints.

$$\alpha_j = \frac{\exp(\omega_{\alpha_j})}{1+\sum_{i=2}^{4}\exp(\omega_{\alpha_j})} \quad j \in \{2,3,4\} \tag{2.25}$$

$$\alpha_5 = \frac{\exp(\omega_{\alpha_5})}{1+\exp(\omega_{\alpha_5})} \tag{2.26}$$

where ω's are parameters to be estimated. Note that α_1 is determined by the normalization of the scale of the utility, $\sum_{j=1}^{4}\alpha_j = 1$.

The cognitive technology parameters, $\{\delta_{1,t}, \delta_{2,t}, \delta_{3,t}, \delta_{4,t}\}$, and the non-cognitive technology parameters, $\{\eta_{1,t}, \eta_{2,t}, \eta_{3,t}, \eta_{4,t}\}$, are age dependent. In other words, the elasticity of cognitive and non-cognitive skills with respect to different inputs depends on the child's age. I assume that the elasticities are log-linear in child's age in order to economize on the number of parameters to estimate.

$$\delta_{j,t} = \exp(\omega_{\delta_j,0} + \omega_{\delta_j,1}t) \quad j \in \{1,2,3,4\} \tag{2.27}$$

$$\eta_{j,t} = \exp(\omega_{\eta_j,0} + \omega_{\eta_j,1}t) \quad j \in \{1,2,3,4\} \tag{2.28}$$

where t is age of the child, and $\omega_{\delta_j,0}$ and $\omega_{\delta_j,1}$ ($\omega_{\eta_j,0}$ and $\omega_{\eta_j,1}$) are the intercept and the slope for the age profile of log elasticity of cognitive (non-cognitive) skills with respect to input j to be estimated.

There are 32 parameters to estimate given the assumptions: 4 parameters for preferences, 12 parameters for technologies of cognitive and non-cognitive skills when one child lives in the household, and 16 parameters for technologies of cognitive and non-cognitive skills when two children live in the household. In order to simplify the estimation, I assume that the technologies of skill formation when one child lives in the household, meaning the periods that second child is not born yet or periods that the first child has moved out of the household, are identical to the technologies of

nitive skills is different from the technology of non-cognitive skills. If the technologies are identical, then all α_5s are observationally equivalent.

skill formation for households with one child. Therefore, I divide the estimation of the model into two steps. In the first step, I use data on households with one child and estimate the technology of cognitive and non-cognitive skills when one child lives in the household. [3] In the second step, I take the technologies when one child lives in the household as given, and I estimate the technologies when two children live in the household and the preferences.

The estimator is based on the Method of Moments since the children's skills are observed every five years and skills between the two measurements require simulation. The algorithm for the Method of Moments works as follows. Start with a sample of households at time t. Then, pick an initial value for the vector of primitive parameters of the model, ω. Next, solve the household's problem for each of the households in the sample. Then, use the household's optimal decisions to simulate children's cognitive and non-cognitive skills from $t+1$ to $t+5$. Finally, compute moments from the simulated and the actual data. The Method of Moments estimator of ω is:

$$\widehat{\omega} = \arg\min_{\omega}(\Omega - \widehat{\Omega}(\omega))' W (\Omega - \widehat{\Omega}(\omega)) \tag{2.29}$$

where Ω is the vector of moments computed from the sample, $\widehat{\Omega}(\omega)$ is the vector of moments computed from the simulated data generated under the parameter vector ω, and W is a symmetric positive-definite weighting matrix.

The moments that identify the model are the mean of cognitive scores conditional on age and birth order, the mean of non-cognitive scores conditional on age and birth order, the mean of alone time investments conditional on age and birth order, the mean of shared time investments conditional on age, the covariance of household income and cognitive scores conditional on age and birth order, and the covariance of household income and non-cognitive scores conditional on age and birth order.

[3] Appendix B.2 describes the model with one child, which is analogous to the model with two children.

2.4 Results

This section discusses the estimates and the counterfactual experiments. The counterfactual experiments show that income transfers reduce the birth order effects, however, they are most effective when they target low-income households. Therefore, the paper documents another efficiency benefit for policies that target children in low-income households.

2.4.1 Estimates

Figures 2.1 and 2.2 demonstrate the model's fit. Figure 2.1 displays the fit for average cognitive skills, average non-cognitive skills, and average alone time investments for the first-born children conditional on age. Figure 2.2 replicates the same moments for the second-born children. The figures demonstrate that the model matches age-profiles of skills and time investments well. In particular, the model is able to replicate the increasing pattern of cognitive and non-cognitive skills and decreasing pattern of time investments in age.

Figures 2.3 and 2.4 plot the estimates of the elasticities of cognitive skills with respect to different inputs for the first and second child, respectively, in households with birth spacing of one year. The elasticity of cognitive skills with respect to monetary and time investment are decreasing in age. Monetary and time investments early in development are more productive in comparison with investments late in development. For example, the elasticity of cognitive skills with respect to alone time investments decreases from approximately 1 at birth to 0.1 at age sixteen for the first child. The kinks are the result of a child entering or exiting the household. For example, if the second child enters the household when the first child is one year old, the elasticity of cognitive skills with respect to alone time investments drops from 1 to 0.5 while the elasticity of cognitive skills with respect to shared time investments rises from 0 to 0.3 for the first child.

Figures 2.5 and 2.6 plot the estimates of the elasticity of non-cognitive skills with respect to different inputs for the first and second child, respectively. The figures demonstrate that the elasticity of non-cognitive skills with respect to monetary investments are decreasing in age while the elasticity with respect to shared time investments are increasing in age. Monetary investments are more productive early in development while shared time investments are more productive late in development. The figure also shows that the elasticity with respect to alone time investment is flat at about zero. Comparison of cognitive and non-cognitive elasticities with respect to alone and shared time investments demonstrates that alone time investment is more productive to cognitive skills while shared time investment is more productive to non-cognitive skills.

Additionally, when considering the preference parameters, the highest weight is on leisure at 0.83, which is not surprising as it includes any time that households do not spent with their children. Furthermore, households have a slight preference for their second child; the weight on the second child's skills is 0.06, while the weight on the first child's skills is 0.05. Even though the difference in preference weights on the two children is negligible, the weight on the second child is 20% higher than the weight on the first child. Note that the ratio of household investments in the two children is a function of the ratio of children's preference weights. Finally, with a weight on cognitive skills equal to 0.51, households equally value cognitive and non-cognitive skills of children. The details of the primitive parameters estimates, ω's, are available in Appendix B.1.

2.4.2 *Counterfactual Experiments*

I use the model estimates to quantify the effect of income transfers, one of the most popular counterfactual policy experiments in child development literature, on the birth order effects. Figure 2.7 plots the birth order effects in units of standard deviation at the end of development, seventeen years of age, against household income. [4] The birth order effects on cognitive and non-cognitive skills range from 0.61 and 0.46 of a standard deviation in households that earn 10,000 USD per year to 0.34 and 0.11 in households that earn 100,000 USD per year. Concavity of the technologies implies that each additional dollar increases the skills at a decreasing rate, and the concavity of the preferences gives the household an incentive to close the skills gap between the children as the household's income increases.

The slope of the curve is the effect of an annual income transfer of 10,000 USD on the birth order effects, which is decreasing in household income. The figure shows that an increase in annual income of 10,000 USD decreases the birth order effects in cognitive and non-cognitive skills, on average, by 0.03 and 0.04 of a standard deviation. The gain, however, is decreasing in household income. For example, an annual transfer of 10,000 USD to households that earn 10,000 USD per year decreases the birth order effects on cognitive and non-cognitive skills by 0.08 and 0.10 of a standard deviation. An identical transfer to households that earn 90,000 USD per year decrease the birth order effects in cognitive and non-cognitive skills only by 0.01 and 0.02 of a standard deviation. Therefore, the model documents another efficiency benefit for policies that target children in low-income households.

The estimates of households preference weights demonstrated that the weight on

[4] The initial skills for the following counterfactual experiments are estimated at age three and are assumed to be the same for the two children. One can estimate the initial conditions conditional on birth order. My approach is more conservative.

the second child's skills is 20% higher than the preference weight on the first child's skills. This implies that the birth order effects are not the result of favorable preferences for the first child, rather they must be the result of the differences in cognitive and non-cognitive elasticities, the life-cycle structure of the time constraints, or some combination of the two. In other words, if the household have had equal preferences for the two children, then the magnitude of the birth order effects would have been larger. To quantify the effect of unequal preference weights on the birth order effects, I compute the birth order effects in a counterfactual world where households have equal preference weights for the first and second child. For this experiment, I set the preference weights on the first and second child equal to the average of the two and re-run the model.

Figure 2.8 plots the birth order effects in units of standard deviation at the end of development against household income. The figure demonstrates that equal preference weight for the two children exacerbates the birth order effects. As the weight on the second child's skills decreases and the weight on the first child's skills increases, households decrease their investments in the second child and increase their investments in the first child. This widens the skills gap between the two children. For example, the birth order effects on cognitive and non-cognitive skills in households that earn 10,000 USD per year increases from 0.61 and 0.46 of a standard deviation, in the baseline, to 0.99 and 0.48 of a standard deviation, in the counterfactual, respectively. The effect of equal preference weights is more noticeable on cognitive skills than non-cognitive skills, which is due to the higher elasticity of cognitive skills with respect to investments.

2.4.3 Discussion

In this section, I compare the gradient of the birth order effects in household's income between the reduced-form and the structural models. Figure 2.9 plots the birth order effects in cognitive and non-cognitive skills by the level of household income. The black line shows the income gradient implied by the reduced-form model, while the gray line shows the gradient implied by the structural model. The figure demonstrates that the income gradient implied by the structural model is steeper than the gradient implied by the reduced-form model. This discrepancy comes from the limitations of each model. The reduced-form model allows for heterogeneity in multiple dimensions, but it has endogeneity problems. [5] In contrast, the structural model addresses the endogeneity of the household's investment choices, but it only allows for heterogeneity in household's income. This modeling choice is for tractability, but if households are heterogeneous in other dimensions that are correlated with income, then income captures their effects. For example, households with a high income may have high ability parents. If high ability parents become more productive after investing in their first-born child relative to low ability parents, then their investments in the second-born child become more productive, which causes the birth order effects to decrease in the parents ability. But if ability is not modeled, then income captures its effects.

2.5 Conclusions

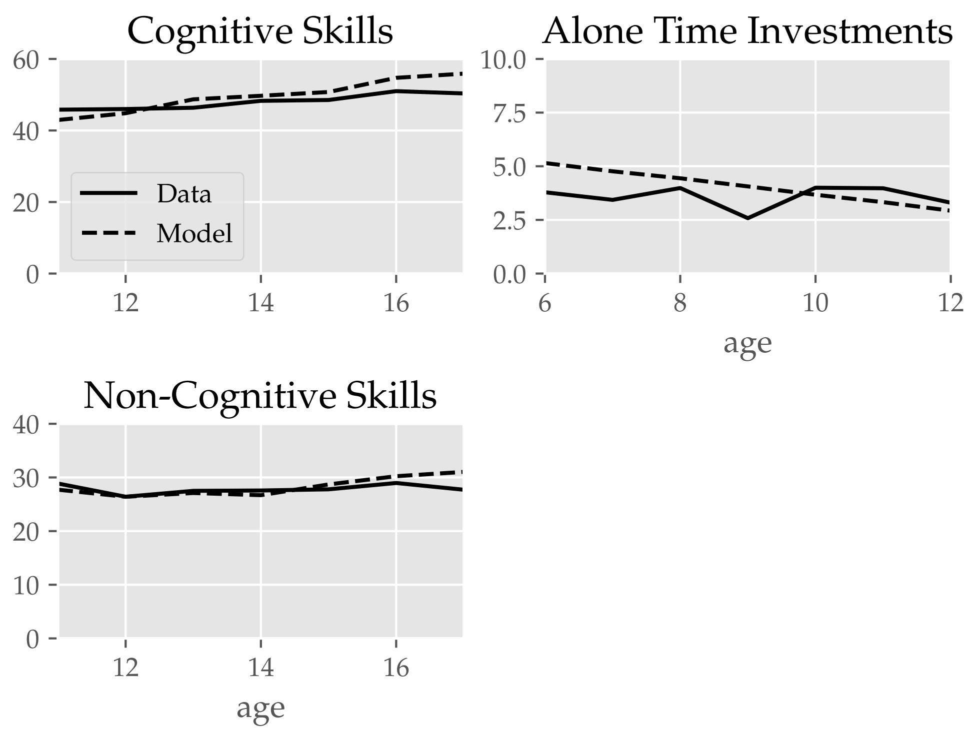
In this chapter, I demonstrate that income transfers can alleviate the birth order effects, but their effects are heterogeneous and decreasing in household income. The counterfactual experiment shows that an annual income transfer of 10,000 USD to

[5] See section 1.4.4

low-income households decreases the birth order effects on cognitive and non-cognitive skills by one-sixth, which is five times bigger than the effect in high-income household. Therefore, the model demonstrates another efficiency benefit for policies that target children in low-income households.

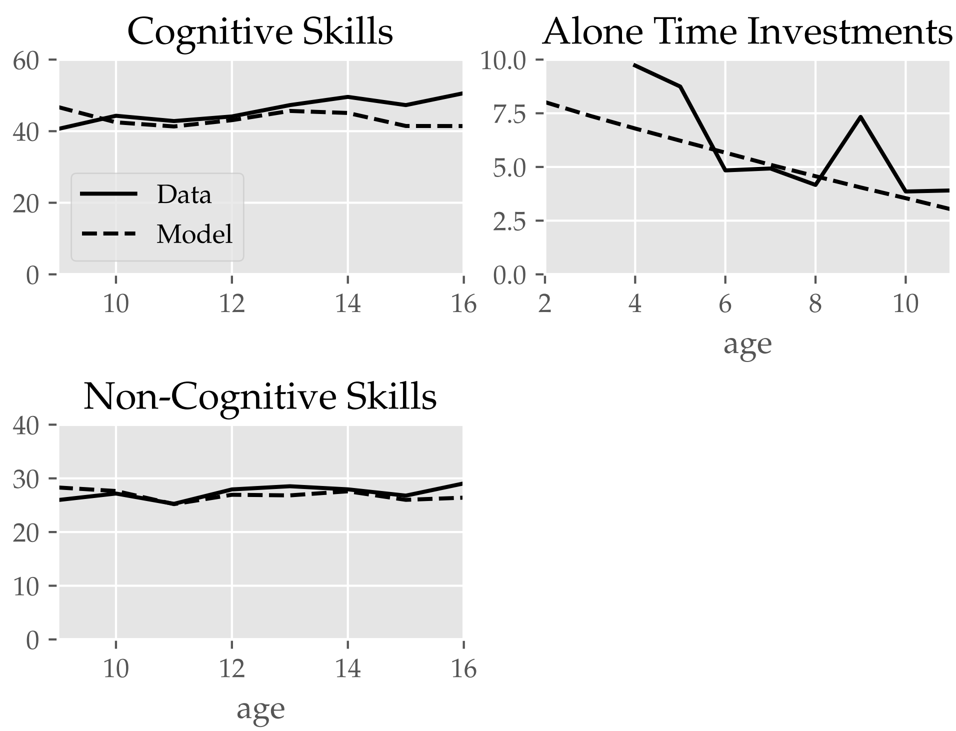
The model has a number of limitations. The model takes fertility as exogenous. There is developing evidence that birth spacing between siblings affects the birth order effects, which is an omitted variable in the model and can bias the findings. Moreover, the model does not address the labor supply decisions, which is a potential margin for households to save time. For example, households can choose to work more in some periods in order to work less in other periods and spend more time with their children. Counterfactuals that do not take labor supply decisions into account are likely to be a lower-bound for the policy effects. Finally, the model allows for heterogeneity only in household's income. If households are heterogeneous in other dimensions, then income captures their effects.

Figure 2.1: Model Fit: Mean Conditional on Age - First Child



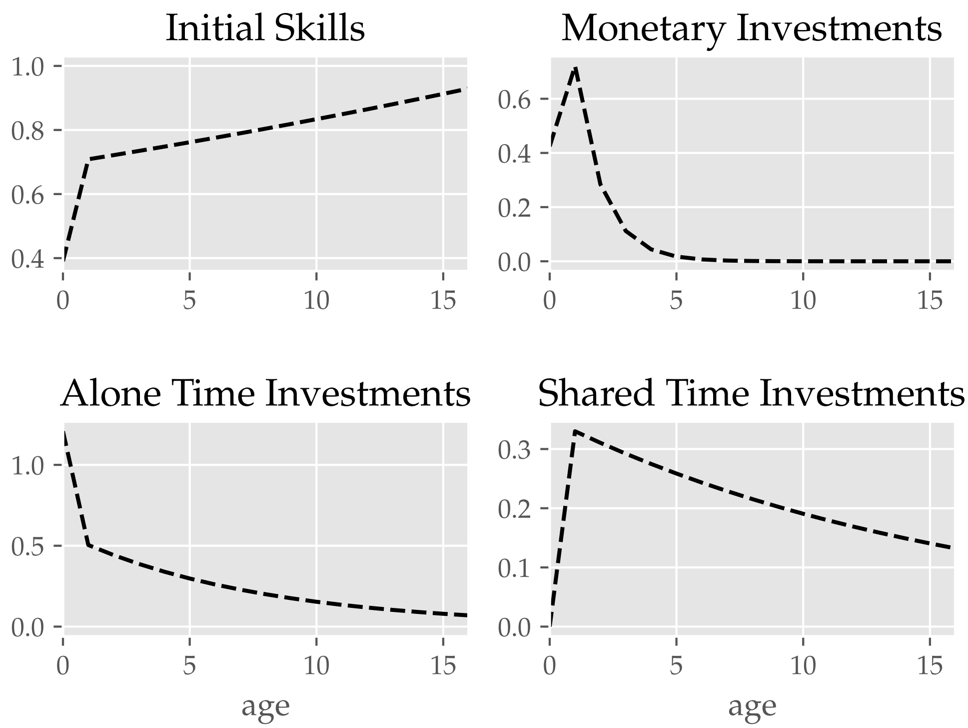
Source: Model estimates using a sample of PSID-CDS data.

Figure 2.2: Model Fit: Mean Conditional on Age - Second Child



Source: Model estimates using a sample of PSID-CDS data.

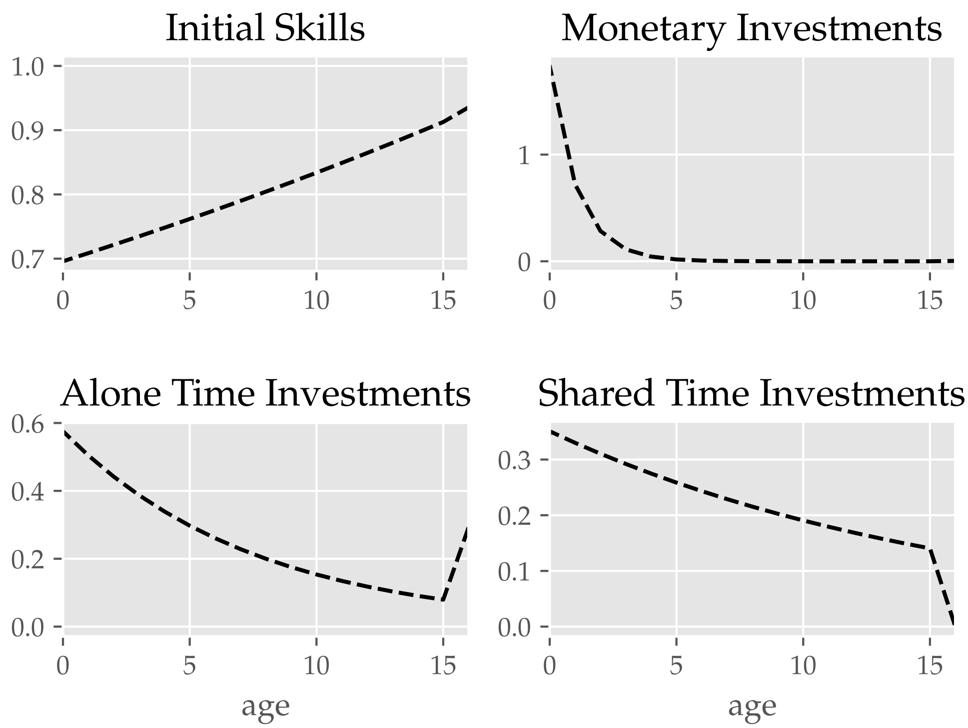
Figure 2.3: Estimates of Cognitive Elasticities - First Child



Source: Model estimates using a sample of PSID-CDS data.

Note: The figure pertains to households with birth spacing of one year.

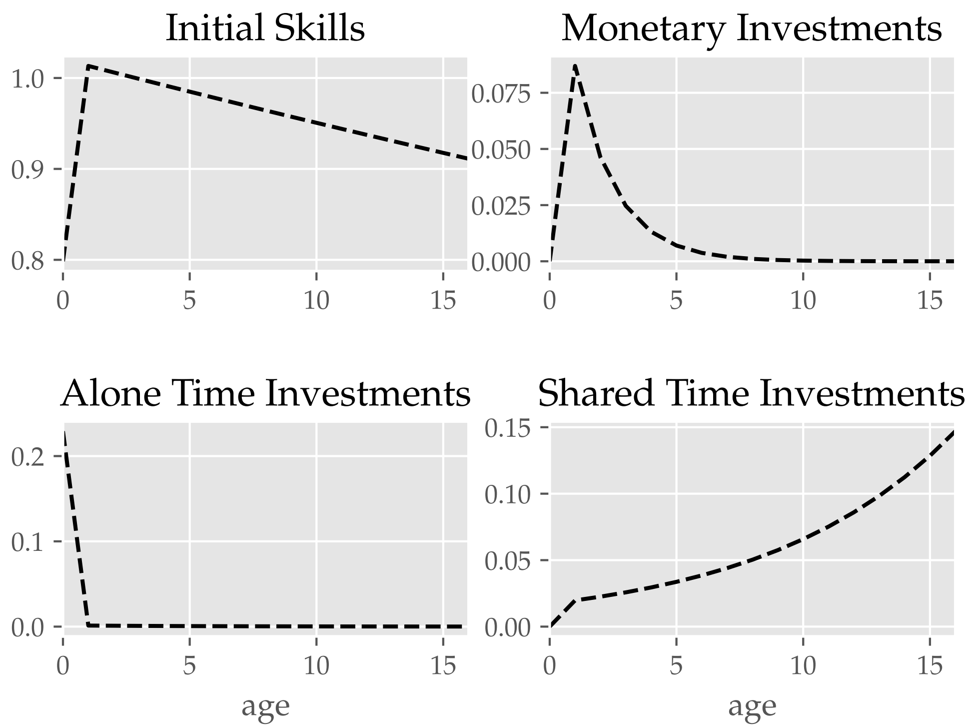
Figure 2.4: Estimates of Cognitive Elasticities - Second Child



Source: Model estimates using a sample of PSID-CDS data.

