Essays on Children's Skill Formation

by

Francesco Agostinelli

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

Approved April 2018 by the
Graduate Supervisory Committee:

Matthew Wiswall, Co-Chair
Daniel Silverman, Co-Chair
Esteban Matias Aucejo
Kevin Reffett
Gregory Francisco Veramendi

ARIZONA STATE UNIVERSITY

May 2018

ABSTRACT

The dissertation is composed by three chapters. In Chapter 2 (coauthored with Matthew Wiswall) I develop new results for the identification and estimation of the technology of children's skill formation when children's skills are unobserved. In Chapter 3 I shed light on the importance of dynamic equilibrium interdependencies between children's social interactions and parental investments decisions in explaining developmental differences between different social environments. In Chapter 4 (coauthored with Giuseppe Sorrenti) I study the effect of family income and maternal hours worked on both cognitive and behavioral child development.

TABLE OF CONTENTS

CHAPTER Page

Page

Page

LIST OF TABLES

Table Page

Table Page

Table | Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Chapter 1

INTRODUCTION

The dissertation is divided in three chapters. In Chapter 2, coauthored with Matthew Wiswall, we develop a new estimator for the process of children's skill formation. The wide dispersion of measured human capital in children and its strong correlation 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, 2014b). However, the empirical challenges in estimating the skill formation process, principally the technology of child development, is hampered by the likely imperfect measures of children's skills we have available. 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 and provide widely differing levels of informativeness about the underlying latent skills of the child. [1] 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.

In this paper, we develop a new method to estimate the skill formation process in children when skills are not observed directly but instead measured with error. Rather than assuming skills are measured perfectly by a particular measure, we accommodate the variety of skills measures used in practice and allow latent skills to be

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

measured with error using a system of arbitrarily located and scaled skill measures. In our framework, we treat the parameters of the measurement model as "nuisance" parameters and use transformations of moments of the measurement data to eliminate them, analogous to the transformations used to eliminate fixed effects with panel data. We show non-parametric identification of the primitive parameters of the production technology, without assuming any particular values for the measurement process parameters or "re-normalizing" latent skills each period.

The heart of our identification analysis is a characterization of the classes of production technologies which can be identified given different assumptions about the measurement process. We introduce the concept of production technologies that have a *known* location and scale, technologies which are implicitly restricted so that the location and scale is already known. These known location and scale (KLS) technologies include the CES production technologies considered in a number of previous papers (Cunha and Heckman, 2007; Cunha *et al.*, 2010; Cunha and Heckman, 2008; Pavan, 2015). Starting with this class of technologies, we show that standard measurement error assumptions non-parametrically identify the primitive production function parameters, up to a normalization on the initial conditions only. Importantly, identification is obtained without restrictions on the later skill measures as imposed in some previous papers, which can bias the production function estimates (see Agostinelli and Wiswall (2016b) for a discussion).

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 estimated already has a known location and scale, as is the case for the technology estimated by Cunha *et al.* (2010). We show that non-parametric 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)).

In an important extension of our baseline results, we develop additional restrictions on the measurement process which are sufficient for identification of more general production technologies, including those exhibiting Hicks neutral total factor productivity (TFP) dynamics and non-constant returns to scale. Using standard assumptions, these more general technologies cannot be identified because the location and scale of the technology cannot be separately identified from the location and scale of the measures. These more general aspects of the skill development formation process are nonetheless potentially important as restricting the technology can reduce the permissible skill dynamics and productivity of investments, substantially changing our inferences about the child development process and our evaluation of policy. Our paper provides the first identification results for these more general models. Our analysis makes clear the key identification tradeoff researchers face: identification of restricted KLS technologies is possible with standard measurement assumptions, but identification of more general technologies requires stronger assumptions. We evaluate the empirical relevance of these additional assumptions, and provide guidance to researchers to evaluate whether the measures available to them satisfy these assumptions.