Note: The figure pertains to households with birth spacing of one year.

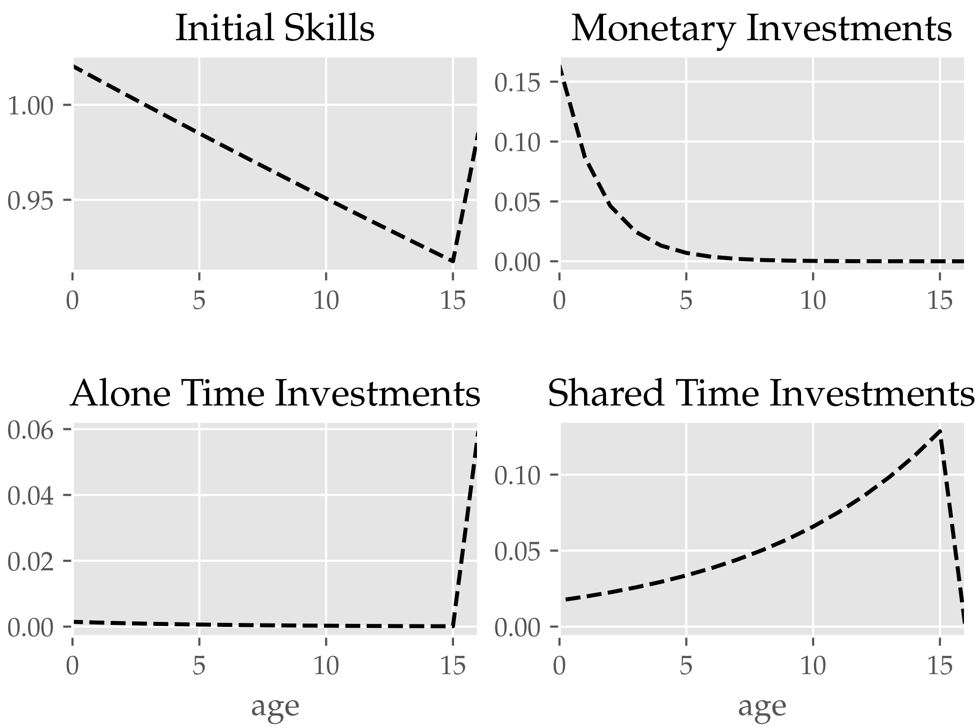
Figure 2.5: Estimates of Non-Cognitive Elasticities - First Child



Source: Model estimates using a sample of PSID-CDS data.

Note: The figure pertains to households with birth spacing of one year.

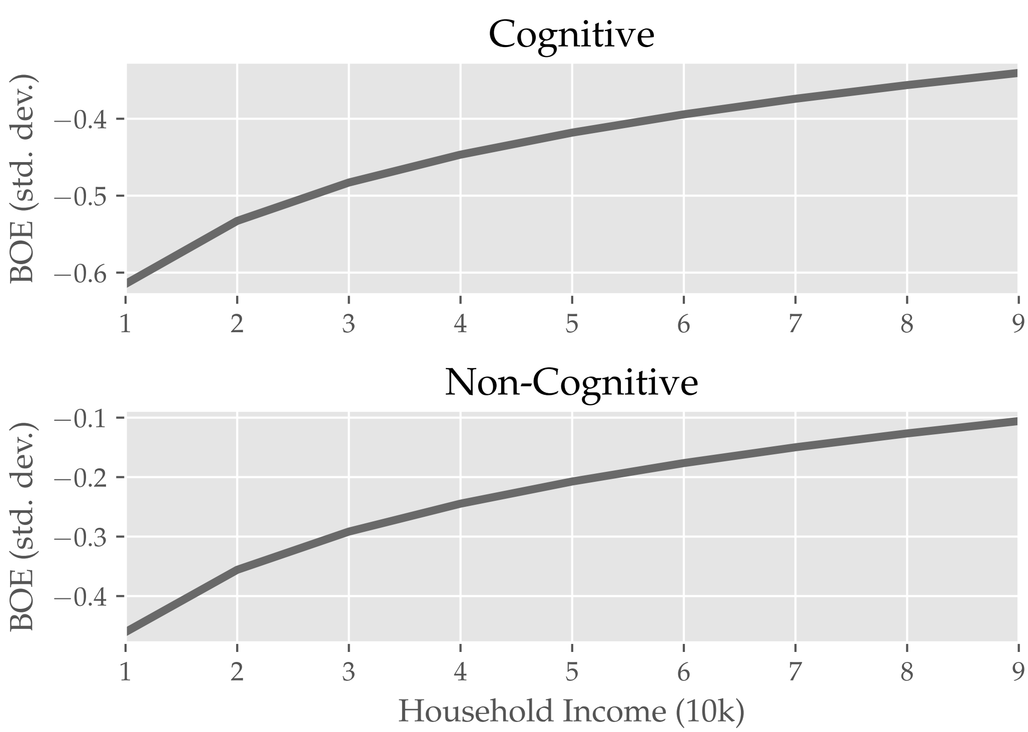
Figure 2.6: Estimates of Non-Cognitive Elasticities - Second Child



Source: Model estimates using a sample of PSID-CDS data.

Note: The figure pertains to households with birth spacing of one year.

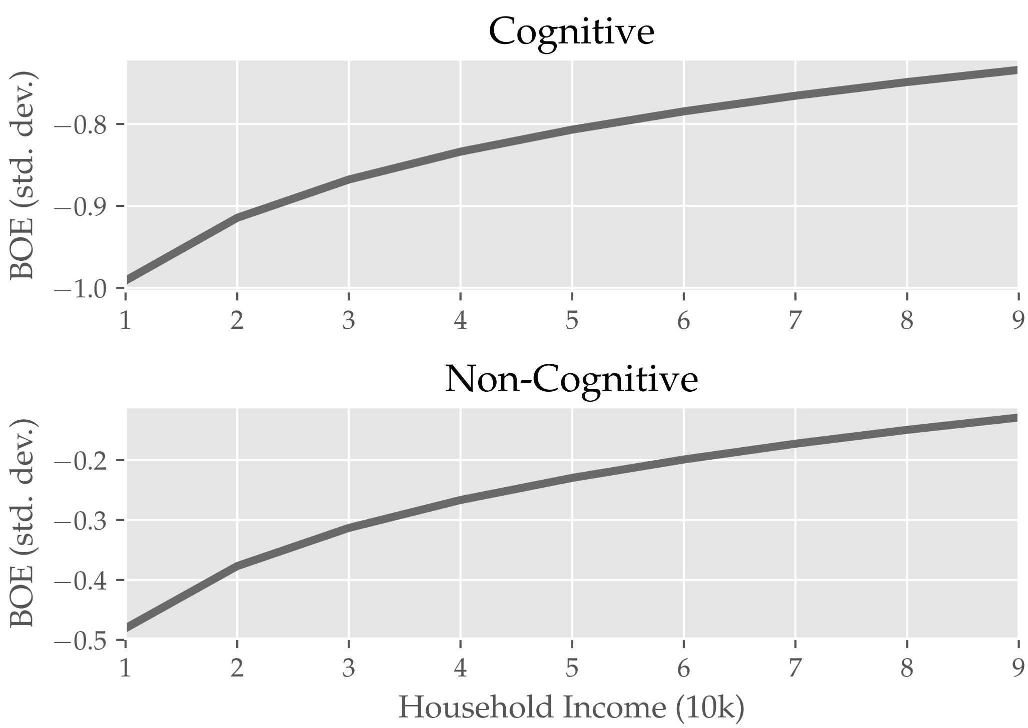
Figure 2.7: Birth Order Effects at Age 17



Source: Model estimates using a sample of PSID-CDS data.

Note: BOE denotes the birth order effect and is defined as the difference between the second and the first child skills.

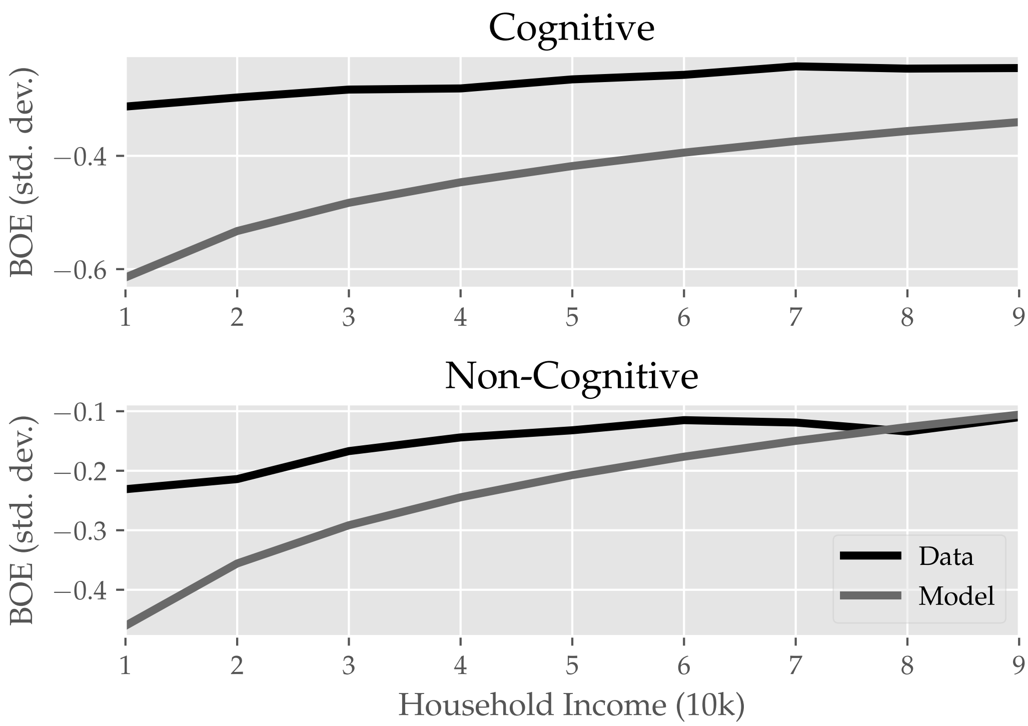
Figure 2.8: Birth Order Effects at Age 17 under Equal Preference Weights



Source: Model estimates using a sample of PSID-CDS data.

Note: BOE denotes the birth order effect and is defined as the difference between the second and the first child skills. The graph plots the BOE had the household had the same preference weights on the first and second child.

Figure 2.9: Birth Order Effects: Reduced-Form vs. Structural Models



Source: Model estimates using a sample of PSID-CDS data.

Note: BOE denotes the birth order effect and is defined as the difference between the second and the first child skills. "Data" and "Model" denote the prediction of the reduced-form and structural models, respectively.

Chapter 3

SCHOOLS, PARENTS, AND CHILD DEVELOPMENT [1]

3.1 Introduction

The wide dispersion of measured human capital in children and its strong relationship with later life outcomes has prompted a renewed interest in understanding the determinants of skill formation among children (for a recent review, see Heckman and Mosso, 2014). Throughout their lives, children are influenced by many factors–parents, neighborhoods, peers, and schools–and quantifying the relative importance of the factors at various points in the development process is a central question. This chapter develops a framework to examine the relative importance of investments at home and at school during an important transition for many children, entering formal schooling at kindergarten, which represents a substantial increase in interactions with caregivers other than family members.

We synthesize two separate and parallel research programs. First, a child development literature, which uses observational data on measures of children's cognitive and non-cognitive skills, assesses the importance of parental investments on the development process (Todd and Wolpin, 2007; Cunha and Heckman, 2007, 2008; Cunha *et al.*, 2010). This analysis is largely silent on the role of schools, and how the distribution of school quality across children affects inequality in children's skills. In contrast, a distinct second literature, which uses large scale administrative data from particular school systems, assesses the importance of schools, classrooms, and teachers (Rockoff, 2004; Rivkin *et al.*, 2005; Aaronson *et al.*, 2007; Jackson, 2012; Chetty *et al.*, 2014a,b;

[1] WITH FRANCESCO AGOSTINELLI AND MATTHEW WISWALL.

Flèche, 2017). This research is mostly silent on the importance of influences outside of school. We develop an empirical framework general enough to "nest" many of the key features of the prior literature. Our framework allows for both classroom and parental influences, imperfect measures of both skills and inputs, cognitive and non-cognitive skills, and complementarities between children's skills and investments from home and school.

Our first empirical challenge in estimating the technology of skill formation is measurement error. While measurement issues exist in many areas of empirical research, they may be particularly salient in research about child development. There exists a number of different measures of children's skills, and each measure can be arbitrarily located and scaled. [2] In the presence of these measurement issues, identification of the underlying latent process of skill development is particularly challenging, but nonetheless essential because ignoring the measurement issues through ad hoc simplifying assumptions could severely bias our inferences. We treat both the home and school as latent, fundamentally unobserved inputs. For the home influences, we use a number of measures, following the dynamic latent factor structure of Cunha *et al.* (2010). For the school influences, we exploit the clustering by classroom in our data, as in much of the recent education production function literature.

Our second empirical challenge is the classic omitted variable bias. Our approach is to use the rich nature of the data, including measures for cognitive skills, non-cognitive skills, and parental investments, and allowing for a general classroom effect. We test this assumption using the rich set of parental characteristics that we have available. This follows approaches such as Chetty *et al.* (2014a) that use information on parental income from tax records to test the validity of teacher value

[2] For a recent analysis of how measurement issues can be particularly salient, see Bond and Lang (see 2013b,a) who analyze the black-white test score gap.

added models. Although parental and household characteristics are strongly correlated with outcomes, once we estimate our models, the estimates for these variables are small. We further test the validity of our models using a number of out-of-sample tests, exploiting extant experimental variation. We "simulate" the effects of the Tennessee Student/Teacher Achievement Ratio (STAR) experiment (Krueger, 1999; Chetty *et al.*, 2011). The intuition for this approach is that if our estimated model can reproduce the experimental results, it would provide a strong suggestion of our model's validity.

Our identification analysis builds on previous work but offers a distinct approach to the empirical challenges. Previous approaches apply the techniques developed for cross-sectional latent factor models (Anderson and Rubin, 1956; Jöreskog and Goldberger, 1975; Goldberger, 1972; Chamberlain and Griliches, 1975; Chamberlain, 1977a,b; Carneiro *et al.*, 2003) to the dynamic latent factor models describing the development of children's skills. In an influential paper applying latent factor modeling to child development, Cunha *et al.* (2010) identify the skill production technology by first "re-normalizing" the latent skill distribution at each period, treating the skills in each period as separate latent factors. While latent skills, which lack a meaningful location and scale, require some normalization (say at the initial period), repeated re-normalization every period is an unnecessary over-identifying restriction if the production function already has a known location and scale. We show that nonparametric identification of this class of KLS production functions is possible without these re-normalization restrictions, and our identification approach avoids imposing restrictions these restrictions because they can bias the estimation (Agostinelli and Wiswall, 2016b).

We find that school investments are an important determinant of children's skills at the end of kindergarten, whereas home investments, although strongly correlated with end-of-kindergarten outcomes, have smaller effects. In addition, we document a negative complementarity between children's skills at kindergarten entry and investments from schools, implying that low-skill children benefit the most from an increase in the quality of schools. The counterfactual policy experiments show that providing all children with the 90th percentile of school investments decreases the standard deviation of the end-of-kindergarten distribution of skills by about 0.2 of a standard deviation, a substantial effect for one year of intervention. Moreover, it decreases the gap between the 10th and the 90th percentile of the end-of-kindergarten distribution of skills by 0.5 of a standard deviation.

Subsequent sections of this chapter are organized as follows. Section 3.2 discusses the econometric model, identification assumptions, and estimation strategy. Section 3.3 describes the data. Section 3.4 reports the estimation and the counterfactual policy experiments results. Section 3.5 provides the validation and robustness exercises. Section 3.6 concludes the chapter.

3.2 Model and Estimation Framework

3.2.1 Skill Development

At each age $t = 0, 1, \ldots, T$ children are characterized by a set of J skills. Let $\theta_{j,i,t}$ be child i's stock of skill j at age t. The collection of J skills for child i is represented by the vector $\Theta_{i,t} = \{\theta_{1,i,t}, \ldots, \theta_{J,i,t}\}$. Skills include both "cognitive" and "non-cognitive" skills, and can include other attributes of the child such as health or personality traits. Skill j in the next period is produced according to this technology:

$$\theta_{j,i,t+1} = f_{j,t}(\Theta_{i,t}, S_{j,i,t}, H_{j,i,t}) \tag{3.1}$$

where $H_{j,i,t}$ is a vector of investments from "home" and $S_{j,i,t}$ is a vector of investments from the "school" the child attends. This formulation of the skill technology specifies skills as a first order Markov process. The child development function (3.1) is indexed by j and t to emphasize that the technology itself is heterogeneous across skills and over ages.

We use the terms “home” and “school” broadly. Investment from home represents all child development activities outside of school, and need not be from interactions with parents, but could involve non-parental caregivers such as after school care. Investment from school can be from any interaction during the school day, including from teachers, other schools staff, and peers. In our empirical specifications we allow for classroom specific investment, which can expose children within the same school to different investments. Another key feature of our model is that investments have both a “quality” and a “quantity” dimension. All else being equal, investment from school increases with the time children spend in school, and investment from home correspondingly decreases. Using child level data on time spent at school, we attempt to ascertain how much of the heterogeneity in school investment is due to quantity of school time relative to quality for a fixed amount of time.

In the child development literature (see perhaps Heckman and Masterov, 2007), (3.1) is typically labeled a “child development technology”. In the education literature (see for example Rivkin *et al.*, 2005; Krueger, 1999), (3.1) is labeled an “education production function”. In the former case, the skills include cognitive and non-cognitive skills measured in survey data, and the investments from parents are the focus of the analysis. In the latter case, skills are typically reading or mathematics skills measured using standardized tests administered in schools, and the productivity of school inputs is the focus. Our specification nests both of these frameworks, and an important emphasis of our approach is to examine the relative importance of home and school factors on child development.

3.2.2 Measurement

Our skill development/education production function is written in terms of latent variables. An important empirical advance is the recognition that children's skills and the various investments in those skills from parents and schools are, in general, unobserved and imperfectly measured. We consider a measurement system which can incorporate several major approaches. We allow for multiple measures of latent variables, and following Cunha and Heckman (2008), we conceptualize each measure M as imperfect, with some measurement error ϵ.

Let ω be a generic latent variable. For each latent variable for a child i, say latent skill stock in period t with $\omega_{i,t} = \theta_{j,i,t}$, we have $m = 1, 2, \ldots, K_{\omega,t}$ measures. The number of measures can vary across latent variables and periods, and depends on the data available. Each scalar measure is denoted $M_{\omega,i,t,m}$ and takes the form:

$$M_{\omega,i,t,m} = b_{\omega,t,m}(\omega_{i,t}, \epsilon_{\omega,i,t,m}) \tag{3.2}$$

where $\epsilon_{\omega,i,t,m}$ is the measurement error for measure m and $b_{\omega,t,m}$ is the measurement function.

One could take a similar approach with school inputs and assume observed classroom characteristics, such as class size or teacher's years of experience, are imperfect measures of classroom quality (Bernal *et al.*, 2016). We take a different and more robust approach, and exploit the clustering by classrooms in our data to identify the distribution of classroom quality and the productivity of these inputs in producing child skills. We detail this approach below.

3.2.3 Baseline Empirical Specification

The general model presented above provides some of the general concepts of our empirical specification. Next we present specific functional forms, identification assumptions, and estimation strategy, which we can take directly to data. We start with a baseline specification, based on a particular specification of the production technology (3.1) and measurement system (3.2). This specification is the most restrictive specification that we consider, and we generalize it in subsequent sections.

The baseline specification assumes a log-linear, Cobb-Douglas, form for the production technology. We specify the skill development function for each skill j (3.1) as

$$\ln \theta_{j,i,t+1} = \ln A_{j,t} + \gamma_{1,j} \ln \theta_{j,i,t} + \gamma_{2,j} \ln S_{j,i,t} + \eta_{j,i,t} \tag{3.3}$$

Skill j in period $t+1$ is produced by the previous period stock $\theta_{j,i,t}$ and school investments $S_{j,i,t}$. The parameter $\gamma_{1,j}$ provides the relative productivity of the existing stock of skills. $\gamma_{2,j}$ provides the productivity of school investments. $\eta_{j,i,t}$ is a skill production "shock," representing all omitted inputs. $\ln A_{j,t}$ is an intercept, representing total factor productivity (TFP).

In the baseline framework, we assume a linear system of measures for the latent skills stocks. For each latent variable ω and period t, we have $m = 1, 2, \ldots, K_{\omega,t}$ measures given by

$$M_{\omega,i,t,m} = \mu_{\omega,t,m} + \lambda_{\omega,t,m} \ln \omega_{i,t} + \epsilon_{\omega,i,t,m} \tag{3.4}$$

where $M_{\omega,i,t,m}$ are the measures, $\epsilon_{\omega,i,t,m}$ are the measurement "errors", and $\mu_{\omega,t,m}$ and $\lambda_{\omega,t,m}$ are the measurement parameters. $\mu_{\omega,t,m}$ and $\lambda_{\omega,t,m}$ provide the location and scale of the measure m. $\lambda_{\omega,t,m}$ is often referred to as a "factor loading" for the latent "factor" $\omega_{i,t}$ and measure m. Given the inclusion of the intercept, we normalize the measurement error to be mean-zero without loss of generality: $E(\epsilon_{\omega,t,m}) = 0$ for all ω, t, m.

For children's skills ($\omega_{i,t} = \theta_{j,i,t}$), the measures M are a combination of available assessments, as discussed in more detail below. Instead of using specific measures of classroom quality, we treat classroom quality as a kind of "common" input experienced by all children in the classroom. This specification is more robust than assuming particularly classroom characteristics, such as class size or teacher's year of experience, provide particular measures of the within school investments.

Finally, note that our baseline framework nests the standard "value added" education production specifications. To see this, consider the special case of this model where a given measure indexed m is assumed to measure the latent skills with no error. This measure is typically an end-of-grade standardized test in a subject such as mathematics or reading. In our notation, the value added specification takes the form:

$$M_{j,i,s,t+1,m} = \gamma_{1,j} M_{j,i,s,t,m} + \delta_{j,s,t} + \eta_{j,i,s,t,m} \tag{3.5}$$

where $\delta_{j,s,t}$ is the common school "effect" for school s. Equation (3.5) shows that the value added specification is a special case of our baseline specification. First, the model in (3.5) can be easily mapped to our baseline model (3.3) as a special case when the school investment is homogeneous among children in the same school, that is $S_{j,i,t} = S_{j,s,t}$ for all i in school s. In fact, in this case it is easy to show that $\delta_{j,s,t} = \gamma_{2,j} \ln S_{j,s,t}$. Moreover, we allow for skills to be unobserved and we do not base our empirical analysis on arbitrarily scaled test scores.

3.2.4 Identification of Baseline Specification

In this section, we describe the identification of our baseline specification. The concepts introduced here apply to the identification of the more general models that we explore next.

The baseline specification is identified up to some initial normalization, and it requires several normalizations given that latent skills and school inputs are not directly observed and have no particular location or scale. We normalize all of the initial period ($t = 0$) latent variables to have mean of zero and variance of one:

Normalization 1 *Initial period ($t = 0$) normalizations:*

(i) $E(\ln \omega_{i,0}) = 0$

(ii) $V(\ln \omega_{i,0}) = 1$

for all $\omega_{i,0} \in \{\Theta_{i,0}, S_{j,i,0}\}$

With these normalization, we treat all latent variables symmetrically, imposing the same normalizations on each in order to ease the interpretation of the estimates. One normalization that appears non-standard relative to the prior literature is that for the school input, which we can write as $\delta_{j,s,0} \equiv \gamma_{2,j} \ln S_{j,s,0}$, the scale normalization implies that $V(\delta_{j,s,0}) = \gamma_{2,j}$. That is, the variance of the school component (typically treated as a school or classroom level fixed effect) is equal to the parameter $\gamma_{2,j}$. We choose this normalization so that the technology parameters in (3.3), which represent the productivity of each factor in producing a child's skills, can easily be compared across the latent variables.

Given the normalizations on the location of the latent variables, the measurement intercepts are identified from the mean of the measures:

$$\mu_{\omega,0,m} = E(M_{\omega,i,0,m}) \text{ for all } \omega_{i,0} \in \{\Theta_{i,0}\}. \tag{3.6}$$

We cannot identify the scaling parameters for the initial period, $\lambda_{\omega,0,m}$, without further restrictions on the measurement errors. We consider the following independence assumptions, commonly used in this literature (e.g. Cunha *et al.*, 2010), that the measurement errors are independent of each other and of the latent variables:

Assumption 1 *Measurement model assumptions:*

(i) $\epsilon_{\omega,i,t,m} \perp \epsilon_{\omega,i,t,m'}$ *for all* t, $m \neq m'$, *and latent variable* ω

(ii) $\epsilon_{\omega,i,t,m} \perp \epsilon_{\omega,i,t',m}$ *for all* $t \neq t'$, *all* m *and* m', *and latent variable* ω

(iii) $\epsilon_{\omega,i,t,m} \perp \omega'$ *for all* t, m, *and latent variable* ω'

Assumption 1 (i) asserts that measurement errors are independent contemporaneously across measures. Assumption 1 (ii) asserts that measurement errors are independent over time. Assumption 1 (iii) asserts that measurement errors are independent of the latent variables for child skills and investments. Although these assumptions are strong in some sense, they are common in the current literature. [3]

Under these assumptions, the initial period (t=0) scaling factors are identified from ratio of covariances between the measures:

$$\lambda_{\omega,0,m} = \sqrt{\frac{Cov(M_{\omega,0,m}, M_{\omega,0,m'})Cov(M_{\omega,0,m}, M_{\omega,0,m''})}{Cov(M_{\omega,0,m'}, M_{\omega,0,m''})}} \tag{3.7}$$

for any $m \neq m' \neq m''$. [4] Given the identification of the measurement parameters for the initial period, we can identify the latent variables up to the measurement errors:

$$\ln \omega_{i,0} = \frac{M_{\omega,i,0,m} - \mu_{\omega,0,m}}{\lambda_{\omega,0,m}} + \frac{\epsilon_{\omega,i,0,m}}{\lambda_{\omega,i,0,m}} = \widetilde{M}_{\omega,i,0,m} + \frac{\epsilon_{\omega,i,0,m}}{\lambda_{\omega,0,m}} \tag{3.8}$$

where $\widetilde{M}_{\omega,i,0,m}$ is the transformed measure using the identified measurement parameters. To analyze the identification of the skill development function (3.3), we substitute the transformed measure (3.8) into (3.3):

$$\ln \theta_{j,i,1} = \ln A_{j,0} + \gamma_{1,j}(\widetilde{M}_{\omega,i,0,m} - \frac{\epsilon_{j,i,0,m}}{\lambda_{j,0,m}}) + \gamma_{2,j} \ln S_{j,i,0} + \eta_{j,i,0} \tag{3.9}$$

We use a measure of the skills at $t = 1$ given by

$$\ln \theta_{j,i,1} = \frac{M_{j,i,1,m} - \mu_{j,1,m}}{\lambda_{j,1,m}} - \frac{\epsilon_{j,i,1,m}}{\lambda_{j,1,m}} \tag{3.10}$$

[3]Our assumption of full independence is sufficient, but not necessary, for at least some of our identification analysis.

[4]In the presence of at least two latent skills, the identification of the measurement systems is guaranteed with only two measures for each latent skill.

substituting (3.10) into (3.9)

$$M_{j,i,1,m} = \mu_{j,1,m} + \lambda_{j,1,m} \ln A_{j,0} + \lambda_{j,1,m}\gamma_{1,j}\widetilde{M}_{j,i,0,m} + \lambda_{j,1,m}\gamma_{2,j} \ln S_{j,i,0} + \kappa_{j,i,0,m} \tag{3.11}$$

where $\kappa_{j,i,0,m}$ is given by

$$\kappa_{j,i,0,m} = \lambda_{j,1,m}\eta_{j,i,0} - \gamma_{1,j}\frac{\lambda_{j,1,m}}{\lambda_{j,0,m}}\epsilon_{j,i,0,m} + \epsilon_{j,i,1,m} \tag{3.12}$$

The equation in (3.11) can be rewritten in terms of the "reduced-form" parameters as

$$M_{j,i,1,m} = \beta_{0,j} + \beta_{1,j}\widetilde{M}_{j,i,0,m} + \beta_{2,j} \ln S_{j,i,0} + \kappa_{j,i,0,m} \tag{3.13}$$

In particular, it is easy to show that the reduced-form parameters map into the structural parameters as

$$\beta_{0,j} = \mu_{j,1,m} + \lambda_{j,1,m} \ln A_{j,0} \tag{3.14}$$

$$\beta_{1,j} = \lambda_{j,1,m}\gamma_{1,j} \tag{3.15}$$

$$\beta_{2,j} = \lambda_{j,1,m}\gamma_{2,j} \tag{3.16}$$

The system in (3.14)-(3.16) includes 3 equations and 5 unknowns. Agostinelli and Wiswall (2016a) show that this under-identification problem can be solved in the presence of age-invariant measures for skills.

Definition 1 *A pair of measures $M_{t,m}$ and $M_{t+1,m}$ is* age-invariant *if $E(M_{t,m}|\theta_t = p) = E(M_{t+1,m}|\theta_{t+1} = p)$ for all $p \in \mathbb{R}_{++}$.*

Intuitively, age-invariance implies that the expectation of the measure M, e.g. a test score, for two children with different ages but equal skills is the same, which implies that the measurement parameters are constant with respect to a child's age ($\mu_{j,1,m}$

$= \mu_{j,0,m}$ and $\lambda_{j,1,m} = \lambda_{j,0,m}$). We are now able to identify the structural parameters from the reduced-form parameters as

$$\ln A_{j,0} = \frac{\beta_{0,j} - \mu_{j,0,m}}{\lambda_{j,0,m}} \tag{3.17}$$

$$\gamma_{\ell,j} = \frac{\beta_{\ell,j}}{\lambda_{j,0,m}} \quad \forall\, \ell \in \{1,2\} \tag{3.18}$$

where the measurement parameters for the initial period ($\mu_{j,0,m}$, $\lambda_{j,0,m}$) are already identified as shown in (3.6) and (3.7). Equations (3.17) and (3.18) show that in order to consistently identify the structural parameters, we need to consistently identify the reduced-form parameters βs.

Assumption 2 *Mean-independence of the production function shock:*

$$E(\eta_{j,0}|\theta_{j,0}, S_{j,0}) = 0$$

Under Assumption 2, Agostinelli and Wiswall (2016a) show that the reduced-form parameters βs are consistently identified using the multiple excluded measures as instrumental variables to adjust for measurement error.

3.2.5 Generalizing: Parental Investments

We generalize our baseline specification in (3.3) by including parental investments. The technology of skill formation for each type of skill j is

$$\ln \theta_{j,i,t+1} = \ln A_{j,t} + \gamma_{1,j} \ln \theta_{j,i,t} + \gamma_{2,j} \ln S_{j,i,t} + \gamma_{3,j} \ln H_{i,t} + \eta_{j,i,t} \tag{3.19}$$

where $H_{i,t}$ is the parental investments, and the parameter $\gamma_{3,j}$ is the productivity of parental investments. Note that we model parental investments as a common input among all skills, but we allow it to a have different productivity in producing each of the skills. Similar to our treatment of the skills, we assume that parental investments is unobserved and imperfectly measured. Further, we assume a linear system of measures for latent parental investments as in (3.4). We have $m = 1, 2, \ldots, K_{H,t}$ measures given by

$$M_{H,i,t,m} = \mu_{H,t,m} + \lambda_{H,t,m} \ln H_{i,t} + \epsilon_{H,i,t,m} \tag{3.20}$$

where we maintain the assumptions on the measurement errors as in (1). Additionally, we assume the structural shock is mean-independent of the latent factors.

Assumption 3 *Mean-independence of the production function shock:*

$$E(\eta_{j,0}|\theta_{j,0}, H_0, S_{j,0}) = 0$$

Following the identification discussion in section 3.2.4, we can write the technology in (3.19) in terms of the measures as

$$\begin{aligned} M_{j,i,1,m} = \mu_{j,1,m} + &\lambda_{j,1,m} \ln A_{j,0} + \lambda_{j,1,m}\gamma_{1,j}\widetilde{M}_{j,i,0,m} + \\ &\lambda_{j,1,m}\gamma_{2,j} \ln S_{j,i,0} + \lambda_{j,1,m}\gamma_{3,j}\widetilde{M}_{H,i,0,m} + \kappa_{j,i,0,m} \end{aligned} \tag{3.21}$$

and it can be written in the reduced-form as

$$M_{j,i,1,m} = \beta_{0,j} + \beta_{1,j}\widetilde{M}_{j,i,0,m} + \beta_{2,j} \ln S_{j,i,0} + \beta_{3,j}\widetilde{M}_{H,i,0,m} + \kappa_{j,i,0,m} \tag{3.22}$$

3.2.6 Generalizing: Complementarities

In this section, we generalize the technology in (3.19) to allow for complementarities between investments and skills. In other words, we allow for heterogeneity in the productivity of home and school investments with respect to initial stock of skills. Following Agostinelli and Wiswall (2016a), we consider a trans-log technology of skill formation with interaction terms between investments and skills. The technology for each type of skill j is

$$\ln \theta_{j,i,t+1} = \ln A_{j,t} + \gamma_{1,j} \ln \theta_{j,i,t} + \gamma_{2,j} \ln S_{j,i,t} + \gamma_{3,j} \ln H_{i,t} + \\ \gamma_{4,j} \ln \theta_{j,i,t} \ln S_{j,i,t} + \gamma_{5,j} \ln \theta_{j,i,t} \ln H_{i,t} + \gamma_{6,j} \ln H_{i,t} \ln S_{j,i,t} + \eta_{j,i,t} \tag{3.23}$$

where the parameter $\gamma_{4,j}$ governs the complementarity between school investments and skills, $\gamma_{5,j}$ governs the complementarity between home investments and skills, and $\gamma_{6,j}$ governs the complementarity between home investments and school investments. Note that the elasticity of each input depends on the level of other inputs. For example, the elasticity of skill j with respect to school investments is

$$\frac{\partial \ln \theta_{j,i,t+1}}{\partial \ln S_{j,i,t}} = \gamma_{2,j} + \gamma_{4,j} \ln \theta_{j,i,t} + \gamma_{6,j} \ln H_{i,t} \tag{3.24}$$

$\gamma_{4,j} > 0$ implies a higher return to school investment for children with high initial skills relative to children with low initial skills, emphasizes the importance of investments prior to time t (dynamic complementarity), and supports policies that target resources to children with high initial skills.. In contrast, $\gamma_{4,j} < 0$ implies a higher return to school investment for children with low initial skill relative to children with high initial skill and supports policies that target resources to children with low initial skills.

Following the identification discussion in section 3.2.4, we can write the technology in (3.23) in terms of the measures as

$$M_{j,i,1,m} = \mu_{j,1,m} + \lambda_{j,1,m} \ln A_{j,0} + \lambda_{j,1,m}\gamma_{1,j}\widetilde{M}_{j,i,0,m} + \lambda_{j,1,m}\gamma_{2,j} \ln S_{j,i,0} + \\ \lambda_{j,1,m}\gamma_{3,j}\widetilde{M}_{H,i,0,m} + \lambda_{j,1,m}\gamma_{4,j}\widetilde{M}_{j,i,0,m} \ln S_{j,i,0} + \\ \lambda_{j,1,m}\gamma_{5,j}\widetilde{M}_{j,i,0,m}\widetilde{M}_{H,i,0,m} + \lambda_{j,1,m}\gamma_{6,j}\widetilde{M}_{H,i,0,m} \ln S_{j,i,0} + \kappa_{j,i,0,m} \tag{3.25}$$

and it can be written in the reduced-form as

$$M_{j,i,1,m} = \beta_{0,j} + \beta_{1,j}\widetilde{M}_{j,i,0,m} + \beta_{2,j} \ln S_{j,i,0} + \beta_{3,j}\widetilde{M}_{H,i,0,m} + \\ \beta_{4,j}\widetilde{M}_{j,i,0,m} \ln S_{j,i,0} + \beta_{5,j}\widetilde{M}_{j,i,0,m}\widetilde{M}_{H,i,0,m} + \beta_{6,j}\widetilde{M}_{H,i,0,m} \ln S_{j,i,0} + \kappa_{j,i,0,m} \tag{3.26}$$

3.2.7 Generalizing: Other Skills

We complete our model by including all the J skills in the technology of skill formation for each type of skill j. Further, we allow for complementarities between investments and all the J skills. The technology for skill j is

$$\begin{aligned}\ln\theta_{j,i,t+1} &= \ln A_{j,t} + \sum_{k=1}^{J} \gamma_{1,j,k} \ln \theta_{k,i,t} + \gamma_{2,j} \ln S_{j,i,t} + \gamma_{3,j} \ln H_{i,t} + \\ &\sum_{k=1}^{J} \gamma_{4,j,k} \ln \theta_{k,i,t} \ln S_{j,i,t} + \sum_{k=1}^{J} \gamma_{5,j,k} \ln \theta_{k,i,t} \ln H_{i,t} + \gamma_{6,j} \ln H_{i,t} \ln S_{j,i,t} + \eta_{j,i,t}\end{aligned} \tag{3.27}$$

The empirical model for the technology in (3.27) is constructed by substituting the transformed measures $\widetilde{M}$ for each of the corresponding latent factors.

3.2.8 School/Classroom Decomposition

We exploit the clustering by classroom in our data to identify the distribution of school investments. We treat school investments as a common input experienced by all children in the same classroom. For schools in which we observe multiple classrooms, the school investment factor can be decomposed into school- and classroom-specific effects:

$$\ln S_{j,i,0} = \ln \delta_{j,s,0} + \ln \delta_{j,c,0} \tag{3.28}$$

where $\delta_{j,s,0}$ and $\delta_{j,c,0}$ represent the school and classroom effects, respectively. These two effects are separately identified up to an additional normalization: classroom effects are mean-zero within each school, $E\left[\delta_{j,c,0}|s\right] = 0$. In the case of schools where we observe only one classroom, the two effects are not separately identified, but we are still able to identify the combined effect, $S_{j,i,0}$.

3.2.9 Estimation

In this section, we provide an algorithm for estimating the technologies. We demonstrate our algorithm using the technology with parental investments and complementarities in (3.26). The algorithm for estimating other technologies includes the same steps. We follow Arcidiacono *et al.* (2012) which develops an iterative algorithm for estimating models that include interactions with fixed effects. Within this framework, we allow for skills and parental investments to be unobserved and imperfectly measured. We adjust for measurement errors by using multiple excluded measures as instrumental variables as in Agostinelli and Wiswall (2016a). Supplementary Monte Carlo experiments are available in Appendix C.1.

- **Step 1:** Start with an initial guess for the parameters, $\hat{\beta}_j^0 = [\hat{\beta}_{0,j}^0, \hat{\beta}_{1,j}^0, \cdots, \hat{\beta}_{6,j}^0]$.
- **Step 2:** Conditional on $\hat{\beta}_j^h$, $h \in \{0, 1, ...\}$, estimate the school/classroom fixed effect for each school/classroom s:

$$\widehat{\ln S}_{j,i,0} = \frac{\sum_{i \in s} \left[M_{j,i,1,m} - \beta_{0,j}^h - \beta_{1,j}^h \widetilde{M}_{j,i,0,m} - \beta_{3,j}^h \widetilde{M}_{H,i,0,m} - \beta_{5,j}^h \widetilde{M}_{j,i,0,m} \widetilde{M}_{H,i,0,m}\right]}{\sum_{i \in s} \left[\beta_{2,j}^h + \beta_{4,j}^h \widetilde{M}_{j,i,0,m} + \beta_{6,j}^h \widetilde{M}_{H,i,0,m}\right]}$$

- **Step 3:** Conditional on $\widehat{\ln S}_{j,i,0}$, estimate $\hat{\beta}_j^{h+1}$ using the equation in (3.26) and Two-Stage Least Squares estimator.
- **Step 4:** If $||\frac{\hat{\beta}_j^h - \hat{\beta}_j^h}{\hat{\beta}_j^{h-1}}||_\infty < 0.01$, stop. Otherwise, return to Step 2.

3.3 Data

The data is from the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS). The ECLS covers a nationally representative cohort of children from kindergarten through eighth grade. Importantly for our analysis, ECLS collects information from a number of students at each school, covering multiple classrooms and including multiple children in each classroom. The ECLS links each student to their classroom, their sampled classmates within that classroom, and the classroom teacher. Moreover, the ECLS has extensive information on home investments and child outcomes.

3.3.1 Sample

We focus on the Fall and Spring of the kindergarten year. We include all classrooms that have at least five children with two mathematics, two reading, and two non-cognitive assessments at the beginning and at the end of the kindergarten year. Further, we require these children to have two measures of home investment during the kindergarten year. Applying these criteria result in 9,474 children within 1,211 classrooms from 668 schools. Our sample of children with complete information is more socio-economically advantaged relative to the full ECLS sample. This is indicated by the higher household income in our sample, which is about $68,700 compared to $62,300 in the full sample. Appendix C.2 provides the detail of the sample selection.

Table 3.1 presents detailed descriptive statistics for our sample at the time of entry into kindergarten. The average age of children at entry is 5.68 years. Sixty-seven percent of the children are white non-Hispanic, 14 percent are black non-Hispanic, 10 percent are Hispanic, and 9 percent are of an other race/ethnicity. Sixty-eight percent of the children are living with both biological parents. Their mothers are on average 34 years old, have an average of 13.8 years of schooling, and work an average of 26.1 hours per week. Examining the kindergarten classrooms and schools, the average class size is about 20 students. Some of the kindergarten classes are half-day, and others full-day. This implies that the average instructional time across all classrooms is 24 hours per week, lower than for later grades. Because these are kindergarten classrooms the teachers are overwhelmingly female. The teachers have an average of 9.6 years of experience in teaching kindergarten. Moreover, 36 percent of them have at least a master's degree. Finally, about 30 percent of the schools in the sample are non-public schools, including secular and religious private schools. We return to how these characteristics relate to classroom and school quality in later sections.

3.3.2 Child Development during Kindergarten

We briefly motivate our analysis by analyzing patterns in child development during the kindergarten period. Figure 3.1 displays the average scores on mathematics, reading, and non-cognitive assessments at the beginning and at the end of the kindergarten year by the level of child's household income. Each score is normalized using its respective mean and standard deviation at kindergarten entry to facilitate interpretation. Figure 3.1 reveals the wide dispersion of mathematics, reading, and non-cognitive scores at kindergarten entry across income deciles. At the time of entry, the gap between the average mathematics, reading, and non-cognitive scores among children at the lowest and the highest income deciles is 1.1, 1.0, and 0.5 of a standard deviation, respectively. Figure 3.2 plots the change in average scores between the beginning and the end of the kindergarten year. The figure indicates that by the end of kindergarten, the dispersion of average mathematics and non-cognitive scores widens by 0.1 of a standard deviation, while the dispersion of reading scores remains unchanged. Specifically, the gap in average mathematics, reading, and non-cognitive scores at the end of the kindergarten are 1.2, 1.0, and 0.6 of a standard deviation, respectively.