In the second part of our paper, we estimate a flexible parametric version of our model using data from the US National Longitudinal Survey of Youth (NLSY). We examine the development of cognitive skills in children from age 5 to age 14, and estimate a model of cognitive skill development allowing for complementarties between parental investment and children's skills; endogenous parental investment responding to the stock of children's skills, maternal skills, and family income; Hicks neutral dynamics in TFP; non-constant returns to scale; and unobserved shocks to the investment process and skill production. Following Cunha *et al.* (2010), our empirical framework treats not only the child's cognitive skills as measured with error, but investment and maternal skills as well.

Constructively derived from our identification analysis, we form a method of moments estimator. Our estimator is not only relatively simple and tractable but also robust because it does not impose parametric distributional assumptions on the distribution of latent skills and measurement errors, as is commonly imposed in previous estimators. We jointly estimate the technology of skill formation, the process of parental investments in children, and the adult distribution of completed schooling and earnings, allowing the production technology and investment process to freely vary as the child ages. Our estimates of high TFP and increasing returns to scale at early ages indicate that investments are particularly productive early in the development period. We also find that the marginal productivity of early investments is substantially higher for children with *lower* existing skills, suggesting the optimal targeting of interventions to disadvantaged children.

Our estimates of the dynamic process of investment and skill development allow us to estimate the heterogeneous treatment effects of some simple policy interventions.

We show that even a modest transfer of family income to families at age 5 would substantially increase children's skills and completed schooling, with the effects larger for low income families. When we compare these estimates to those using models which restrict the technology or ignore measurement error, we estimate policy effects which are substantially smaller, indicating that the generalities we allow are important quantitatively to answering key policy questions.

The paper is organized as follows. In the next two sections, we develop the model of skill development and the measurement process. The next sections analyze the identification of this model, first under weak assumptions about the measurement process, and then under stronger assumptions about measurement which allows the identification of more general technology specifications, including those with TFP dynamics and non-constant returns to scale. The remainder of the paper develops our estimator and discusses our estimation results.

In Chapter 3, I study the effects of social interactions on the dynamics of children's skills. This paper analyzes the effect of social interactions on skill formation in children. In particular, I build and estimate a model of child development, where children grow up in different *environments*, which are defined by: peers' composition, neighborhood quality and school quality. The dynamics of skills is governed by a technology of skill formation, which depends upon parental investments, the current child's skills and the environment-specific inputs. In this framework, I shed light on the importance of the dynamic effects of children's endogenous social interactions and the parental investment decisions in explaining developmental differences between different environments. A growing consensus in the literature emphasizes the importance of neighborhoods in shaping children's opportunities later in life (Chetty and Hendren, 2016a,b; Chetty *et al.*, 2016a,b). However, despite extensive research, the mechanisms behind these results remain unexplained. This paper reconciles the previous findings of childhood exposure to neighborhood with the role of children's social interactions in child development.

This project advances the current literature of child development by building and estimating a dynamic equilibrium model of children's skill formation with two innovative empirically grounded features. First, within different environments, children endogenously select their peer groups based on their preferences for their peers' characteristics. Social interactions can exhibit the tendency of children to become friends with others who share similar characteristics: a phenomenon called homophily bias. Second, parental investments respond to changes in peer groups. Decisions regarding parental investments depend upon a child's current peers, as well as on expectations about future peer groups. Equilibrium effects arise from the socially determined aspects of parental investments. In this framework, parental investments not only directly affect a child's skills, but also affect the development of the child's peers through social interactions. Consequently, the individual return on investing in children is affected by the equilibrium parental engagement within each environment.

Skills are formed dynamically through a technology of skill formation, which defines the complementarities between parental investments and the other inputs of child development in producing a child's skills: the current endowment of skills, the skills of peers, the school quality and the neighborhood quality. In this framework, there are two main channels through which peers affect parental behavior. First, contemporaneous changes in current peers and parental investments are related to the *static* complementarity between the two inputs. Second, permanent changes in peer composition affect parental behavior through the *dynamic* complementarity in skill formation. In other words, a permanent change in peer composition affect the return of parental investments through the dynamic aspect of skill formation.

The model is estimated using data on U.S. adolescents from the National Longitudinal Study of Adolescent Health (Add Health). Add Health provides information about friendships within each school, which is key for analyzing the formation of peer groups. Moreover, information about child achievements and parental investments are available.