3.4 Estimates

3.4.1 Measurement Parameters

Table 3.3 presents the estimates of the measurement parameters for mathematics, reading, and non-cognitive skills and home investments. The location of a measure is the expected value of the measure at the average value of the latent. The scale of a measure is the effect of one standard deviation increase in the latent on the measure. The signal-to-noise ratio of a measure is the fraction of the variance of the measure that is explained by the latent. Note that a higher signal-to-noise ratio indicates that the measure is more informative about the latent. The table shows that the signal-to-noise ratios are high, particularly for mathematics and reading skills.

3.4.2 Production Technology

Table 3.4 presents the estimates of the baseline technology. The table demonstrates that all three skills are important in the production of each of the skills. In addition, the table shows that self-elasticities are considerably higher than cross-elasticities. For example, the elasticity of end-of-kindergarten reading skills with respect to initial reading skills is 0.543, while elasticities with respect to initial mathematics and initial non-cognitive skills are 0.225 and 0.073, respectively. Furthermore, the table demonstrates that school investments have a sizeable effect on each of the skills. The elasticity of end-of-kindergarten mathematics, reading, and non-cognitive skills with respect to school investments are 0.323, 0.387, and 0.517, respectively. These are sizeable effects. In particular, they are 0.4, 0.7, and 0.7 times the mathematics, reading, and non-cognitive self-elasticities.

Table 3.5 reports the estimates of the baseline technology including parental investments. The table shows that including parental investments in the technology has negligible effects on the baseline skill formation estimates. For example, the elasticity of end-of-kindergarten non-cognitive skills with respect to initial non-cognitive skills changes from 0.764 to 0.762, with respect to initial mathematics skills remains unchanged at 0.142, with respect to initial reading skills changes from 0.032 to 0.022, and with respect to school investments changes from 0.517 to 0.518. Moreover, the effects of parental investments on skills are small. The elasticities of end-of-kindergarten mathematics, reading and non-cognitive skills with respect to parental investments are -0.040, 0.044, and 0.047, respectively, which are only one-tenth of the elasticities with respect to school investments. [5]

Table 3.6 presents the estimates of the value added technology (the baseline technology without the non-cognitive skills). The table demonstrates that failing to include the non-cognitive skills in the technology upwardly biases self- and cross-elasticities of the end-of-kindergarten mathematics and reading skills. After including the non-cognitive skills in the technology, the elasticity of the end-of-kindergarten mathematics skills with respect to initial mathematics skills decreases by 5%, from 0.796 to 0.759, and with respect to initial reading skills decreases by 23%, from 0.074 to 0.057.

3.4.3 Complementarity

In this section, we present the estimates for the models including complementarities. Table 3.7 shows the estimates for the baseline model. [6] The table documents a negative complementarity between initial skills and school investments. This implies that school investments are more productive for children with low initial skills

[5] Appendix C.3 demonstrates that measures of home investment are highly correlated with end-of-kindergarten skills, but, once we control for children's initial skills, their correlation with end-of-kindergarten skills goes to zero.

[6] Estimates for the baseline model including parental investments and the value-added model are available in Appendix C.4.

relative to children with high initial skills, and supports policies that target resources to disadvantaged children.

Figure 3.3 displays the implied elasticities of school investments against the deciles of child's initial skills. [7] The figure demonstrates that the elasticities of school investments are decreasing in the child's skills at kindergarten entry. For example, a 1% increase in school investments increases the end-of-kindergarten mathematics, reading, and non-cognitive skills of a child in the lowest decile of initial mathematics skills (the lowest decile of initial reading skills) [the lowest decile of initial non-cognitive skills] by 0.39, 0.46, and 0.59 percent (0.40, 0.50, and 0.59 percent) [0.33, 0.41, and 0.59 percent], respectively. Alternatively, the same increase in the school investments increases skills of a child in the highest decile of initial mathematics skills (the highest decile of initial reading skills) [the highest decile of initial non-cognitive skills] by only 0.20, 0.25, and 0.41 percent (0.18, 0.19, and 0.41 percent) [0.31, 0.37, and 0.45 percent], respectively. In other words, the effect of the 1% increase in school investments on end-of-kindergarten mathematics, reading, and non-cognitive skills for children in the lowest decile of initial mathematics skills (the lowest decile of initial reading) [the lowest decile of initial non-cognitive] are 2.0, 1.8, and 1.4 (2.2, 2.6, and 1.4) [1.1, 1.2, and 1.3] times larger than the effect on children in the highest decile.

Figure 3.4 displays the elasticities of school investments against deciles of child's household income for the baseline model, the baseline model including parental investments, and the value added model. The figure demonstrates that including parental investments in the baseline model have negligible effects on the elasticities of school investments. In addition, the figure shows that failing to include non-cognitive skills in the technology upwardly biases the elasticities of school investments. Moreover, the

[7] Appendix C.5 provides descriptive statistics for the elasticities of skills with respect to different inputs.

figure demonstrates that the elasticities of school investments are decreasing in the child's household income. For example, 1 percent increase in school investments increases end-of-kindergarten mathematics, reading, and non-cognitive skills of a child in the lowest decile of household income by 0.36, 043, and 0.55 percent, respectively, whereas the same increase in the school investments increases skills of a child in the highest decile of household income by 0.28, 0.34, and 0.48 percent, respectively. This implies that the elasticities of school investments for children in low-income households are 1.3, 1.3, 1.1 times larger than the elasticities for children in high-income households.

3.4.4 Classroom Sorting

Table 3.8 presents the estimates of the variance-covariance matrix of initial skills and school investments. Since the standard deviation of all inputs are normalized to one, covariances are equal to correlations. The table shows that all three of the initial skills are positively correlated. The correlation between mathematics and reading skills is particularly high and equal to 0.70. The table also demonstrates that school investments in mathematics, reading, and non-cognitive skills are positively correlated with each other. In particular, the correlation between school investments in mathematics and reading skills is 0.45, and is considerably higher than the correlation between other school investments. Note that the correlation between school investments are relatively small – smaller than correlation between skills.

Figure 3.5 plots the average school investments in mathematics, reading, and non-cognitive skills by the level of child's household income. The figure demonstrates that school investments in mathematics and reading skills have a positive gradient in household income, while school investments in non-cognitive skills does not have a gradient in household income. The gap in average school investments in mathematics and reading skills between children at the lowest and the highest income decile is 0.38 and 0.24 of a standard deviation, respectively, while the corresponding gap in the average school investments in non-cognitive skills is only 0.05 of a standard deviation.

3.4.5 Decomposing Child Skill Development

In this section, we examine the effects of an exogenous change in the level of school investments on the distribution of end-of-kindergarten mathematics, reading, and non-cognitive skills. We quantify by how much providing children with higher levels of school investments reduces the inequality among them at the end of the kindergarten year. We run three experiments. In the first experiment, we provide all children with the average school investments. In this experiment, children in lower than the average schools get higher quality schools while children in higher than the average schools get lower quality schools. In the second experiment, we provide the average school investments to all children in lower than the average schools. In the final experiment, we provide all children with the 90th percentile of school investments.

Table 3.9 presents the results of these experiments. The first experiment shows that providing all children with the average school investments decreases the standard deviation of the end-of-kindergarten distribution of mathematics, reading, and non-cognitive skills by 0.09, 0.13, and 0.09 of a standard deviation, respectively. These are equal to 9%, 14%, and 9% reduction in inequality relative to the baseline, which are substantial effects for one year of intervention. This experiment also decreases the gap between the 10th and 90th percentile of the end-of-kindergarten distribution of mathematics, reading, and non-cognitive skills by 0.22, 0.33, and 0.24 of a standard deviation, respectively. The second experiment demonstrates that the reduction in inequality is not because of the losses made by children in higher than the average quality schools, rather because of the gains made by children in lower than the average quality schools. For example, providing average school investments only to children in lower than the average schools decreases the standard deviation (the gap between the 10th and the 90th percentile) of the end-of-kindergarten distribution of mathematics, reading, and non-cognitive skills by 0.08, 0.11, and 0.08 (0.19, 0.28, and 0.23) of a standard deviation, respectively. The third experiment demonstrates that providing all children with the 90th percentile of school investments decreases the standard deviation of the end-of-kindergarten distribution of mathematics, reading, and non-cognitive skills by 0.16, 0.23, and 0.15 of a standard deviation, respectively. These are equal to 17%, 26%, and 16% reduction in inequality relative to the baseline. This experiment also decreases the gap between the 10th and the 90th percentile of the end-of-kindergarten distribution of mathematics, reading, and non-cognitive skills by 0.40, 0.59, and 0.41 of a standard deviation, respectively.

3.5 Validation and Robustness Exercises

3.5.1 Long-Term Effects

In this section, we examine the long-term effects of our estimated school investments during kindergarten. In particular, we examine the effects of school investments during kindergarten on mathematics, reading, and non-cognitive scores in the Spring of first, third, fifth, and eighth grades. Table 3.10 presents the effect of investments during the kindergarten year on future mathematics scores. The dependent and independent variables are rescaled by their respective standard deviations to facilitate interpretation of units. The table shows that one standard deviation increase in school investments in mathematics skills during the kindergarten year is associated with 0.15, 0.13, 0.12, and 0.08 of a standard deviation increase in mathematics score at the end of the first, third, fifth, and eighth grades, respectively. In addition, the table demonstrates that school investments in reading skills during the kindergarten year is not associated with an increase in mathematics score in future grades. In contrast, one standard deviation increase in school investments in non-cognitive skills during the kindergarten year is associated with 0.03, 0.05, 0.05, and 0.06 of a standard deviation increase in mathematics score at the end of the first, third, fifth, and eighth grades, respectively. Notice that the effects of school investments in mathematics skills during the kindergarten year decreases over the grades while the effects of school investments in non-cognitive skills increases over the grades. In addition, the effects of school investments during kindergarten are large. For example, one standard deviation increase in school investments in mathematics skills during kindergarten is associated with 0.08 of a standard deviation increase in mathematics score at the end of eighth grade, which is one-fifth of the effect of one standard deviation increase in mathematics score at the kindergarten entry. Appendix C.6 documents similar patterns for the effect of our estimated school investments during kindergarten on reading and non-cognitive scores in future grades.

3.5.2 Proxies for Unobservables

One of our main concerns is that unobservable components to skill development are correlated with school quality. We consider using a proxy for these unobservables that is derived from household characteristics. We include previously omitted observable characteristics in the technology of skills formation, such as child's household income, and examine their effects on the estimates. This is similar to Chetty *et al.* (2014a) that uses parental characteristics derived from matched federal tax data to ascertain whether teacher value added estimated using administrative records from a large school district are correlated with unobservables.

Table 3.11 presents the results for the baseline model. The table demonstrates that the effects of household income is small. For example, a 100,000 USD increase in household income (an increase of 2.2 standard deviations) is associated with only 0.08, 0.04, and 0.08 of a standard deviation increase in mathematics, reading, and non-cognitive skills at the end of kindergarten. Comparison between the estimates in Table 3.11 and the estimates for the baseline model in Table 3.4 shows that including a child's household income in the technology of skill formation has negligible effects on the elasticities of other inputs. For example, the elasticity of end-of-kindergarten mathematics, reading, and non-cognitive skills with respect to initial skills changes by less than 3%, and with respect to school investments changes by less than 1%. Table 3.12 shows the effect of household income on skills in more general models. Panel A demonstrates that household income is highly correlated with end-of-kindergarten skills, but as we generalize the technology of skill formation and include initial skills (Panel B), school investments (Panel C), parental investments (Panel D), and complementarities (Panel E), their correlation with end-of-kindergarten skills goes to zero.

3.5.3 Replication of Tennessee STAR

In this section, we replicate the Tennessee Student/Teacher Achievement Ratio (STAR) experiment using our model to compare the average treatment effects from our model with the experimental results from the STAR. The STAR was a four-year longitudinal class-size study in which over 11,000 students and their teachers from 79 schools were randomly assigned into one of three interventions: small class (13 to 17 students per teacher), regular class (22 to 25 students per teacher), and regular-with-aide class (22 to 25 students with a full-time teacher's aide). The interventions were initiated as the students entered kindergarten in the 1985-1986 school year and continued through third grade. [8]

We replicate the STAR by estimating a counterfactual using the ECLS. In particular, we change the class size to 15 students and estimate the average treatment effect on children's end-of-kindergarten skills using our estimated model. There are two challenges in replicating the STAR. First, we need an estimate for the effect of class size on school investments. We regress our estimated school investments on the observable characteristics of the schools to obtain an estimate for the effect of class size. Second, we need a metric that is comparable across the different assessments available in the two datasets. In particular, considering the measurement equation in (3.4), we see that children's skills map into test scores as

$$M^{ECLS}_{j,i,t,m} = \mu^{ECLS}_{j,t,m} + \lambda^{ECLS}_{j,t,m} \ln \theta_{j,i,t} + \epsilon^{ECLS}_{j,i,t,m} \tag{3.29}$$

$$M^{STAR}_{j,i,t,m'} = \mu^{STAR}_{j,t,m'} + \lambda^{STAR}_{j,t,m'} \ln \theta_{j,i,t} + \epsilon^{STAR}_{j,i,t,m'} \tag{3.30}$$

where the location and the scale of the ECLS and STAR test scores can be different from each other. The average treatment effect of the STAR on the test scores is the expected difference in test scores between the treatment group (τ=1) and the control group (τ=0), which maps into the average treatment effect in skills as

$$\begin{aligned} E\left[TE^{STAR}\right] &= E\left[M^{STAR}_{j,i,1,m'}|\tau=1\right] - E\left[M^{STAR}_{j,i,1,m'}|\tau=0\right] \\ &= \lambda^{STAR}_{j,1,m'}\Bigg(E\left[\ln\theta_{j,i,1}|\tau=1\right] - E\left[\ln\theta_{j,i,1}|\tau=0\right]\Bigg) \end{aligned} \tag{3.31}$$

Note that the treatment effect depends on the scale of the test score, $\lambda^{STAR}_{j,t,m'}$, which creates a problem when comparing the findings from the STAR experiment with the estimated counterfactuals from the ECLS. However, the relative average treatment effect between two subpopulations is free from the loading factor and can be compared between the two datasets. To see this, consider the binary variable X,

[8] See Krueger (1999) for a comprehensive summary of the experiment.

which takes a value of 1, for example, if the child qualifies for free or reduced-priced lunch, and 0 otherwise. We define the relative average treatment effect for these two subpopulations as

$$
\begin{aligned}
\frac{E\left[TE^{STAR}|X=1\right]}{E\left[TE^{STAR}|X=0\right]} &= \frac{E\left[M^{STAR}_{j,i,1,m'}|X=1,\,\tau=1\right]-E\left[M^{STAR}_{j,i,1,m'}|X=1,\,\tau=0\right]}{E\left[M^{STAR}_{j,i,1,m'}|X=0,\,\tau=1\right]-E\left[M^{STAR}_{j,i,1,m'}|X=0,\,\tau=0\right]} \\
&= \frac{E\left[\ln\theta_{j,i,1}|X=1,\,\tau=1\right]-E\left[\ln\theta_{j,i,1}|X=1,\,\tau=0\right]}{E\left[\ln\theta_{j,i,1}|X=0,\,\tau=1\right]-E\left[\ln\theta_{j,i,1}|X=0,\,\tau=0\right]} \qquad (3.32)
\end{aligned}
$$

which is independent of the loading factor. Notice that if the average treatment effect in subpopulation 1 is equal to (greater than) [smaller than] the average treatment effect in subpopulation 0, then the relative average treatment effect is equal to (greater than) [smaller than] 1.

We estimate the relative average treatment effect on children's end-of-kindergarten skills using our estimated model as

$$
\frac{E\left[TE^{ECLS}|X=1\right]}{E\left[TE^{ECLS}|X=0\right]} = \frac{E\left[\ln\theta_{j,i,1}|X=1,\,\tau=1\right]-E\left[\ln\theta_{j,i,1}|X=1,\,\tau=0\right]}{E\left[\ln\theta_{j,i,1}|X=0,\,\tau=1\right]-E\left[\ln\theta_{j,i,1}|X=0,\,\tau=0\right]} \qquad (3.33)
$$

where the average treatment effect for each subpopulation is given by

$$
\begin{aligned}
E[TE^{ECLS}|X] = {} & \gamma_{2,j}\Big(E\left[\ln S_{j,i,0}|X,\,\tau=1\right]-E\left[\ln S_{j,i,0}|X,\,\tau=0\right]\Big) \qquad (3.34) \\
& + \sum_{k=1}^{J}\gamma_{4,j,k}\Big(E\left[\widetilde{M}_{k,i,0,m}\ln S_{j,i,0}|X,\,\tau=1\right]-E\left[\widetilde{M}_{k,i,0,m}\ln S_{j,i,0}|X,\,\tau=0\right]\Big) \\
& + \gamma_{6,j}\Big(E\left[\widetilde{M}_{H,i,0,m}\ln S_{j,i,0}|X,\,\tau=1\right]-E\left[\widetilde{M}_{H,i,0,m}\ln S_{j,i,0}|X,\,\tau=0\right]\Big)
\end{aligned}
$$

We compare the relative average treatment effect from the STAR in (3.32) with the ECLS in (3.33). The emphasis is on whether the relative average treatment effect from both the STAR and the ECLS are on the same side of 1. [9]

Table 3.13 demonstrates that reducing class size by one student is associated with 0.014 of a standard deviation increase in school investments in both mathematics

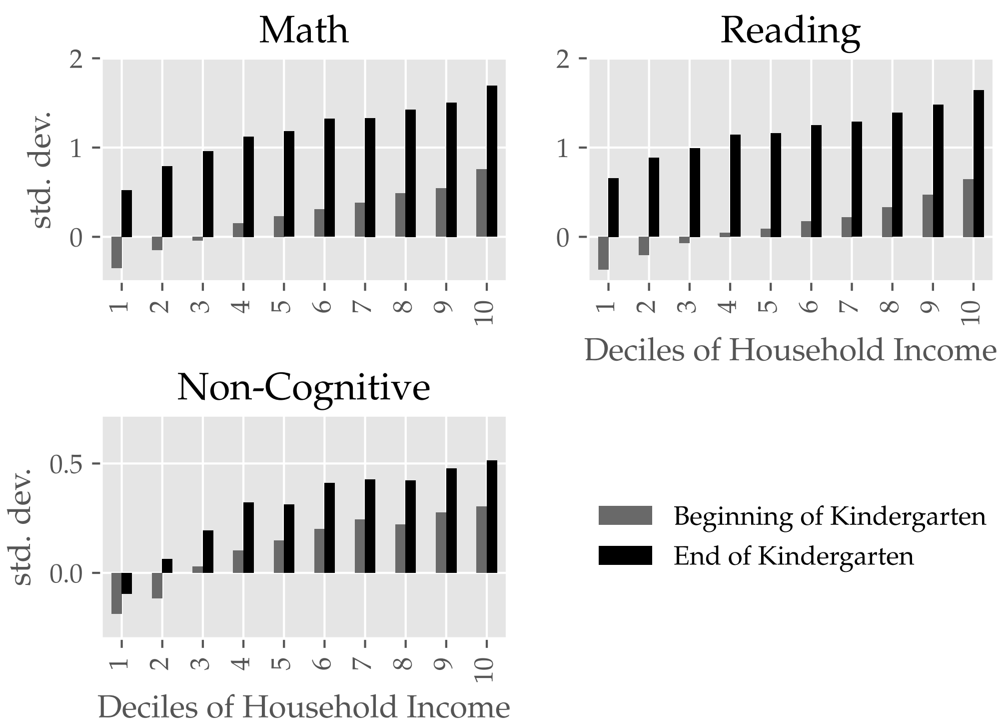
[9] Note that the STAR provides the total effect of an exogenous change in the class size, while the model provides the partial effect of change in class size (see Todd and Wolpin, 2003).

and reading skills. Interestingly, the effect of reducing class size by one student is about the same as effect of increasing instructional time by one hour per week for school investments in mathematics and two hours per week for school investments in reading. Table 3.14 reports the relative average treatment effects estimated using the ECLS and the STAR. The top panel displays the results for the subpopulation of children that qualify for free or reduced-priced lunch and children who do not. The bottom panel displays the results for subpopulations of white and non-white children. The top panel demonstrates that the relative average treatment effect on math and reading skills for children who qualify for free or reduced-priced lunch to children who do not in the STAR are 1.18 and 1.58, respectively, while the corresponding estimates from the model are 1.35 and 1.33. The bottom panel shows that the relative average treatment effect on math and reading skills for non-white to white children in the STAR are 1.19 and 1.99, respectively, while the corresponding estimates from the model are 1.44 and 1.41.

3.6 Conclusion

We develop an empirical framework that is general enough to nest many of the key features of two previously separate and parallel research programs, the Child Development literature and the Education Production Function literature. Our framework allows for both classroom and parental in influences, imperfect measures of both skills and investments, cognitive and non-cognitive skills, and complementarities between children's skills and investments from home and school. We find that investment from schools are an important determinant of children's skills at the end of kindergarten, whereas parental investments, although strongly correlated with end-of-kindergarten outcomes, have smaller effects. In addition, we document a negative complementarity between children's skills at kindergarten entry and investments from schools, implying that low-skill children benefit the most from an increase in the quality of schools. The counterfactual policy experiments show that providing all children with the 90th percentile of school investments decreases the standard deviation of the end-of-kindergarten distribution of skills by about 0.2 of a standard deviation, and decreases the gap between the 10th and the 90th percentile of the end-of-kindergarten distribution of skills by 0.5 of a standard deviation.

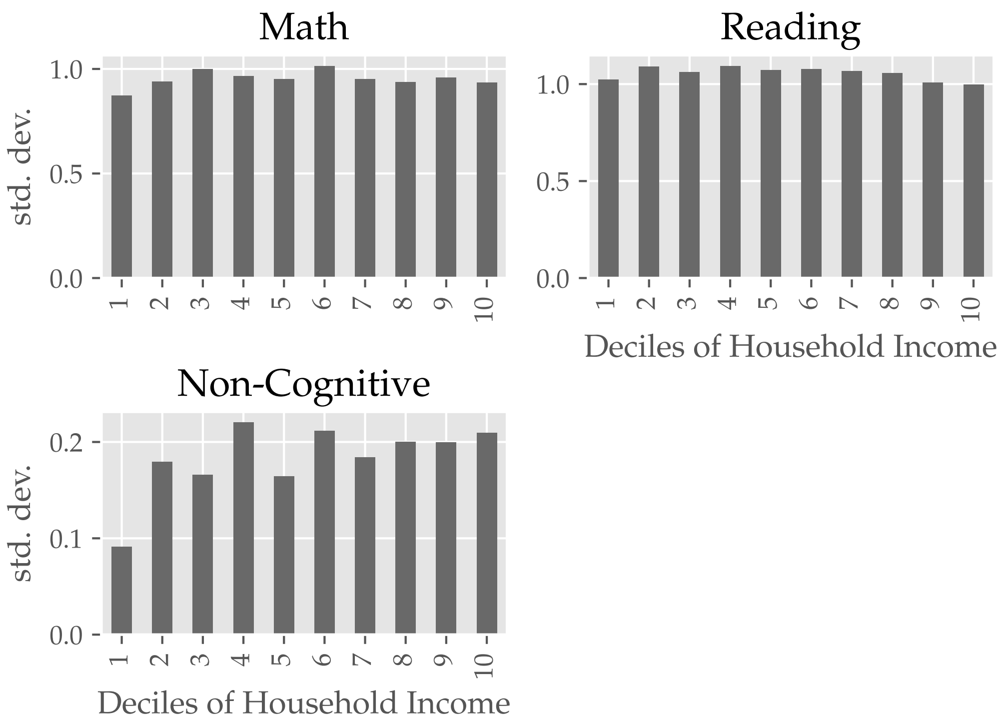
Figure 3.1: Cognitive and Non-cognitive Scores Over Income Deciles



Source: the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99.

Notes: The Math and Reading scores are raw. The non-cognitive score is the approaches to learning evaluated by the teacher. The scores are standardized using the mean and the standard deviation of scores at k-entry. (Let $M_{i,t}$ denote a score for child i in round $t = 0, 1$ (0 is entry, 1 is exit). Then, the child's standardized score is $\hat{M}_{i,t} = (M_{i,t} - \mu_0)/\sigma_0$, where μ_0 is the estimated sample mean at k-entry, and σ_0 is the estimated standard deviation at k-entry).

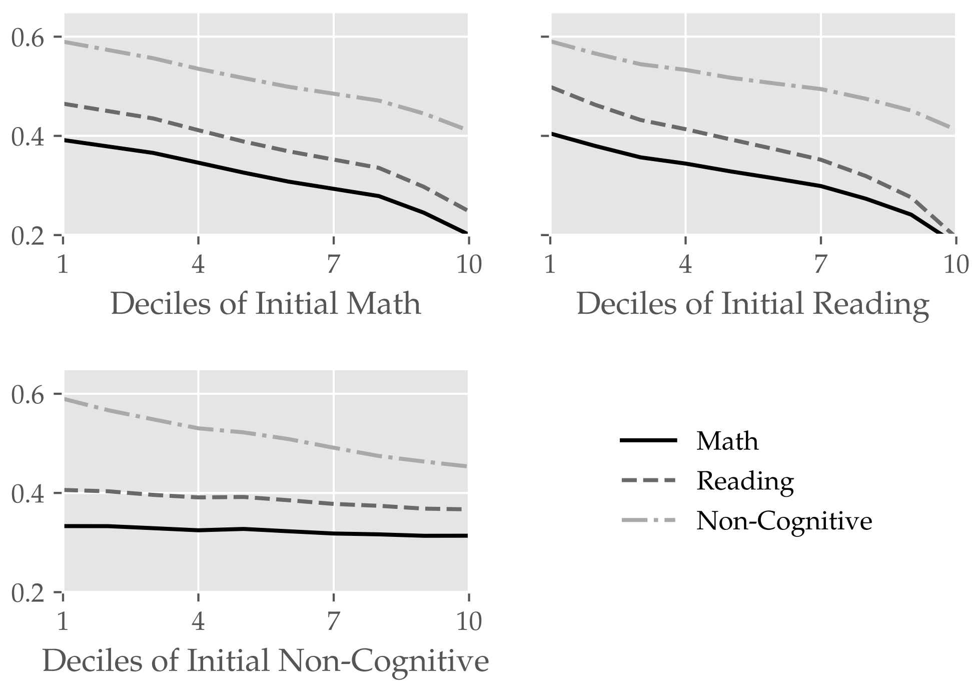
Figure 3.2: Change in Cognitive and Non-cognitive Scores During K



Source: the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99.

Notes: The Math and Reading scores are raw. The non-cognitive score is the approaches to learning evaluated by the teacher. The scores are standardized using the mean and the standard deviation of scores at k-entry. (Let $M_{i,t}$ denote a score for child i in round $t = 0, 1$ (0 is entry, 1 is exit). Then, the child's standardized score is $\hat{M}_{i,t} = (M_{i,t} - \mu_0)/\sigma_0$, where μ_0 is the estimated sample mean at k-entry, and σ_0 is the estimated standard deviation at k-entry). The change is the difference between k-exit and k-entry standardized scores ($\hat{M}_{i,1} - \hat{M}_{i,0}$).

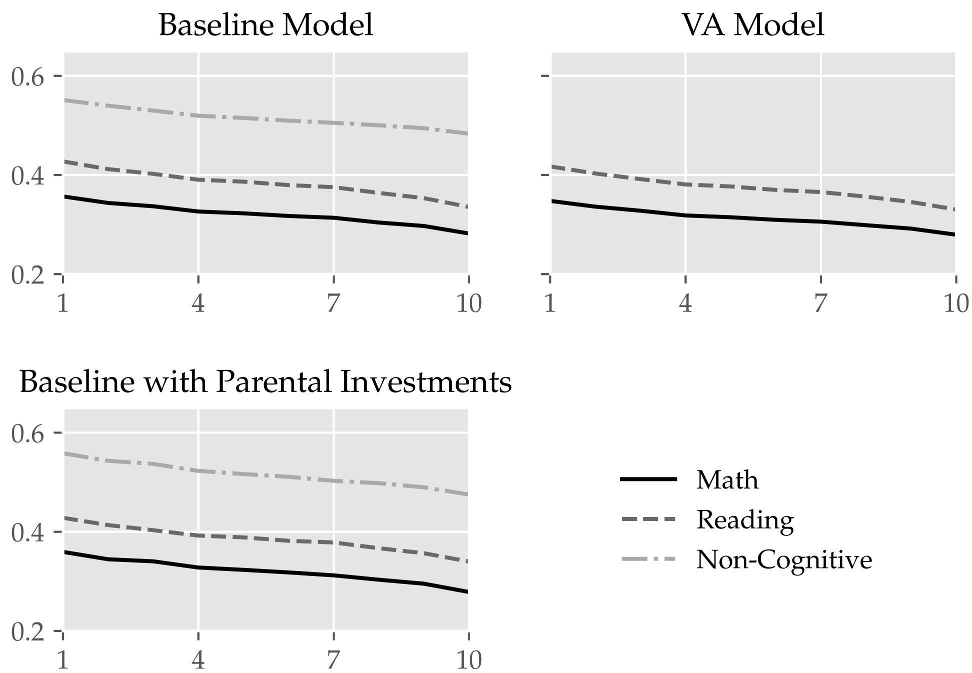
Figure 3.3: Elasticities of Investments from School



Source: Model estimates using a sample of ECLS data.

Note: Results pertain to the baseline model with complementarities (see Table 3.7).

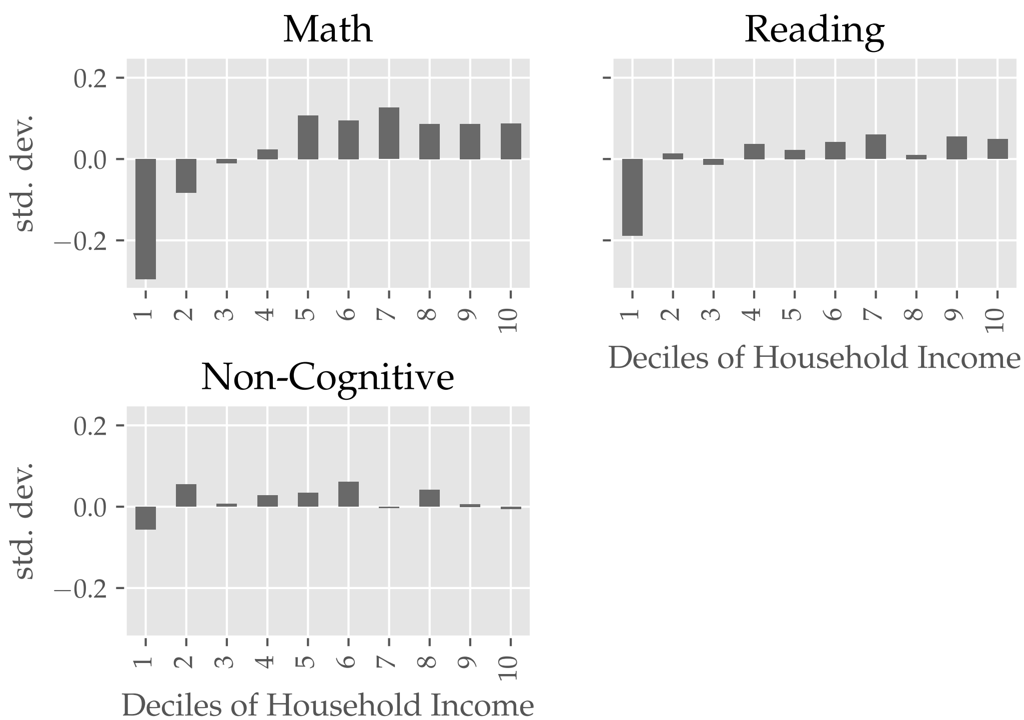
Figure 3.4: Elasticities of Investments from School Against Household Income



Source: Model estimates using a sample of ECLS data.

Note: Results pertain to the models with complementarities (see Tables 3.7, C.4, and C.5).

Figure 3.5: School/Classroom Sorting: Estimated Investments



Source: Model estimates using a sample of ECLS data.

Notes: School Investments pertain to the model with parental investments and complementarities (see Table C.4).

3.8 Tables

Table 3.1: Descriptive Statistics at Kindergarten Entry

	Mean	Std. Dev.
A: Characteristics of Children (N=9474)		
Age	5.68	0.36
Fraction male	0.51	
Fraction White, Non-Hispanic	0.67	
Fraction Black, Non-Hispanic	0.14	
Fraction Hispanic	0.10	
Fraction living with both biological parents	0.68	
Mother's age	33.81	6.41
Mother's years of schooling	13.83	2.24
Mother's hours worked	26.08	19.12
Household income (10,000 2005USD)	6.87	4.47
B: Characteristics of Classrooms (N=1211)		
Class size	20.11	4.60
Instructional time (hours/week)	24.18	9.25
Teacher's years of experience teaching K	9.57	7.82
Fraction of teachers having at least a master's degree	0.36	
C: Characteristics of Schools (N=668)		
Fraction public school	0.70	
Fraction of students receiving free or reduced-price lunch	0.27	0.27

Source: the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS).

Table 3.2: Children Scores and Home Investments at K-Entry

	Obs	Mean	Std. Dev.	Min	Max
A: Mathematics Skills					
math; raw	9,474	5.06	2.95	0	16
math; i.r.t.	9,474	27.43	9.24	10.5	96.0
B: Reading Skills					
reading; raw	9,474	6.18	3.96	0	20
reading; i.r.t.	9,474	36.11	10.39	21.5	138.5
C: Non-cognitive Skills					
approaches to learning	9,474	3.03	0.66	1	4
self-control	9,474	3.10	0.61	1	4
interpersonal skills	9,226	3.01	0.63	1	4
D: Home Investments					
reading to the child	9,474	5.16	2.02	0	7
child reading picture book	9,474	5.30	2.06	0	7
child reading	9,474	4.35	2.35	0	7

Source: the Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS).

Notes: The math and reading raw scores are a count of the number of items a child answers correctly on the routing test. The math and reading IRT score are an estimate of the number of items that the child answers correctly if she takes all of the questions on all forms. Non-cognitive scores are rated by the teacher on a scale of 1 to 4: 1=Never, 2=Sometimes, 3=Often, 4=Very Often. Home investments are the frequency of an activity during a representative week: 0=Never, 1.5=Once or Twice per Week, 4.5=Three to Six Times per Week, 7=Every Day.

Table 3.3: Measurement Parameters Estimates at K-Entry

Latent	Parameter	Measure 1	Measure 2	Measure 3
Mathematics	location	5.06	27.43	
	scale	2.85	8.86	
	signal to noise ratio	0.93	0.92	
Reading	location	6.18	36.11	
	scale	3.93	9.48	
	signal to noise ratio	0.98	0.83	
Non-Cognitive	location	3.03	3.10	3.01
	scale	0.50	0.53	0.56
	signal to noise ratio	0.58	0.76	0.81
Home Investments	location	5.30	5.16	4.35
	scale	1.61	1.02	1.24
	signal to noise ratio	0.61	0.25	0.28

Source: Model estimates using a sample of ECLS data.

Notes: The estimates pertain to the initial period ($t = 0$). For each measure m of latent ω at time t, the location is $\mu_{\omega,t,m}$, the scale is $\lambda_{\omega,t,m}$, and the signal-to-noise ratio is $1 - var(\epsilon_{\omega,t,m})/var(M_{\omega,t,m})$. See the measurement equation 3.4 for more details.

Table 3.4: Baseline Skill Formation Estimates

	$\log M_1$	$\log R_1$	$\log N_1$
$\log M_0$	0.759***	0.225***	0.142***
	[0.732,0.782]	[0.197,0.253]	[0.111,0.174]
$\log R_0$	0.057***	0.543***	0.032**
	[0.028,0.083]	[0.515,0.566]	[0.003,0.060]
$\log N_0$	0.080***	0.073***	0.764***
	[0.060,0.102]	[0.048,0.098]	[0.732,0.797]
$\log S_0$	0.323***	0.387***	0.517***
	[0.307,0.339]	[0.370,0.408]	[0.491,0.555]
N-Children	9474	9474	9474
N-Classroom	1211	1211	1211

Source: Model estimates using a sample of ECLS data.

Notes: M, R, N, and S represent mathematics skills, reading skills, non-cognitive skills, and school investments, respectively. Subscript 0 and 1 represent the beginning and the end of the kindergarten year. All models control for age, age squared, and the number of days between the two rounds of the data collection. Bootstrapped classroom-clustered 95% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 3.5: Baseline Skill Formation Estimates w. Parental Investments

	$\log M_1$	$\log R_1$	$\log N_1$
$\log M_0$	0.752***	0.222***	0.142***
	[0.726,0.778]	[0.195,0.250]	[0.108,0.175]
$\log R_0$	0.059***	0.532***	0.022
	[0.030,0.085]	[0.506,0.555]	[-0.007,0.055]
$\log N_0$	0.081***	0.071***	0.762***
	[0.062,0.103]	[0.046,0.096]	[0.733,0.795]
$\log H_0$	-0.040***	0.044***	0.047***
	[-0.064,-0.012]	[0.023,0.069]	[0.013,0.071]
$\log S_0$	0.322***	0.389***	0.518***
	[0.307,0.338]	[0.372,0.410]	[0.492,0.556]
N-Children	9474	9474	9474
N-Classroom	1211	1211	1211

Source: Model estimates using a sample of ECLS data.

Notes: M, R, N, H, and S represent mathematics skills, reading skills, non-cognitive skills, home investments, and school investments, respectively. Subscript 0 and 1 represent the beginning and the end of the kindergarten year. All models control for age, age squared, mother's years of education, and the number of days between the two rounds of the data collection. Bootstrapped classroom-clustered 95% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 3.6: VA Skill Formation Estimates

	$\log M_1$	$\log R_1$
$\log M_0$	0.796***	0.258***
	[0.778,0.820]	[0.231,0.288]
$\log R_0$	0.074***	0.559***
	[0.046,0.100]	[0.531,0.580]
$\log S_0$	0.317***	0.381***
	[0.301,0.334]	[0.366,0.402]
N-Children	9474	9474
N-Classroom	1211	1211

Source: Model estimates using a sample of ECLS data.

Notes: M, R, and S represent mathematics skills, reading skills, and school investments, respectively. Subscript 0 and 1 represent the beginning and the end of the kindergarten year. All models control for child's age, age squared, and the number of days between the two rounds of the data collection. Bootstrapped classroom-clustered 95% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 3.7: Skill Formation Estimates w. Complementarities

	$\log M_1$	$\log R_1$	$\log N_1$
$\log M_0$	0.757***	0.219***	0.141***
	[0.729,0.780]	[0.195,0.247]	[0.109,0.173]
$\log R_0$	0.064***	0.552***	0.030*
	[0.036,0.089]	[0.525,0.576]	[-0.002,0.059]
$\log N_0$	0.081***	0.074***	0.764***
	[0.061,0.102]	[0.048,0.098]	[0.729,0.797]
$\log S_0$	0.324***	0.387***	0.518***
	[0.307,0.340]	[0.371,0.412]	[0.495,0.555]
$\log M_0 \times \log S_0$	-0.037*	-0.025*	-0.032**
	[-0.076,0.002]	[-0.046,0.001]	[-0.071,-0.005]
$\log R_0 \times \log S_0$	-0.044***	-0.073***	-0.018
	[-0.078,-0.012]	[-0.092,-0.055]	[-0.044,0.015]
$\log N_0 \times \log S_0$	0.018	0.018	-0.019
	[-0.006,0.046]	[-0.006,0.049]	[-0.053,0.018]
N-Children	9474	9474	9474
N-Classroom	1211	1211	1211

Source: Model estimates using a sample of ECLS data.

Notes: M, R, N, and S represent mathematics skills, reading skills, non-cognitive skills, and school investments, respectively. Subscript 0 and 1 represent the beginning and the end of the kindergarten year. All models control for age, age squared, and the number of days between the two rounds of the data collection. Bootstrapped classroom-clustered 95% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 3.8: Initial Conditions Estimates

		(1)	(2)	(3)	(4)	(5)	(6)
M_0	(1)	1.00					
R_0	(2)	0.70	1.00				
N_0	(3)	0.54	0.48	1.00			
$S_{M,0}$	(4)	0.10	0.11	-0.07	1.00		
$S_{R,0}$	(5)	0.07	0.11	-0.09	0.45	1.00	
$S_{N,0}$	(6)	-0.02	-0.04	-0.12	0.08	0.11	1.00

Source: Model estimates using a sample of ECLS data.

Notes: M_0, R_0, and N_0 represent mathematics, reading, and non-cognitive skills at kindergarten entry, respectively. $S_{M,0}$, $S_{R,0}$, and $S_{N,0}$ represent school investments in mathematics, reading, and non-cognitive skills, respectively. Results pertain to the baseline model with complementarities (see Table 3.7).

Table 3.9: Effect of School Investments on the Distribution of Skills at K-Exit

Counterfactual	Std. Dev.	10−90 Gap
A: Mathematics Skills		
baseline	0.95	2.42
mean quality for all classrooms	0.86	2.20
mean quality for classrooms below the mean	0.87	2.23
90pct quality for all classrooms	0.79	2.02
B: Reading Skills		
baseline	0.90	2.30
mean quality for all classrooms	0.77	1.97
mean quality for classrooms below the mean	0.79	2.02
90pct quality for all classrooms	0.67	1.71
C: Non-cognitive Skills		
baseline	0.96	2.48
mean quality for all classrooms	0.87	2.24
mean quality for classrooms below the mean	0.88	2.25
90pct quality for all classrooms	0.81	2.07

Source: Model estimates using a sample of ECLS data.

Note: Results pertain to the baseline model with complementarities (see Table 3.7).

Table 3.10: Effect of Investments in Kindergarten on Future Mathematics Outcomes

	Spring G1	Spring G3	Spring G5	Spring G8
M_0	0.508***	0.553***	0.535***	0.443***
	(0.011)	(0.013)	(0.015)	(0.017)
R_0	0.043***	0.035***	0.030**	0.071***
	(0.010)	(0.012)	(0.014)	(0.017)
N_0	0.162***	0.129***	0.114***	0.085***
	(0.011)	(0.012)	(0.013)	(0.015)
$\log S_{M,0}$	0.148***	0.130***	0.122***	0.080***
	(0.013)	(0.014)	(0.017)	(0.019)
$\log S_{R,0}$	0.034***	0.001	-0.010	0.006
	(0.011)	(0.013)	(0.015)	(0.017)
$\log S_{N,0}$	0.032***	0.048***	0.054***	0.065***
	(0.010)	(0.011)	(0.014)	(0.015)
N	8019	6934	5539	4697

Source: Model estimates using a sample of ECLS data.

Notes: M_0, R_0, and N_0 represent mathematics, reading, and non-cognitive scores at kindergarten entry, respectively. Mathematics and reading scores are the raw scores. Non-cognitive score is the Approaches to Learning evaluated by the teacher. $S_{M,0}$, $S_{R,0}$, and $S_{N,0}$ represent school investments in mathematics, reading, and non-cognitive skills, respectively. School investments corresponds to the baseline model with complementarities (see Table 3.7). Scores are rescaled to have standard deviation of 1. Classroom-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 3.11: Selection on Observables - Baseline Technology

	$\log M_1$	$\log R_1$	$\log N_1$
$\log M_0$	0.755***	0.222***	0.139***
	[0.728,0.779]	[0.194,0.251]	[0.107,0.172]
$\log R_0$	0.056***	0.542***	0.031**
	[0.027,0.082]	[0.514,0.566]	[0.002,0.059]
$\log N_0$	0.080***	0.072***	0.764***
	[0.060,0.101]	[0.046,0.098]	[0.731,0.797]
$\log S_0$	0.321***	0.387***	0.517***
	[0.306,0.337]	[0.370,0.408]	[0.491,0.555]
income (100k)	0.076***	0.037*	0.075***
	[0.031,0.118]	[-0.007,0.079]	[0.021,0.137]
N-Children	9474	9474	9474
N-Classroom	1211	1211	1211

Source: Model estimates using a sample of ECLS data.