The identification of the model comes with two main challenges: (i) unobserved heterogeneity in how peer groups are endogenously formed; and (ii) children's skills and parental investments are unobserved. Ignoring these issues by using correlational relationships would cause the model's estimates and subsequent quantitative analysis to be biased.

The first challenge presents itself from the fact that peer groups may be formed based on additional unobserved heterogeneity, which can cause correlation between peer groups' realization and the residual unexplained variation in skill formation. To address this concern, I implement a standard instrumental variable (IV) approach in the literature. This identification strategy exploits random variations in cohort composition within school / across cohorts. The idea behind this identification strategy is simple: random changes in cohort compositions affect the opportunities for friendships between children. These shifts in the formation of peer groups affect the return of parental investments and the subsequent parental decisions. [2]

In addressing the second challenge, Cunha *et al.* (2010) illustrate that even the classical measurement error in measuring a child's skills can cause important biases in estimating the technology of children's skill formation. Following the approach in Cunha *et al.* (2010) and Agostinelli and Wiswall (2016a), I implement a dynamic

[2]For previous use of similar source of identifying variation, see Hoxby (2000); Hanushek *et al.* (2003); Ammermueller and Pischke (2009); Lavy and Schlosser (2011); Lavy *et al.* (2012); Bifulco *et al.* (2011); Burke and Sass (2013); Card and Giuliano (2016); Carrell *et al.* (2016); Olivetti *et al.* (2016); Patacchini and Zenou (2016)

latent factor model, which allows me to identify the joint distribution of latent skills and investments by exploiting multiple measurements in the data.

I estimate the model via simulated method of moments (SMM). I find that parental investments and peers are substitute inputs in producing children's skills. At the same time, I find a strong dynamic complementarity between parental investments and future expected peers. As a result of these two findings, a permanent change in peer composition has two opposing effects on parental investments. On one hand, "better" peers generate contemporaneous substitution effects in investment decisions due to the high substitutability in the production function. On the other hand, higher expected future skills for peers produce an "income" effect through the dynamic complementarity of skill formation. Parents have the incentive to invest more in their children because a higher-skilled child benefits more from higher-skilled peers in the future.

Furthermore, my estimates suggest that the formation of peer groups displays an extensive degree of homophily bias. I show evidence of homophily bias with respect to a child's race and level of latent skills. A child who is in the lower quartile of the skill distribution and belongs to a minority group is four times more likely to befriend a same-race child than a different-race child. In addition, the same child is two times more likely to befriend a same-skill and same-race child than a same-race child in the upper quartile of skill distribution.

I first use the estimated model to analyze the extent to which growing up in different environments accounts for the variation in children's outcomes. I find sizable effects for children moving to better environments. The effects are in proportion to the exposure time. The earlier children are moved, the higher the effect. A child who is moved at age 12 to an environment where children have 1 percentile higher skills at age 16 exhibits, on average, an improvement in her skills rank at age 16 by

0.63 percentiles. The average effect is 0.48 percentiles if the child is moved at age 15. As model validation, I show that my findings track (out-of-sample) the quasi-experimental findings of childhood exposure effects of neighborhoods for the U.S. from Chetty and Hendren (2016a). In addition, my model allows me to decompose these effects. I find that *peers* account for more than half of the exposure effects.

The relative importance of peers for the exposure effects underlines the role of policies that change peers' composition and promote socioeconomic integration in environments, as a way to improve outcomes for disadvantaged children. I find that by moving the most disadvantaged children (in the lower quartile of skill distribution) from a low-income environment to a high-income environment generates important dynamic equilibrium effects, with heterogeneous treatment effects for both the moved and receiving children. I first consider a large-scale policy, i.e. a policy that moves a sizable fraction of disadvantaged children into a higher-income environment (approximately 5% of the population of the receiving cohort). I find that the policy increases the skills of the moved population of 16-year-old children, on average, by approximately 0.40 standard deviations. On average, I do not find any adverse effect for receiving children. On the contrary, when the fraction of moved population increases to 30%, I find that the policy generates *winners* and *losers*. First, I find that the policy increases the skills of the moved population of 16-year-old children on average by 0.22 standard deviations. In contrast, there is an adverse effect for receiving children, with the skills of 16-year-old children decreasing, on average, by 0.15 standard deviations. Additionally, I find that children who remained in the sending environment benefit from the outflow of the most disadvantaged companions, with an average increase in skills at age 16 of 0.17 standard deviations.