Notes: M, R, N, and S represent mathematics skills, reading skills, non-cognitive skills, and school investments, respectively. Subscript 0 and 1 represent the beginning and the end of the kindergarten year. All models control for age, age squared, and the number of days between the two rounds of the data collection. Bootstrapped classroom-clustered 95% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 3.12: Selection on Observables - Generalized Technologies

	$\log M_1$	$\log R_1$	$\log N_1$
A: with No Controls			
income (100k)	0.730***	0.605***	0.500***
	[0.667,0.793]	[0.542,0.668]	[0.419,0.581]
B: with Initial Skills			
income (100k)	0.094***	0.039	0.063**
	[0.061,0.133]	[-0.009,0.089]	[0.005,0.138]
C: with Initial Skills and Schools			
income (100k)	0.076***	0.037*	0.075***
	[0.031,0.118]	[-0.007,0.079]	[0.021,0.137]
D: with Initial Skills, Schools, and Parental Investments			
income (100k)	0.058***	0.022	0.069***
	[0.014,0.102]	[-0.023,0.063]	[0.014,0.129]
E: with Initial Skills, Schools, Parental Investments, and Comp.			
income (100k)	0.054***	0.017	0.071***
	[0.009,0.097]	[-0.026,0.060]	[0.017,0.132]

Source: Model estimates using a sample of ECLS data.

Notes: Each entry is the estimate from a separate regression. M_1, R_1, and N_1 represent mathematics, reading, non-cognitive skills at the end of the kindergarten, respectively. All models control for age, age squared, and the number of days between the two rounds of the data collection. Models in Panels D and E control for mother's years of education. Bootstrapped classroom-clustered 95% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 3.13: Determinants of Investments from School

	$S_{M,0}$	$S_{R,0}$	$S_{N,0}$
instructional time (hours/week)	0.011***	0.028***	-0.005
	(0.004)	(0.004)	(0.004)
number of students in class	-0.016**	-0.015**	-0.000
	(0.008)	(0.008)	(0.007)
teacher has a graduate degree	-0.079	0.024	0.025
	(0.068)	(0.067)	(0.068)
teacher's years of experience	-0.001	0.003	0.005
	(0.004)	(0.004)	(0.004)
N-Classrooms	1211	1211	1211

Source: Model estimates using a sample of ECLS data.

Notes: $S_{M,0}$, $S_{R,0}$, and $S_{N,0}$ represent school investments in mathematics, reading, and non-cognitive skills, respectively. School investments corresponds to the baseline model including parental investments and complementarities (see Table C.4). Regressions include indicators for observations with missing values. School-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table 3.14: Comparison of the Model Predictions with the STAR

	Model		STAR	
	Math	Reading	Math	Reading
ATE | Non-Free Lunch	0.023	0.028	8.326	5.159
ATE | Free Lunch	0.031	0.037	9.816	8.128
Ratio	1.345	1.326	1.179	1.575
ATE | White	0.021	0.026	8.280	4.833
ATE | Non-White	0.030	0.037	9.838	9.609
Ratio	1.441	1.405	1.188	1.988

Source: Model estimates using a sample of ECLS data and the STAR.

Notes: ATE denotes the average treatment effect of the STAR.

Aaronson, D., L. Barrow and W. Sander, 'Teachers and student achievement in the chicago public high schools', Journal of Labor Economics **25**, 1, 95–135 (2007).

Agostinelli, F. and M. Wiswall, 'Estimating the technology of children's skill formation', NBER Working Paper , 22442 (2016a).

Agostinelli, F. and M. Wiswall, 'Identification of dynamic latent factor models: The implications of re-normalization in a model of child development', NBER Working Paper , 22441 (2016b).

Álvarez, B. and D. Miles, 'Gender effect on housework allocation: Evidence from spanish two-earner couples', Journal of Population Economics **16**, 2, 227–242 (2003).

Anderson, W. and H. Rubin, 'Statistical inference in factor analysis', Proc. Third Berkeley Symp. on Math. Statist. and Prob **5**, 111–150 (1956).

Arcidiacono, P., G. Foster, N. Goodpaster and J. Kinsler, 'Estimating spillovers using panel data, with an application to the classroom', Quantitative Economics **3**, 3, 421–470 (2012).

Argys, L. M., D. I. Rees, S. L. Averett and B. Witoonchart, 'Birth order and risky adolescent behavior', Economic Inquiry **44**, 2, 215–233 (2006).

Barclay, K. and M. Kolk, 'Birth order and mortality: A population-based cohort study', Demography **52**, 2, 613–639 (2015).

Becker, G. S. and H. G. Lewis, 'On the interaction between the quantity and quality of children', Journal of political Economy **81**, 2, Part 2, S279–S288 (1973).

Becker, G. S. and N. Tomes, 'Child endowments and the quantity and quality of children', Journal of political Economy **84**, 4, Part 2, S143–S162 (1976).

Bernal, P., N. Mittag and J. A. Qureshi, 'Estimating effects of school quality using multiple proxies', Labour Economics **39**, 1 – 10 (2016).

Black, S. E., P. J. Devereux and K. G. Salvanes, 'The more the merrier? the effect of family size and birth order on children's education', The Quarterly Journal of Economics **120**, 2, 669–700 (2005).

Black, S. E., P. J. Devereux and K. G. Salvanes, 'Healthy (?), wealthy, and wise: Birth order and adult health', Economics & Human Biology **23**, 27–45 (2016).

Black, S. E., E. Grönqvist and B. Öckert, 'Born to lead? the effect of birth order on non-cognitive abilities', Tech. rep., National Bureau of Economic Research (2017).

Blake, J., 'Family size and the quality of children', Demography **18**, 4, 421–442 (1981).

Bond, T. N. and K. Lang, ‘The black-white education-scaled test-score gap in grades k-7’, NBER Working Paper , 19243 (2013a).

Bond, T. N. and K. Lang, ‘The evolution of the black-white test score gap in grades k3: The fragility of results’, Review of Economics and Statistics **95**, 5, 1468–1479 (2013b).

Booth, A. L. and H. J. Kee, ‘Birth order matters: the effect of family size and birth order on educational attainment’, Journal of Population Economics **22**, 2, 367–397 (2009).

Breining, S. N., J. J. Doyle Jr, D. N. Figlio, K. Karbownik and J. Roth, ‘Birth order and delinquency: Evidence from denmark and florida’, Tech. rep., National Bureau of Economic Research (2017).

Browning, M., F. Bourguignon, P.-A. Chiappori and V. Lechene, ‘Income and outcomes: A structural model of intrahousehold allocation’, Journal of political Economy **102**, 6, 1067–1096 (1994).

Bu, F., ‘Sibling configurations, educational aspiration and attainment’, Tech. rep., ISER Working Paper Series (2014).

Buckles, K. and S. Kolka, ‘Prenatal investments, breastfeeding, and birth order’, Social Science & Medicine **118**, 66–70 (2014).

Buckles, K. S. and E. L. Munnich, ‘Birth spacing and sibling outcomes’, Journal of Human Resources **47**, 3, 613–642 (2012).

Carneiro, P., K. T. Hansen and J. J. Heckman, ‘Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college’, International Economic Review (2003).

Chamberlain, G., ‘Education, income, and ability revisited’, Journal of Econometrics pp. 241–257 (1977a).

Chamberlain, G., ‘An instrumental variable interpretation of identification in variance components and mimic models’, in Paul Taubman, ed., Kinometrics: Determinants of Socio-Economic SuccessWithin and Between Families (Amsterdam: North-Holland) (1977b).

Chamberlain, G. and Z. Griliches, ‘Unobservables with a variance-components structure: Ability, schooling, and the economic success of brothers’, International Economic Review **16**, 422–449 (1975).

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach and D. Yagan, ‘How does your kindergarten classroom affect your earnings? evidence from project star *’, The Quarterly Journal of Economics **126**, 4, 1593–1660 (2011).

Chetty, R., J. N. Friedman and J. E. Rockoff, ‘Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates’, American Economic Review **104**, 9 (2014a).

Chetty, R., J. N. Friedman and J. E. Rockoff, 'Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood', American Economic Review **104**, 9, 2633–79 (2014b).

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

Cunha, F. and J. J. Heckman, 'Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation', Journal of human resources **43**, 4, 738–782 (2008).

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

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

Del Boca, D., C. Flinn and M. Wiswall, 'Transfers to households with children and child development', The Economic Journal **126**, 596 (2016).

Flèche, S., 'Teacher quality, test scores and non-cognitive skills', CEP Discussion Papers (2017).

Goldberger, A. S., 'Structural equation methods in the social sciences', Econometrica **40**, 979–1001 (1972).

Hao, J. V. H., Lingxin and G. Z. Jin, 'Games parents and adolescents play: Risky behaviour, parental reputation and strategic transfers', The Economic Journal **118**, 528, 515–555 (2008).

Heckman, J. J. and D. V. Masterov, 'The productivity argument for investing in young children', Applied Economic Perspectives and Policy **29**, 3, 446–493 (2007).

Heckman, J. J. and S. Mosso, 'The economics of human development and social mobility', NBER Working Paper , 19925 (2014).

Heckman, J. J. and Y. Rubinstein, 'The importance of noncognitive skills: Lessons from the ged testing program', The American Economic Review **91**, 2, 145–149 (2001).

Heckman, J. J., J. Stixrud and S. Urzua, 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', Journal of Labor economics **24**, 3, 411–482 (2006).

Heiland, F., 'Does the birth order affect the cognitive development of a child?', Applied Economics **41**, 14, 1799–1818 (2009).

Hersch, J. and L. S. Stratton, 'Housework, wages, and the division of housework time for employed spouses', The American Economic Review **84**, 2, 120–125 (1994).

Hotz, J. V. and J. Pantano, 'Strategic parenting, birth order, and school performance', Journal of population economics **28**, 4, 911–936 (2015).

Jackson, K. C., 'Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in north carolina', NBER Working Paper , 18624 (2012).

Jöreskog, K. G. and A. S. Goldberger, 'Estimation of a model with multiple indicators and multiple causes of a single latent variable', Journal of the American Statistical Association **70**, 631–639 (1975).

Kantarevic, J. and S. Mechoulan, 'Birth order, educational attainment, and earnings an investigation using the psid', Journal of human resources **41**, 4, 755–777 (2006).

Kleinman, J. C. and J. H. Madans, 'The effects of maternal smoking, physical stature, and educational attainment on the incidence of low birth weight', American journal of epidemiology **121**, 6, 843–855 (1985).

Krueger, A. B., 'Experimental estimates of education production functions*', The Quarterly Journal of Economics **114**, 2, 497–532 (1999).

Lehmann, J.-Y. K., A. Nuevo-Chiquero and M. Vidal-Fernandez, 'The early origins of birth order differences in childrens outcomes and parental behavior', Journal of Human Resources pp. 0816–8177 (2016).

Lise, J. and S. Seitz, 'Consumption inequality and intra-household allocations', The Review of Economic Studies **78**, 1, 328–355 (2011).

Monfardini, C. and S. G. See, 'Birth order and child outcomes: does maternal quality time matter?', IZA Discussion Paper **6825** (2012).

Pavan, R., 'On the production of skills and the birth order effect', Journal of Human Resources (2015).

Price, J., 'Parent-child quality time does birth order matter?', Journal of Human Resources **43**, 1, 240–265 (2008).

Rivkin, S. G., E. A. Hanushek and J. F. Kain, 'Teachers, schools, and academic achievement', Econometrica **73**, 2, 417–458 (2005).

Rockoff, J. E., 'The impact of individual teachers on student achievement: Evidence from panel data', The American Economic Review **94**, 2, 247–252 (2004).

Rodgers, J. L., D. C. Rowe and D. F. Harris, 'Sibling differences in adolescent sexual behavior: Inferring process models from family composition patterns', Journal of Marriage and the Family pp. 142–152 (1992).

Rosenzweig, M. R., 'Birth spacing and sibling inequality: Asymmetric information within the family', International Economic Review pp. 55–76 (1986).

Rostila, M., J. Saarela and I. Kawachi, 'Birth order and suicide in adulthood: evidence from swedish population data', American journal of epidemiology **179**, 12, 1450–1457 (2014).

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

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

Zajonc, R. B., 'Family configuration and intelligence', Science **192**, 4236, 227–236 (1976).

Zajonc, R. B., 'The family dynamics of intellectual development.', American Psychologist **56**, 6-7, 490 (2001).

Zajonc, R. B. and G. B. Markus, 'Birth order and intellectual development', Psychological review **82**, 1, 74 (1975).

APPENDIX A

THE EFFECT OF BIRTH ORDER ON CHILDHOOD DEVELOPMENT

A.1 Ceiling Effects

This section examines the sensitivity of the birth order effects on cognitive and non-cognitive outcomes to the ceiling effects. Note that only a handful of children score at the ceiling on cognitive outcomes. Specifically, 9 out of 2251, 1 out of 1960, and 0 out of 2244 children score at the ceiling on the Letter-Word Identification, the Paragraph Comprehension, and the Applied Problems test, respectively. In contrast, a substantial proportion of children score at the ceiling on non-cognitive outcomes. In particular, 1074 out of 2586 and 1927 out of 2586 score at the ceiling on the Dependent and the Peer Problems scale, respectively. To examine the sensitivity of the birth order effects to the ceiling effects, for each outcome I define an indicator that takes the value of 1 if the child's outcome is above the median conditional on age, and 0 otherwise. Note that the defined dependent variables are not sensitive to the range of the outcomes. I then re-estimate models in 1.1 and in 1.6 with the new dependent variables. The tables show that the findings are qualitatively the same, although some of the coefficients are imprecisely estimated.

Table A.1: Effect of Being Born Second on Outcomes

Outcome	OLS	FE
A: Cognitive Outcomes		
Letter-Word Identification	-0.083***	-0.179***
	(0.025) [2251]	(0.049) [2251]
Paragraph Comprehension	-0.068***	-0.064
	(0.025) [1960]	(0.056) [1960]
Applied Problems	-0.036	0.009
	(0.024) [2244]	(0.050) [2244]
B: Non-Cognitive Outcomes		
Dependent	-0.053**	-0.088**
	(0.021) [2586]	(0.044) [2586]
Peer Problems	-0.026	-0.051
	(0.021) [2586]	(0.043) [2586]

Source: the PSID-CDS.
Notes: Each entry is the estimate from a separate regression. The scale of non-cognitive outcomes is reversed so that a higher score is a better outcome. Each outcome is an indicator for whether child's test score is greater than the median conditional on age. Models control for child's gender, child's age and age squared, mother's years of education, and mother's age at birth. Number of children are in brackets. Household-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table A.2: Heterogeneity in Birth Order Effects

	Cognitive	Non-Cognitive
A: Estimates		
1{birth order=2}	-0.179***	-0.089*
	(0.049)	(0.050)
	[-0.264,-0.097]	[-0.165,-0.005]
$q \times$ 1{birth order=2}	0.011	0.007
	(0.009)	(0.011)
	[-0.001,0.029]	[-0.007,0.026]
B: Marginal Effects		
Average	-0.18	-0.09
Standard Deviation	0.11	0.01
N-Children	2251	2586
N-Households	1529	1747

Source: the PSID-CDS.

Notes: Cognitive outcome is Letter-Word identification test. Non-Cognitive outcome is Dependent scale. The scale of the non-cognitive outcome is reversed so that a higher score is a better outcome. Each outcome is an indicator for whether child's test score is greater than the median conditional on age. q is the household fixed effect and is normalized to have mean of zero standard deviation of one. Regressions control for child's gender, child's age, and child's age squared. Bootstrapped standard errors are in parentheses. 90% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

A.2 Adult Outcomes

This section merges the PSID-CDS with the Transition into Adulthood Supplement (TAS) to examine the birth order effects on adult outcomes. The TAS is a follow-up to the CDS. Its sample includes PSID participants who have completed their CDS interviews and are between 18-28 years of age. It collects information on education, employment, and income among other variables. After merging TAS with PSID-CDS, I estimate models identical to the model in 1.1 with college attendance, years of education, and log hourly wages as the dependent variable. The results show no birth order effects on adult outcomes.

Table A.3: Effect of Being Born Second on Adult Outcomes

	College Attendance	Years of Education	Log Hourly Wages
1{birth order=2}	-0.013	0.006	-0.002
	(0.026)	(0.079)	(0.072)
N-Children	1298	1517	928
N-Households	616	640	473

Source: the PSID-CDS-TAS.
Notes: Each entry is the estimate from a separate regression. All models control for family fixed effects, child's gender, child's age and age squared. Standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

A.3 Parental Passive Time

This section examines the birth order effects on parental passive time. The results are qualitatively similar to the birth order effects on parental active time presented in the main text.

Table A.4: Effect of Being Born Second on Passive Parental Time

	Alone	Shared
A: Maternal Time		
1{birth order=2}	-11.13***	13.27***
	(1.633)	(1.405)
1{birth order=2} × age	1.66***	-2.25***
	(0.327)	(0.300)
1{birth order=2} × age^2	-0.03**	0.06***
	(0.015)	(0.015)
Average Marginal Effects	1.880	-2.180
B: Paternal Time		
1{birth order=2}	-7.98***	5.49***
	(1.081)	(1.060)
1{birth order=2} × age	1.00***	-0.71***
	(0.225)	(0.237)
1{birth order=2} × age^2	-0.01	0.01
	(0.011)	(0.012)
Average Marginal Effects	0.714	-1.223
N-Children	2435	2435
N-Households	1744	1744

Source: the PSID-CDS.
Notes: Time is measured in hours per week. Models control for child's gender, child's age and age squared, mother's years of education, and mother's age at birth. Household-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

A.4 Alternative Measures of Parental Investments

This section examines the effect of birth order on other measures of parental investments available in the CDS, namely, the Household Tasks and the Parental Warmth instruments. The Household Tasks is designed to measure parental interaction with the children. The Parental Warmth is designed to measure the warmth of the relationship between parents and children. Both of these measures are categorical variables and ask parents for the frequency of a selected activities in a month. The results demonstrate that parents read more to first-born children or help them more with homework relative to second-born children. In addition, they have warmer relationships with first-born children relative to second-born children.

Table A.5: Effect of Being Born Second on Parental Investments

Investment	OLS	FE
Household Tasks Index 1	-0.070*	-0.060
(reading and homework)	(0.041)	(0.071)
	[2587]	[2587]
Household Tasks Index 2	0.097**	0.146**
(household chores)	(0.040)	(0.063)
	[2587]	[2587]
Parental Warmth Index 1	-0.078*	0.007
(appreciative words)	(0.043)	(0.060)
	[2862]	[2862]
Parental Warmth Index 2	-0.140***	-0.124*
(participation in child activities)	(0.039)	(0.066)
	[2862]	[2862]

Source: the PSID-CDS.
Notes: Each entry is the estimate from a separate regression. Investments are rescaled by standard deviation at age seven. Models control for child's gender, child's age and age squared, mother's years of education, and mother's age at birth. Number of children are in brackets. Household-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

A.5 Children Allocation of Time

This section examines the effect of birth order on children's allocation of time. Second-born children spend about 1.2 hours per week more on watching TV relative to their first-born siblings, which is about the amount of time that second-born children receive less from their parents.

Table A.6: Effect of being Born Second on Children's Allocation of Time

	OLS
Sleeping	0.01
	(0.39)
School	-0.82
	(0.59)
Watching TV	1.17***
	(0.41)
Housework	0.31
	(0.29)
Personal Care	-0.42
	(0.42)
Socializing	0.03
	(0.25)
Leisure	-0.38
	(0.47)
Other	0.10
	(0.20)
N-Children	2435
N-Households	1744

Source: the PSID-CDS.
Notes: Each entry is the estimate from a separate regression. Time is measured in hours per week. Models control for child's gender, child's age and age squared, mother's years of education, and mother's age at birth. Household-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

A.6 Augmented Sample

This section examines the effect of birth order on cognitive outcomes, non-cognitive outcomes, and parental investments using an augmented sample. The Augmented sample is defined as the baseline sample (households with two children) plus households with one or three children. The results demonstrates that the birth effects are qualitatively the same in the augmented sample.

Table A.7: Descriptive Statistics in 1997

	Mean	Std. Dev.
A: Characteristics of Children		
Number of children	2583	
Age	6.66	3.75
Fraction male	0.50	
Fraction White, Non-Hispanic	0.52	
Fraction Black, Non-Hispanic	0.38	
Fraction Hispanic	0.06	
Fraction other Race/Ethnicity	0.05	
Fraction living with both biological parents	0.64	
Fraction living with only biological mother	0.32	
Fraction living with only biological father	0.02	
Fraction living with no biological parents	0.01	
B: Characteristics of Households		
Number of households	1804	
Fraction having one child	0.17	
Fraction having two children	0.48	
Fraction having three childrens	0.35	
Mother's age	33.44	7.14
Mother's age at first pregnancy	23.91	5.44
Mother's years of schooling	13.75	2.13
Mother's hours worked per week	24.34	17.14
Household income (10,000 1996USD)	4.72	4.64

Source: the PSID-CDS.