I find that large-scale changes in peers' composition generate important equilibrium feedback effects, and as a result amplify the policy effects. Ignoring equilibrium effects would lead to large biases in counterfactual policy predictions for children's final skills. In the case of the larger policy, I find that the policy predictions for the children's skills in the receiving environment would be approximately seven times smaller. Part of the bias is due to the dynamic-equilibrium feedback effects on parental investments. In fact, in the absence of dynamic-equilibrium feedback effects, the *static* complementarity between parents and peers dominates the dynamic effects of the policy.

I find that policy effects for receiving and remaining children reduce in magnitude as the fraction of moved children decreases. An increase of inflow of the most disadvantaged children from the low-income environment to the high-income environment increases the probability of the receiving children becoming friends with the new companions. For the same reason, an increase of the outflow of the moved population benefits children who remain in the sending environment. For children who were moved, the opposite is true. The higher the outflow of disadvantaged companions, the higher the chances that the moved children remain friends with each other in the new environment.

My structural model allows me to analyze the distributional policy effects. I find that large-scale changes in peers' composition exhibit heterogeneous treatment effects as a result of the endogenous formation of new peer groups. Children with lower skills (in the first quartile of the skills distribution in each subpopulation): (i) benefit the most in leaving disadvantaged social environments; (ii) benefit the most amongst the children who remained in the sending environment; (iii) are the ones who are more adversely affected in receiving the new peers. Furthermore, I find stronger policy effects for minorities, with detrimental effects in black and Hispanic children living in the receiving environment. This is explained by the fact that most of the moved children are minorities, and as a result, the minority children from the receiving environment are more likely to interact with the new companions because of the race effects in the endogenous formation of peer groups. In line with this result, previous empirical studies pointed out that peer effects seem to be stronger *intra*-race and for minorities (see Hoxby, 2000; Angrist and Lang, 2004; Imberman *et al.*, 2012)

The paper will be presented as follows. In Section 3.2, I discuss the related literature. In Section 3.3, I present the data used for the empirical work and preliminary empirical results. In Section 3.4 and 3.5, I present the model. In Section 3.6, I describe the identification strategy. Section 3.7 contains a discussion of the structural estimation and results. Section 3.8 and Section 3.9 discuss the quantitative analysis and the model validation. Section 3.10 concludes.

In Chapter 4, coauthored with Giuseppe Sorrenti, we study the effect of family income and maternal hours worked on child development. Poverty represents one of the major threats to child development. In 2015, about 15 million children in the United States (21 percent of all children) were living in families with incomes below the federal poverty threshold (National Center for Children in Poverty, 2015). What effect does growing up in a disadvantaged family have on a child's achievements, and how can living conditions be improved to promote child development?

Support programs such as the Earned Income Tax Credit (EITC), the Food Stamp Program, and the Child Tax Credit attempt to reduce family poverty and especially that experienced by children. Many of these programs (e.g. the EITC) provide cash transfers on the condition that the recipient works (conditional cash transfers). Such conditions might shape child development by introducing a trade-off between the *income* effect, due to a surge in family income, and the *substitution* effect, due possibly to parental labor supply responses and a decrease in time parents spend with their child.

The arising trade-off poses an important question: is the change in family income more important than time spent with parents in shaping child development? In this study we answer this question by appraising the contemporaneous effect of changes in family income and maternal labor supply on cognitive and behavioral development of children. We implement an instrumental variable (IV) approach exploiting changes in the EITC benefits over time and shocks in the local labor demand as instruments for family income and maternal labor supply. In this sense, we bridge the gap between the literature dealing with the estimate of the effect of family income on child development and the literature on the effect of maternal labor supply and child-with-parents time. Moreover, we provide important insights on what policies can foster maternal employment and child development contemporaneously.

Family income is an important predictor of a child's success and future opportunities. Figure 4.1 shows the wide dispersion in children's achievements by family income. Both cognitive (Panel A) and behavioral (Panel B) development measures exhibit a steep income gradient, with high-achieving children placed in the top deciles of the after-tax family income distribution. The impact of family income on child development has been widely debated by economists. Previous studies such as Duncan *et al.* (2011), Levy and Duncan (1999), and Blau (1999) have found a positive relation between family economic conditions during childhood and child achievements. More recently, works such as Løken *et al.* (2012) and Dahl and Lochner (2012b) employ instrumental variable techniques to confirm this positive effect in Norway and in the U.S., respectively.

In addition to studies regarding the income effect, a vast body of economic literature associates maternal labor supply during childhood with possible negative effects on child development and future opportunities (Baum, 2003; Ruhm, 2004; Bernal, 2008; Carneiro and Rodriguez, 2009; Bernal and Keane, 2011; Hsin and Felfe, 2014;

Carneiro *et al.*, 2015; Del Bono *et al.*, 2016; Fort *et al.*, 2017). As examples, according to Bernal and Keane (2011) each year of child care (versus maternal time input) before age 6 decreases test scores by 2.1 percent (0.11 standard deviations). Similarly, Carneiro *et al.* (2015) estimate that the probability of dropping out of high school decreases by 2 percent and wages increase by 5 percent at age 30 with the more time mothers spend with their children in the first months of life.

This paper reconciles these strands of the literature. For most families, an increase in income is due to an increase in maternal labor supply. In this case, a surge in monetary resources is associated with a potential decline in the time the mother spends with her offspring. To understand the possible trade-off between family income and maternal labor supply, we build upon the empirical model in Dahl and Lochner (2012b) by considering not only the role of family income but also the role of maternal hours worked in shaping child development. [3] The work by Dahl and Lochner (2012b) exploits quasi-experimental variation in the EITC to analyze the causal effect of family income on child achievement. However, the EITC is designed to incentivize individuals (including mothers) to work. [4] Mothers, and especially single mothers, are usually the main target group of these welfare programs and are most responsive to incentives (Meyer, 2002; Blundell and Hoynes, 2004; Blundell *et al.*, 2016). This affects the maternal allocation of time between working and parenting, with potential effects on children's test scores. More precisely, endogenous labor supply responses affect child development through two channels. An increase in maternal hours worked generates an income effect, with additional resources coming from a boost in labor

[3] Dahl and Lochner (2017), after the analysis by Lundstrom (2017), adjust for a coding error in their previous work in the creation of the after-tax total family income. The results of the original and reviewed studies are similar.

[4] Hotz and Scholz (2003) and Nichols and Rothstein (2016) summarize theoretical and empirical findings about the effect of the EITC on maternal labor supply. Blundell *et al.* (2016) analyze the case of the U.K. and find substantial elasticities for women's labor supply (in particular for the group of single mothers).

income. At the same time, changes in maternal hours worked can also generate a substitution effect, with changes in the time that mothers allocate to child care (Heckman and Mosso, 2014a; Del Boca *et al.*, 2014a). Moreover, this paper is related to previous works that consider the effect of time and monetary resources on children by estimating a structural model of household choices and child development (see Del Boca *et al.*, 2014a; Mullins, 2016).

An additional contribution of our study relates to the broad definition used for child development. While many works (see Dahl and Lochner, 2012b; Del Boca *et al.*, 2014a) exclusively focus on test scores for cognitive achievements, we extend the analysis to proxies for child noncognitive development. [5] As stated by Heckman and Rubinstein (2001), standard test scores only capture some of the multiple skills determining individual success and well-being. Moreover, early childhood interventions that boost personal traits such as self-discipline or motivation are usually deemed as extremely effective (Heckman, 2000). Socio-emotional skills are often more predictive of later-life success than cognitive skills. [6]

Our empirical analysis is based on the National Longitudinal Study of Youth 1979 (NLSY79) data set matched with its Children (NLSY79-C) section. This combined data set provides longitudinal information about measures of child development, family income, and hours worked by the mother. At the same time, the longitudinal structure allows us to account for individual unobserved heterogeneity through child fixed effects. Cognitive development is measured through children's achievements on the Peabody Individual Achievement Test (PIAT), a set of tests assessing proficiency

[5] We also explore features related to early childhood development (1–7 years old).