Table A.8: Effect of Being Born Second on Outcomes

	(1)	(2)	(3)	(4)	(5)
A: OLS					
1{birth order = 2}	-0.100***	-0.101***	-0.138***	-0.114***	-0.091**
	(0.030)	(0.031)	(0.044)	(0.031)	(0.042)
1{birth order = 3}	-0.264***	-0.233***	-0.325***	-0.293***	-0.204***
	(0.055)	(0.054)	(0.081)	(0.057)	(0.073)
B: FE					
1{birth order = 2}	-0.118**	-0.150***	-0.191***	-0.161***	-0.057
	(0.053)	(0.058)	(0.070)	(0.058)	(0.081)
1{birth order = 3}	-0.295***	-0.337***	-0.353***	-0.423***	-0.152
	(0.103)	(0.111)	(0.136)	(0.112)	(0.157)
N-Children	4659	4062	4650	5320	5320
N-Households	3452	3080	3446	3921	3921

Source: the PSID-CDS.
Notes: Cognitive outcomes are: 1=Letter-Word, 2=Passage Comprehension, 3=Applied Problems. Non-cognitive outcomes are: 4=Dependent, 5=Peer Problems. The scale of non-cognitive outcomes is reversed so that a higher score is a better outcome. Outcomes are rescaled by standard deviation at age seven. Models control for child's gender, child's age and age squared, number of children in the household, mother's years of education, and mother's age at birth. Household-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table A.9: Effect of Being Born Second on Active Parental Time

	(1)	(2)	(3)	(4)
1{birth order = 2}	-12.732***	9.926***	-7.840***	3.859***
	(0.991)	(0.726)	(0.681)	(0.513)
1{birth order = 3}	-12.296***	8.106***	-8.684***	3.143***
	(1.400)	(1.358)	(0.834)	(1.034)
1{birth order = 2} × age	1.071***	-0.905***	0.597***	-0.395***
	(0.077)	(0.057)	(0.052)	(0.042)
1{birth order = 3} × age	1.188***	-0.974***	0.732***	-0.529***
	(0.108)	(0.099)	(0.068)	(0.071)
N-Children	4984	4984	4984	4984
N-Households	3637	3637	3637	3637

Source: the PSID-CDS.
Notes: 1=Maternal/Alone, 2=Maternal/Shared, 3=Paternal/Alone, 4=Paternal/Shared. Parental times are measured in hours per week. Models control for child's gender, child's age and age squared, number of children in the household, mother's years of education, and mother's age at birth. Household-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

APPENDIX B

THE EFFECT OF CASH TRANSFERS ON BIRTH ORDER EFFECTS

B.1 Details of the Estimates

This section provides the estimates of the underlying parameters of the model with their standard errors, ω.

Table B.1: Estimates of Primitive Parameters (ω)

	One-Child		Two-Child	
	Estimate	Std. Err.	Estimate	Std. Err.
A: Cognitive Technology				
last period skills intercept	-0.9630	(0.1102)	-0.3626	(0.0061)
last period skills slope	0.0556	(0.0062)	0.0181	(0.0005)
monetary investment intercept	-0.8994	(0.5276)	0.6049	(0.2123)
monetary investment slope	-0.2950	(0.0587)	-0.9316	(0.1463)
alone time investment intercept	0.2062	(0.1610)	-0.5522	(0.0449)
alone time investment slope	-0.0887	(0.0086)	-0.1321	(0.0068)
shared time investment intercept			-1.0476	(0.0406)
shared time investment slope			-0.0610	(0.0032)
B: Non-Cognitive Technology				
last period skills intercept	-0.2367	(0.0734)	0.0203	(0.0048)
last period skills slope	0.0134	(0.0041)	-0.0071	(0.0009)
monetary investment intercept	-37.1606	(90.2161)	-1.8129	(0.5190)
monetary investment slope	-0.0244	(0.1571)	-0.6294	(0.2204)
alone time investment intercept	-1.4940	(0.3146)	-6.5315	(2.4542)
alone time investment slope	-0.0812	(0.0227)	-0.1675	(0.0793)
shared time investment intercept			-4.0604	(0.0749)
shared time investment slope			0.1339	(0.0094)
C: Preferences				
leisure			2.6891	(0.0589)
first child skills			-0.1772	(0.0593)
second child skills			0.0007	(0.0004)
cognitive skills			0.0351	(0.0222)
terminal payoff to first child			0.5229	(0.1441)
terminal payoff to second child			-0.6220	(0.3628)

Source: Model estimates using a sample of PSID-CDS data.
Notes: Bootstrapped standard errors are in parentheses.

B.2 Model with One Child

This section describes the model with one child. In addition, it provides an algorithm to solve the model. Consider a household that has one child. Development of the child takes M periods. Household's life begins at $t_1 = 1$ with birth of the child, and ends at $t_2 = M + 1$ when the development of the child ends. The timeline of household's life events is:

Preferences

Household's preferences in time t follows a Cobb-Douglas form:

$$u(c_t, \ell_t, k_{t+1}, n_{t+1}) = \alpha_1 \log c_t + \alpha_2 \log \ell_t + \alpha_3 \alpha_5 \log k_{t+1} + \alpha_3(1 - \alpha_5) \log n_{t+1}$$

where c is consumption, ℓ is time allocated to leisure, k is the cognitive skills, and n is the non-cognitive skills. I assume that $\alpha_j > 0$ for $j \in \{1, 2, 3\}$, $\sum_{j=1}^{3} \alpha_j = 1$, and $\alpha_5 \in (0, 1)$.

Technology

Cognitive and non-cognitive skills evolve according to the following technologies:

$$\begin{aligned} k_{t+1} = &\quad k_t^{\delta_{1,t}}\, e_t^{\delta_{2,t}}\, a_t^{\delta_{3,t}} \\ n_{t+1} = &\quad n_t^{\eta_{1,t}}\, e_t^{\eta_{2,t}}\, a_t^{\eta_{3,t}} \end{aligned}$$

where e is monetary investment and a is alone time investment.

Household's Problem

Household's problem in the sequence form is:

$$\max_{\{c_t, \ell_t, e_t, a_t\}_{t=1}^{M}} \quad \sum_{t=1}^{M} \beta^{t-1} u(c_t, \ell_t, k_{t+1}, n_{t+1})$$

$$\sum_{t=1}^{M} R^{t-1}(c_t + e_t) = \mathcal{I}$$

$$\ell_t + a_t = \tau \ \forall t$$

where β is discount factor, R is the gross interest rate, $\mathcal{I}$ is the sum of present value of household's income, and τ is the time endowment.

Optimal Decisions

The optimal decisions are:

$$c_t = \frac{\varphi(c_t)}{\sum_{t=1}^{t_2}[\varphi(c_t) + \varphi(e_t)]} \sum_{t=1}^{T} R^{-(t-1)} \mathrm{I}_t$$

$$e_t = \frac{\varphi(e_t)}{\sum_{t=1}^{t_2}[\varphi(c_t) + \varphi(e_t)]} \sum_{t=1}^{T} R^{-(t-1)} \mathrm{I}_t$$

$$\ell_t = \frac{\alpha_2}{\alpha_2 + \varphi(a_t)} \tau$$

$$a_t = \frac{\varphi(a_t)}{\alpha_2 + \varphi(a_t)} \tau$$

where φs are the lifetime marginal benefits,

$$\varphi(c_t) = \beta^{t-1} R^{t-1} \alpha_1$$

$$\varphi(e_t) = \beta^{t-1} R^{t-1} \big(\alpha_3 \alpha_5 \delta_{2,t}(1 + \phi_t) + \alpha_3(1 - \alpha_5)\eta_{2,t}(1 + \psi_t)\big)$$

$$\varphi(a_t) = \big(\alpha_3 \alpha_5 \delta_{3,t}(1 + \phi_t) + \alpha_3(1 - \alpha_5)\eta_{3,t}(1 + \psi_t)\big)$$

where x is the future flows of an additional investment in cognitive skills, and y is the future flows of an additional investments in non-cognitive skills.

$$\phi_t = \sum_{i=t+1}^{T} \beta^{i-t} \prod_{j=t+1}^{i} \delta_{1,j}$$

$$\psi_t = \sum_{i=t+1}^{T} \beta^{i-t} \prod_{j=t+1}^{i} \eta_{1,j}$$

Solving for the Optimal Decisions

In this part I demonstrate how to solve the model. For simplicity, consider a model with three periods or $t_2 = 3$. The household solves:

$$\max_{\{c_t, \ell_t, e_t, a_t\}_{t=1}^{3}} \sum_{t=1}^{3} \beta^{t-1} u(c_t, \ell_t, k_{t+1}, n_{t+1})$$

$$c_1 + e_1 + R^{-1}(c_2 + e_2) + R^{-2}(c_3 + e_3) = \mathcal{I}$$

$$\ell_t + a_t = \tau \ \forall t$$

The household's objective function is:

$$
\begin{aligned}
U = \sum_{t=1}^{3} \beta^{t-1} u =& \alpha_1 \sum_{t=1}^{3} \beta^{t-1} \log c_t + \alpha_2 \sum_{t=1}^{3} \beta^{t-1} \log \ell_t \\
& + \alpha_3 \alpha_5 \sum_{t=1}^{3} \beta^{t-1} \log k_{t+1} + \alpha_3 (1-\alpha_5) \sum_{t=1}^{3} \beta^{t-1} \log n_{t+1}
\end{aligned}
$$

I use the lifetime monetary budget constraint, time budget constraints, and the cognitive and non-cognitive technologies to convert the constrained optimization problem into an unconstraint optimization problem. First, using the lifetime budget constraint we have:

$$
\begin{aligned}
\sum_{t=1}^{3} \beta^{t-1} \log c_t =& \log c_1 + \beta \log c_2 + \beta^2 \log c_3 \\
=& \log(\mathcal{I} - e_1 - R^{-1}(c_2 + e_2) - R^{-2}(c_3 + e_3)) + \beta \log c_2 + \beta^2 \log c_3
\end{aligned}
$$

Second, using the time budget constraints substitute out leisure:

$$
\begin{aligned}
\sum_{t=1}^{3} \beta^{t-1} \log \ell_t =& \log l_1 + \beta \log l_2 + \beta^2 \log l_3 \\
=& \log(\tau - a_1) + \beta \log(\tau - a_2) + \beta^2 \log(\tau - a_3)
\end{aligned}
$$

Third, for the cognitive skills we have:

$$
\begin{aligned}
\sum_{t=1}^{3} \beta^{t-1} \log k_{t+1} =& (1 + \beta\delta_{1,2} + \beta^2 \delta_{1,3}\delta_{1,2})(\delta_{2,1} \log e_1 + \delta_{3,1} \log a_1) \\
& + (\beta + \beta^2 \delta_{1,3})(\delta_{2,2} \log e_2 + \delta_{3,2} \log a_2) \\
& + \beta^2 (\delta_{2,3} \log e_3 + \delta_{3,3} \log a_3)
\end{aligned}
$$

similarly, for the non-cognitive skills:

$$
\begin{aligned}
\sum_{t=1}^{3} \beta^{t-1} \log n_{t+1} =& (1 + \beta\eta_{1,2} + \beta^2 \eta_{1,3}\eta_{1,2})(\eta_{2,1} \log e_1 + \eta_{3,1} \log a_1) \\
& + (\beta + \beta^2 \eta_{1,3})(\eta_{2,2} \log e_2 + \eta_{3,2} \log a_2) \\
& + \beta^2 (\eta_{2,3} \log e_3 + \eta_{3,3} \log a_3)
\end{aligned}
$$

Finally, substitute each element in the objective function:

$$
\begin{aligned}
U = \sum_{t=1}^{3} \beta^{t-1} u =& \alpha_1 \sum_{t=1}^{3} \beta^{t-1} \log c_t + \alpha_2 \sum_{t=1}^{3} \beta^{t-1} \log \ell_t \\
&+ \alpha_3\alpha_5 \sum_{t=1}^{3} \beta^{t-1} \log k_{t+1} + \alpha_3(1-\alpha_5) \sum_{t=1}^{3} \beta^{t-1} \log n_{t+1} \\
=& \alpha_1 \log(\mathcal{I} - e_1 - R^{-1}(c_2 + e_2) - R^{-2}(c_3 + e_3)) + \beta\alpha_1 \log c_2 + \beta^2\alpha_1 \log c_3 \\
&+ \alpha_2 \log(\tau - a_1) + \beta\alpha_2 \log(\tau - a_2) + \beta^2\alpha_2 \log(\tau - a_3) \\
&+ \alpha_3\alpha_5(1 + \beta\delta_{1,2} + \beta^2\delta_{1,3}\delta_{1,2})(\delta_{2,1} \log e_1 + \delta_{3,1} \log a_1) \\
&+ \alpha_3\alpha_5(\beta + \beta^2\delta_{1,3})(\delta_{2,2} \log e_2 + \delta_{3,2} \log a_2) \\
&+ \alpha_3\alpha_5\beta^2(\delta_{2,3} \log e_3 + \delta_{3,3} \log a_3) \\
&+ \alpha_3(1-\alpha_5)(1 + \beta\eta_{1,2} + \beta^2\eta_{1,3}\eta_{1,2})(\eta_{2,1} \log e_1 + \eta_{3,1} \log a_1) \\
&+ \alpha_3(1-\alpha_5)(\beta + \beta^2\eta_{1,3})(\eta_{2,2} \log e_2 + \eta_{3,2} \log a_2) \\
&+ \alpha_3(1-\alpha_5)\beta^2(\eta_{2,3} \log e_3 + \eta_{3,3} \log a_3)
\end{aligned}
$$

First Order conditions

The first order conditions for consumption, monetary expenditures, leisure, and parental time choices are listed below. These conditions along with the lifetime budget constraint and the time budget constraints determine the optimal allocation of resrouces.

$$
\begin{aligned}
\frac{\partial U}{\partial c_2} =& -\frac{\alpha_1 R^{-1}}{c_1} + \frac{\alpha_1 \beta}{c_2} \\
\frac{c_1}{c_2} =& \frac{R^{-1}}{\beta} \\
\frac{\partial U}{\partial c_3} =& -\frac{\alpha_1 R^{-2}}{c_1} + \frac{\alpha_1 \beta^2}{c_3} \\
\frac{c_1}{c_3} =& \frac{R^{-2}}{\beta^2}
\end{aligned}
$$

$$\frac{\partial U}{\partial e_1} = -\frac{\alpha_1}{c_1} + \frac{\alpha_3\alpha_5}{e_1}(1+\beta\delta_{1,2}+\beta^2\delta_{1,3}\delta_{1,2})\delta_{2,1} + \frac{\alpha_3(1-\alpha_5)}{e_1}(1+\beta\eta_{1,2}+\beta^2\eta_{1,3}\eta_{1,2})\eta_{2,1}$$

$$\frac{c_1}{e_1} = \frac{\alpha_1}{\alpha_3\alpha_5(1+\beta\delta_{1,2}+\beta^2\delta_{1,3}\delta_{1,2})\delta_{2,1} + \alpha_3(1-\alpha_5)(1+\beta\eta_{1,2}+\beta^2\eta_{1,3}\eta_{1,2})\eta_{2,1}}$$

$$\frac{\partial U}{\partial e_2} = -\frac{\alpha_1 R^{-1}}{c_1} + \frac{\alpha_3\alpha_5}{e_2}(\beta+\beta^2\delta_{1,3})\delta_{2,2} + \frac{\alpha_3(1-\alpha_5)}{e_2}(\beta+\beta^2\eta_{1,3})\eta_{2,2}$$

$$\frac{c_1}{e_2} = \frac{\alpha_1 R^{-1}}{\alpha_3\alpha_5(\beta+\beta^2\delta_{1,3})\delta_{2,2} + \alpha_3(1-\alpha_5)(\beta+\beta^2\eta_{1,3})\eta_{2,2}}$$

$$\frac{\partial U}{\partial e_3} = -\frac{\alpha_1 R^{-2}}{c_1} + \frac{\alpha_3\alpha_5}{e_3}\beta^2\delta_{2,3} + \frac{\alpha_3(1-\alpha_5)}{e_3}\beta^2\eta_{2,3}$$

$$\frac{c_1}{e_3} = \frac{\alpha_1 R^{-2}}{\alpha_3\alpha_5\beta^2\delta_{2,3} + \alpha_3(1-\alpha_5)\beta^2\eta_{2,3}}$$

$$\frac{\partial U}{\partial a_1} = -\frac{\alpha_2}{\ell_1} + \frac{\alpha_3\alpha_5}{a_1}(1+\beta\delta_{1,2}+\beta^2\delta_{1,3}\delta_{1,2})\delta_{3,1} + \frac{\alpha_3(1-\alpha_5)}{a_1}(1+\beta\eta_{1,2}+\beta^2\eta_{1,3}\eta_{1,2})\eta_{3,1}$$

$$\frac{\ell_1}{a_1} = \frac{\alpha_2}{\alpha_3\alpha_5(1+\beta\delta_{1,2}+\beta^2\delta_{1,3}\delta_{1,2})\delta_{3,1} + \alpha_3(1-\alpha_5)(1+\beta\eta_{1,2}+\beta^2\eta_{1,3}\eta_{1,2})\eta_{3,1}}$$

$$\frac{\partial U}{\partial a_2} = -\frac{\alpha_2\beta}{\ell_2} + \frac{\alpha_3\alpha_5}{a_2}(\beta+\beta^2\delta_{1,3})\delta_{3,2} + \frac{\alpha_3(1-\alpha_5)}{a_2}(\beta+\beta^2\eta_{1,3})\eta_{3,2}$$

$$\frac{\ell_2}{a_2} = \frac{\alpha_2}{\alpha_3\alpha_5(1+\beta\delta_{1,3})\delta_{3,2} + \alpha_3(1-\alpha_5)(1+\beta\eta_{1,3})\eta_{3,2}}$$

$$\frac{\partial U}{\partial a_3} = -\frac{\alpha_2\beta^2}{\ell_3} + \frac{\alpha_3\alpha_5}{a_3}\beta^2\delta_{3,3} + \frac{\alpha_3(1-\alpha_5)}{a_3}\beta^2\eta_{3,3}$$

$$\frac{\ell_3}{a_3} = \frac{\alpha_2}{\alpha_3\alpha_5\delta_{3,3} + \alpha_3(1-\alpha_5)\eta_{3,3}}$$

B.3 Monte Carlo Experiments

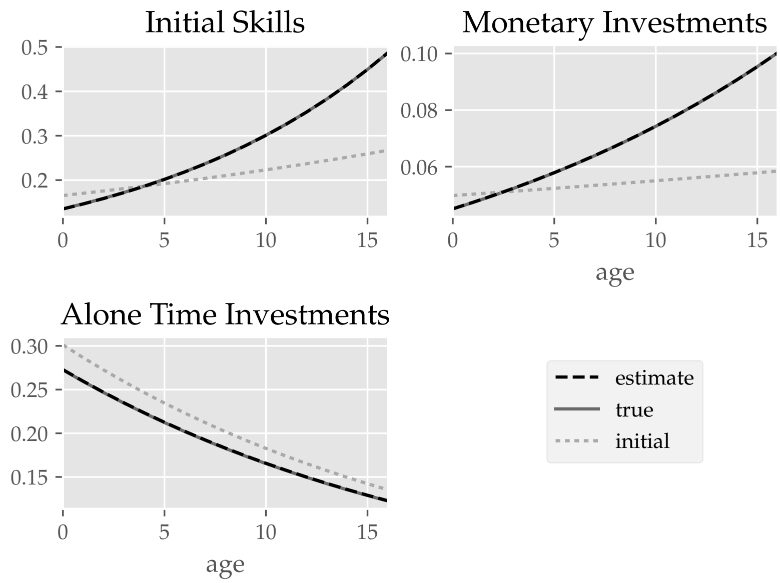
This section uses Monte Carlo exercises to examine the properties of the estimator. I construct the true data generating process in two steps. First, I get the initial conditions from the data. Then, I solve for the optimal choices given the initial conditions, the preferences, and the technologies. In order to make observables comparable with the actual data, I discard all information that are not observed in the data. In particular, I only use the initial conditions at time t, time investments at time t, cognitive and non-cognitive skills at time t and $t+5$, and household lifetime income. Table B.2 displays the initial conditions. Figures B.1 and B.2 plot that the results for one-child households. Figures B.3, B.4, and B.5 plot the results for two-child households.

Table B.2: Initial Conditions

	Obs	Mean	Std. Dev.	Min	Max
A: One-Child Households					
age, first child	215	8.30	2.57	3.00	12.00
cognitive skills, first child	215	30.19	15.09	1.00	55.00
non-cognitive skills, first child	215	2.57	0.41	1.00	3.00
lifetime income (10,000 USD)	215	6.48	3.21	0.50	21.36
B: Two-Child Households					
age, first child	186	9.55	1.98	4.00	12.00
spacing	186	2.34	0.92	1.00	4.00
cognitive skills, first child	186	38.04	11.57	5.00	55.00
non-cognitive skills, first child	186	2.69	0.39	1.25	3.00
cognitive skills, second child	186	24.98	13.76	1.00	57.00
non-cognitive skills, second child	186	2.61	0.41	1.00	3.00
lifetime income (10,000 USD)	186	4.79	2.31	1.02	12.34

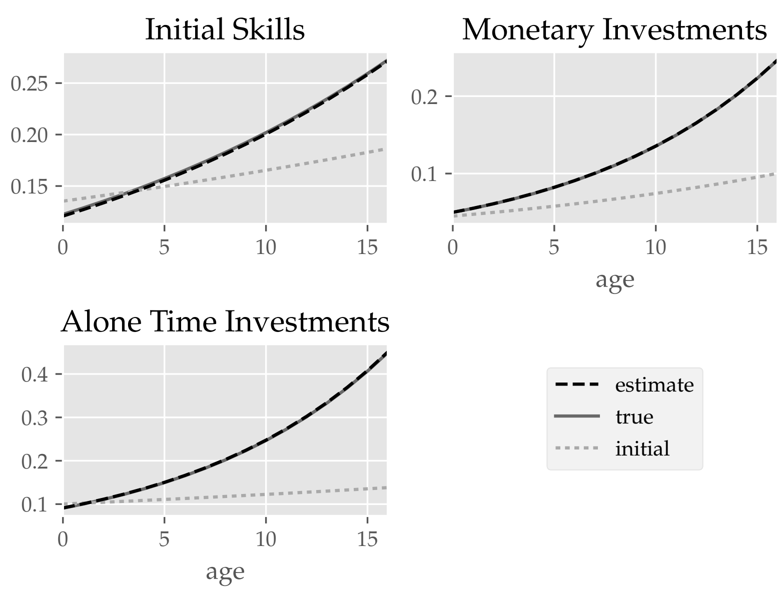
Source: the PSID-CDS.

Figure B.1: One-Child Model; Technology of Cognitive Skills



Source: Monte Carlo results.

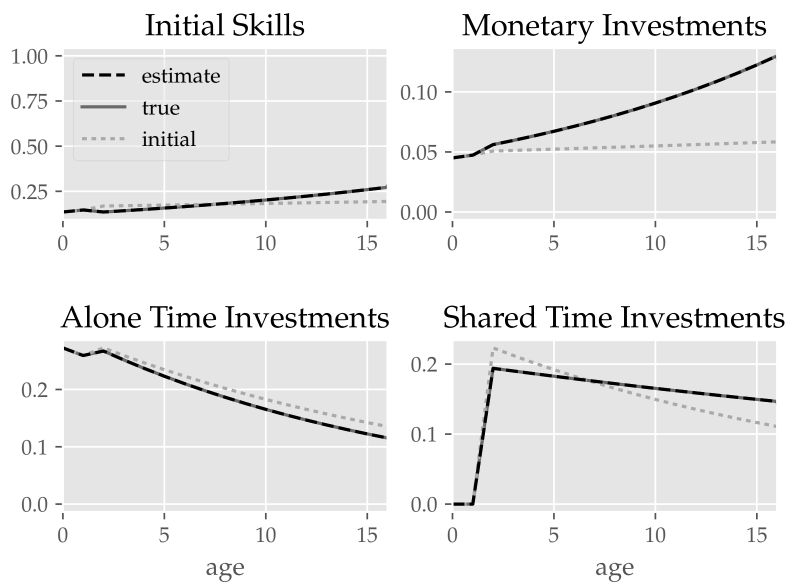
Figure B.2: One-Child Model; Technology of Non-Cognitive Skills



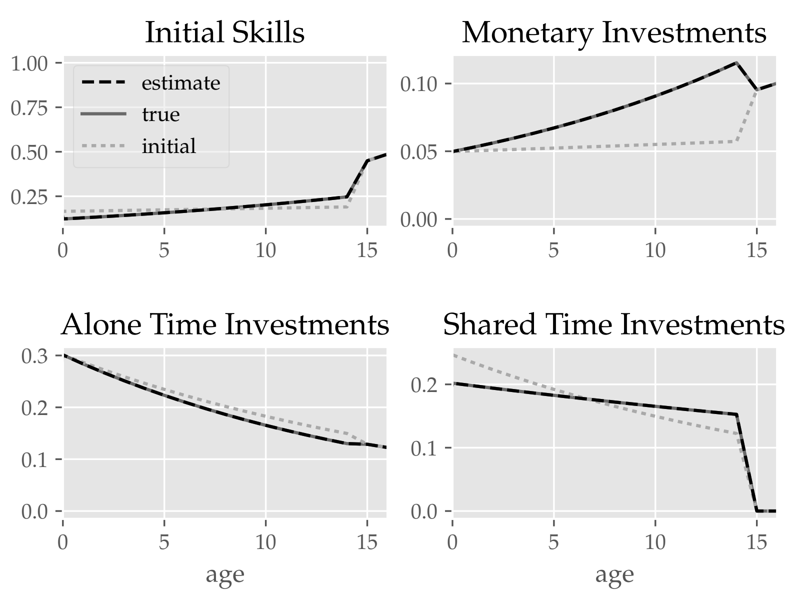
Source: Monte Carlo results.

Figure B.3: Two-Child Model; Technology of Cognitive Skills

(a) First Born



(b) Second Born

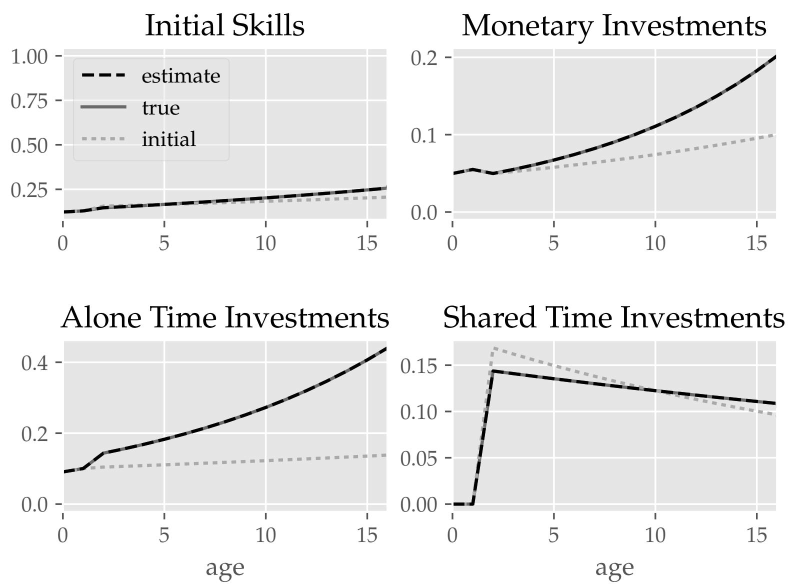


Source: Monte Carlo results.

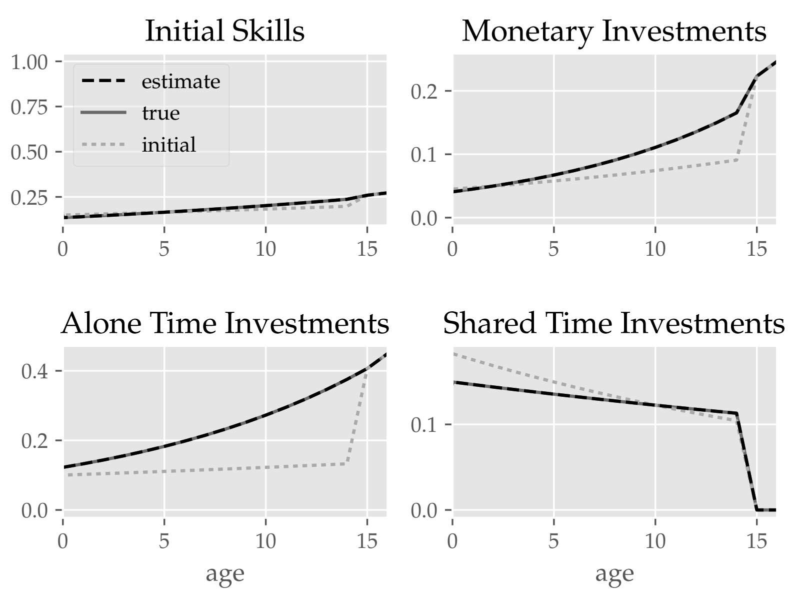
Note: The figure pertains to households with birth spacing of one year.

Figure B.4: Two-Child Model; Technology of Non-Cognitive Skills

(a) First Born



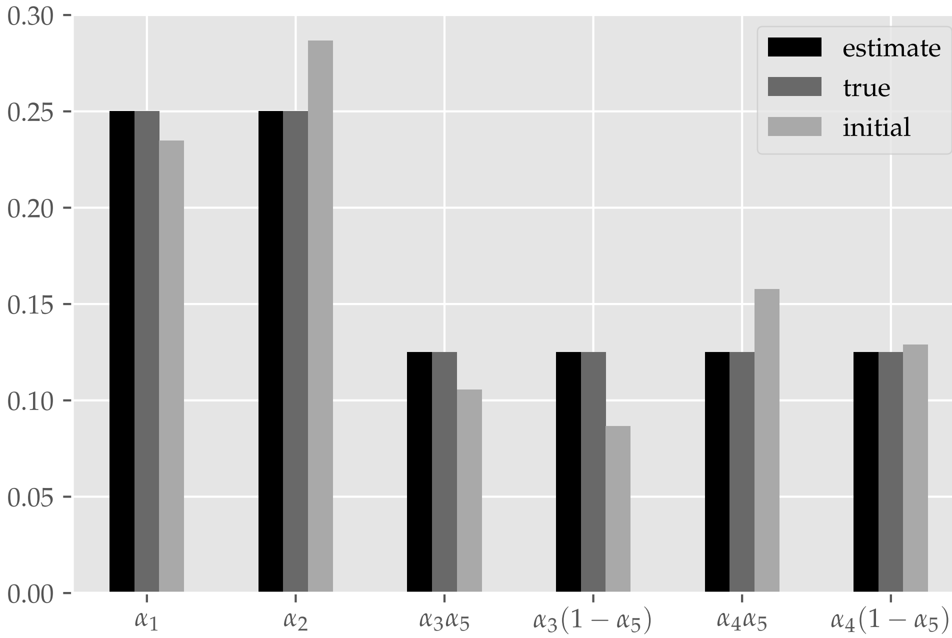
(b) Second Born



Source: Monte Carlo results.

Note: The figure pertains to households with birth spacing of one year.

Figure B.5: Preference Parameters



Source: Monte Carlo results.

Notes: α_1 is the weight on consumption, α_2 is the weight on leisure, α_3 is the weight on first child, α_4 is the weight on second child, α_5 is the weight on cognitive skills, and $1-\alpha_5$ is the weight on non-cognitive skills.

APPENDIX C

SCHOOLS, PARENTS, AND CHILD DEVELOPMENT

C.1 Monte Carlo Experiments

This section uses Monte Carlo exercises to examine the properties of the estimator. The true data generating process is assumed to be:

$$\ln\theta_{i,t+1} = \gamma_0 + \gamma_1 \ \ln\theta_{i,t} + \gamma_2 \ \ln S_{i,t} + \gamma_3 \ \ln\theta_{i,t}\ln S_{i,t} + \eta_{i,t} \tag{C.1}$$
$$\ln S_{i,t} \sim N(0,1) \tag{C.2}$$
$$\ln\theta_{i,t}|\ln S_{i,t} \sim N(0.6\ln S_{i,t}, 1) \tag{C.3}$$
$$\eta_{i,t} \sim N(0,0.3) \tag{C.4}$$

where $\theta_{i,t}$ is the skill of child i at time t, $\ln S_{i,t}$ is the investment from school, and $\epsilon_{i,t}$ is the error term. We assume that children are clustered by classrooms and we observe multiple children per classroom. Further, we assume that investment from school is the same for all children in the same classroom. We assume that $\ln\theta_{i,t}$ is not observed, but we observe two noisy measures of it:

$$\ln y_{i,t} = 0.95\ln\theta_{i,t} + 0.3\epsilon_{i,t} \tag{C.5}$$
$$\ln z_{i,t} = 0.95\ln\theta_{i,t} + 0.3\varepsilon_{i,t} \tag{C.6}$$
$$\epsilon_{i,t} \sim N(0,1) \tag{C.7}$$
$$\varepsilon_{i,t} \sim N(0,1) \tag{C.8}$$

where we have set the signal-to-noise ratio to 90%.

We fix some values for the number of classrooms, N, number of children per classroom, n, and the true parameters; then, we generate 100 simulated dataset of N classrooms with n children from the data generating process described above. Tables C.1 reports the mean estimates and their 95% confidence intervals.

Table C.1: Monte Carlo Results

	γ_0	γ_1	γ_2	γ_3
A: Data Generating Process				
	1.000	0.300	2.000	0.150
B: Model Estimates				
$N = 100$, $n = 10$	1.044	0.376	2.093	0.119
	[1.021,1.069]	[0.335,0.418]	[2.038,2.165]	[0.054,0.185]
$N = 100$, $n = 20$	1.030	0.339	2.092	0.131
	[1.009,1.051]	[0.305,0.372]	[2.053,2.128]	[0.092,0.172]
$N = 100$, $n = 50$	1.016	0.314	2.095	0.143
	[1.002,1.031]	[0.289,0.342]	[2.071,2.123]	[0.114,0.170]

Source: Monte Carlo results.
Notes: N and n denote number of classrooms and number of children per classroom, respectively. 95% confidence intervals are in parentheses.

C.2 Sample Selection Algorithm

We select classrooms that have at least five children with two mathematics, two reading, and two non-cognitive scores in the beginning and at the end of the kindergarten year, and two measures of home investments during the kindergarten year. We begin our selection by dropping all non-responses in Fall (round 1) and Spring (round 2) of the kindergarten year, and children who have changed school during the year.

1. Drop all children who have a non-response status in round 1 or 2; i.e. children with missing School ID.

2. Drop all children who have changed school between round 1 and 2; i.e. children whose School ID in round 1 is different from School ID in round 2.

Next, we select children that have all the required information.

3. Drop all children who do not have two mathematics, two reading, and two non-cognitive scores in rounds 1 and 2, and two measures of homes investments in round 1.

We define a classroom to be the set of all children who have the same circumstances in rounds 1 and 2. In the data, a classroom is a pair of Teacher ID and Class Time (AM/PM/All Day). Teacher ID and/or Class Time may contain missing information in a round, which means a teacher non-response. We allow for a missing Teacher ID/Class Time to be a type in order to maximize the sample size. Finally, we select the classrooms that have at least five observations:

4. Group Teacher ID 1, Class Time 1, Teacher ID 2, and Class Time 2; Then, drop all classrooms that have fewer than five children.

Table C.2: Tabulation of Children by Algorithm Steps

Algorithm Step	Children
did not respond in round 1, 2, or both	760
changed school between rounds 1 and 2	714
do not have full information	5,631
are in classrooms with fewer than 5 observations	4,830
remainder	9,474
total	21,409

Source: the ECLS.

C.3 Measures of Home Investments

This section demonstrates that measures of investment from home are highly correlated with the end-of-kindergarten skills, but once we control for children's skills at the beginning of the kindergarten year, their correlation goes to zero.

Table C.3: Correlation between Investments from Home and Future Outcomes

	$\log M_1$	$\log R_1$	$\log N_1$
A: No Controls			
$\log H_0$	0.131***	0.222***	0.215***
	[0.097,0.167]	[0.186,0.251]	[0.166,0.252]
B: Controlling for Initial Skills			
$\log M_0$	0.781***	0.218***	0.166***
	[0.753,0.808]	[0.191,0.244]	[0.127,0.204]
$\log R_0$	0.071***	0.583***	0.007
	[0.047,0.093]	[0.564,0.603]	[-0.024,0.041]
$\log N_0$	0.059***	0.054***	0.716***
	[0.036,0.076]	[0.028,0.080]	[0.680,0.750]
$\log H_0$	-0.022	0.029***	0.048***
	[-0.042,0.002]	[0.010,0.055]	[0.018,0.073]
N-Children	9474	9474	9474
N-Classroom	1211	1211	1211

Source: Model estimates using a sample of ECLS data.

Notes: M, R, N, and H represent mathematics skills, reading skills, non-cognitive skills, and home investments, respectively. Subscript 0 and 1 represent the beginning and the end of the kindergarten year. All models control for age, age squared, mother's years of education, and the number of days between the two rounds of the data collection. Bootstrapped Classroom-clustered 95% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

C.4 Models with Complementarities

This section provides the estimates for the baseline model including parental investments and the value-added model with complementarities.

Table C.4: Skill Formation Estimates w. Parental Investments and Comp.

	$\log M_1$	$\log R_1$	$\log N_1$
$\log M_0$	0.750***	0.215***	0.140***
	[0.723,0.777]	[0.192,0.241]	[0.106,0.175]
$\log R_0$	0.065***	0.541***	0.022
	[0.038,0.088]	[0.515,0.566]	[-0.009,0.052]
$\log N_0$	0.083***	0.073***	0.762***
	[0.062,0.103]	[0.045,0.098]	[0.730,0.796]
$\log H_0$	-0.043***	0.041***	0.046***
	[-0.067,-0.013]	[0.021,0.066]	[0.011,0.071]
$\log S_0$	0.324***	0.389***	0.520***
	[0.307,0.339]	[0.372,0.414]	[0.496,0.556]
$\log M_0 \times \log S_0$	-0.039**	-0.022	-0.036**
	[-0.083,-0.003]	[-0.042,0.005]	[-0.080,-0.007]
$\log R_0 \times \log S_0$	-0.036***	-0.075***	-0.012
	[-0.069,-0.009]	[-0.096,-0.058]	[-0.043,0.023]
$\log N_0 \times \log S_0$	0.019	0.017	-0.016
	[-0.006,0.044]	[-0.008,0.047]	[-0.049,0.021]
$\log H_0 \times \log S_0$	-0.027	0.008	-0.043**
	[-0.051,0.005]	[-0.014,0.033]	[-0.076,-0.006]
N-Children	9474	9474	9474
N-Classroom	1211	1211	1211

Source: Model estimates using a sample of ECLS data.

Notes: M, R, N, H, and S represent mathematics skills, reading skills, non-cognitive skills, home investments, and school investments, respectively. Subscript 0 and 1 represent the beginning and the end of the kindergarten year. All models control for age, age squared, mother's years of education, and the number of days between the two rounds of the data collection. Bootstrapped classroom-clustered 95% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table C.5: VA Skill Formation Estimates w. Complementarities

	$\log M_1$	$\log R_1$
$\log M_0$	0.792***	0.252***
	[0.768,0.817]	[0.227,0.280]
$\log R_0$	0.080***	0.567***
	[0.052,0.103]	[0.542,0.589]
$\log S_0$	0.316***	0.378***
	[0.299,0.333]	[0.365,0.397]
$\log M_0 \times \log S_0$	-0.025	-0.013
	[-0.063,0.009]	[-0.034,0.015]
$\log R_0 \times \log S_0$	-0.038**	-0.070***
	[-0.068,-0.006]	[-0.088,-0.053]
N-Children	9474	9474
N-Classroom	1211	1211

Source: Model estimates using a sample of ECLS data.

Notes: M, R, and S represent mathematics skills, reading skills, and school investments, respectively. Subscript 0 and 1 represent the beginning and the end of the kindergarten year. All models control for child's age, age squared, and the number of days between the two rounds of the data collection. Bootstrapped classroom-clustered 95% confidence intervals are in brackets. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

C.5 Details of the Elasticities

This section presents the descriptive statistics for the implied elasticities of skills with respect to different inputs by the most general model.

Table C.6: Implied Elasticities

	Mean	Std. Dev.	10th-pct	90th-pct
Elasticity of M_1 w.r.t.				
M_0	0.750	0.039	0.700	0.799
R_0	0.065	0.036	0.019	0.110
N_0	0.083	0.019	0.059	0.107
H_0	-0.043	0.027	-0.077	-0.008
S_0	0.325	0.070	0.235	0.414
Elasticity of R_1 w.r.t.				
M_0	0.215	0.022	0.187	0.244
R_0	0.541	0.075	0.445	0.638
N_0	0.073	0.017	0.052	0.095
H_0	0.041	0.008	0.031	0.052
S_0	0.389	0.084	0.281	0.497
Elasticity of N_1 w.r.t.				
M_0	0.140	0.036	0.094	0.185
R_0	0.022	0.012	0.006	0.037
N_0	0.762	0.016	0.741	0.783
H_0	0.046	0.043	-0.010	0.102
S_0	0.520	0.076	0.423	0.617

Source: Model estimates using a sample of ECLS data.

Notes: M, R, N, H, and S represent mathematics skills, reading skills, non-cognitive skills, home investments, and school investments, respectively. Subscript 0 and 1 represent the beginning and the end of the kindergarten year. Elasticities pertain to model with parental investments and complementarities (see table C.4).

C.6 Effect of Investments in Kindergarten on Future Outcomes

This section examines the effects of the school investments during the kindergarten year on children's future outcomes. The results demonstrates that our estimated investments in kindergarten predict children's future outcomes.

Table C.7: Effect of Investments in Kindergarten on Future Reading Outcomes

	Spring G1	Spring G3	Spring G5	Spring G8
M_0	0.253***	0.337***	0.353***	0.354***
	(0.012)	(0.014)	(0.016)	(0.017)
R_0	0.261***	0.194***	0.225***	0.124***
	(0.012)	(0.013)	(0.015)	(0.017)
N_0	0.196***	0.135***	0.103***	0.121***
	(0.012)	(0.013)	(0.013)	(0.016)
$\log S_{M,0}$	0.091***	0.051***	0.058***	0.046**
	(0.013)	(0.016)	(0.016)	(0.018)
$\log S_{R,0}$	0.143***	0.070***	0.070***	0.021
	(0.012)	(0.014)	(0.015)	(0.017)
$\log S_{N,0}$	0.038***	0.048***	0.031**	0.046***
	(0.012)	(0.011)	(0.013)	(0.015)
N	8021	6897	5537	4678

Source: Model estimates using a sample of ECLS data.

Notes: M_0, R_0, and N_0 represent mathematics, reading, and non-cognitive scores at kindergarten entry, respectively. Mathematics and reading scores are the raw scores. Non-cognitive score is the Approaches to Learning evaluated by the teacher. $S_{M,0}$, $S_{R,0}$, and $S_{N,0}$ represent school investments in mathematics, reading, and non-cognitive skills, respectively. School investments corresponds to the baseline model with complementarities (see Table 3.7). Scores are rescaled to have standard deviation of 1. Classroom-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.

Table C.8: Effect of Investments in Kindergarten on Future Non-Cognitive Outcomes

	Spring G1	Spring G3	Spring G5
M_0	0.171***	0.177***	0.139***
	(0.015)	(0.016)	(0.017)
R_0	0.096***	0.057***	0.028
	(0.014)	(0.017)	(0.017)
N_0	0.376***	0.351***	0.327***
	(0.013)	(0.015)	(0.017)
$\log S_{M,0}$	0.020	0.026	0.012
	(0.015)	(0.016)	(0.019)
$\log S_{R,0}$	0.025	0.049***	0.029
	(0.016)	(0.017)	(0.019)
$\log S_{N,0}$	0.103***	0.060***	0.083***
	(0.013)	(0.016)	(0.017)
N	7473	6023	5342

Source: Model estimates using a sample of ECLS data.

Notes: M_0, R_0, and N_0 represent mathematics, reading, and non-cognitive scores at kindergarten entry, respectively. Mathematics and reading scores are the raw scores. Non-cognitive score is the Approaches to Learning evaluated by the teacher. $S_{M,0}$, $S_{R,0}$, and $S_{N,0}$ represent school investments in mathematics, reading, and non-cognitive skills, respectively. School investments corresponds to the baseline model with complementarities (see Table 3.7). Scores are rescaled to have standard deviation of 1. Classroom-clustered standard errors are in parentheses. ***, **, and * denote significantly different from zero at 1%, 5%, and 10% levels, respectively.