[6] For example, data from the Perry Preschool Program, a high-quality U.S. preschool education program, suggest that increased academic motivation generates 30 percent of the effects on achievement and 40 percent on employment for females. Reduced externalizing behavior decreases lifetime violent crime by 65 percent, lifetime arrests by 40 percent, and unemployment by 20 percent. Visit heckmanequation.org/resource/early-childhood-education-quality-and-access-pay-off/ for a discussion of these results.

in mathematics and reading. To study noncognitive development, we take advantage of the Behavior Problems Index (BPI). This comprehensive index is comprised of several different indicators for behavior such as aggressiveness or hyperactivity that are likely to shape children's future life opportunities.

Given the strong interdependence between maternal labor supply and family income, there is no suitable identifying source of variation that is likely to exclusively affect one variable of interest. Hence, in order to identify the single causal effect of either family income or maternal labor supply on child development, it is necessary to allow for the endogeneity of both inputs. To deal with this challenge, we exploit two instrumental variables. The first instrument is based on the longitudinal changes in monetary benefits of the EITC, one of the largest U.S. federal income support programs. This variation provides us with exogenous changes in family resources to allocate in child development. At the same time, only working people are eligible for EITC benefits, creating incentives for mothers to work. The second instrument is constructed by using longitudinal shocks in the local labor market demand. Shifts in local demand for labor affect equilibrium prices (wages) and, subsequently, the family income resources and the equilibrium labor quantity. [7]

Our instrumental variable analysis suggests different results for cognitive and behavioral development. An additional $1,000 in family income improves cognitive development by 4.4 percent of a standard deviation. [8] The same income change has no effect on child behavioral development. An additional $1,000 improves behavioral

[7]We provide evidence throughout the paper that both identifying sources of variation do not confound other contemporaneous state-specific factors, like state-specific trends in children's achievements or changes in the per-pupil financial resources of schools in different states. Moreover, in the spirit of Goldsmith-Pinkham *et al.* (2017), we assess the validity of our labor demand shock instrument by formally testing for any parallel pre-trends between the instrument and child development. We reject the hypothesis of the existence of any pre-trends.

[8]This result is in line with the findings of Dahl and Lochner (2012b) and Dahl and Lochner (2017).

development by 1.3 percent of a standard deviation, and the result is not statistically significant.

We find that the income effect is counterbalanced by a negative effect of hours worked by the mother on child development. An increase in maternal labor supply of 100 hours per year causes a statistically significant decrease in both child cognitive and behavioral development by approximately -6 percent and -5 percent of a standard deviation. The strong negative impact of the number of hours worked by the mother, both in terms of cognitive test scores and behavioral problems, encourages the debate in a new dimension: how to address concerns about the effect of maternal employment on child development.

We attempt to answer this question in the last part of our study. By using the time diary component of the American Time Use Survey (ATUS), we illustrate the mechanism underlying the negative impact of hours worked by the mother on child development. Similar to Sayer *et al.* (2004), Guryan *et al.* (2008), and Fox *et al.* (2013), we find that working mothers, conditional on income, invest less time in their children. As a consequence, labor market conditions play a role in shaping the effect of labor supply on child development.

We focus on the role of wages and show that, according to our results, an after-tax hourly wage up to $13.50 makes the substitution effect (less maternal time with the child) dominant over the income effect (higher earnings). With higher earnings, families face the option of substituting their decreased time investment with better and more productive alternatives (e.g. nonparental care, additional schooling, youth clubs, music activities, etc.).

We look for possible heterogeneous effects in different subgroups in order to highlight the potential importance of alternative inputs in the child development process. Behavioral development does not display evidence of heterogeneous impacts of income or hours worked by the mother. On the contrary, the negative effect of hours worked by the mother on cognitive development only appears in less educated, low-skilled, or single mothers. More educated and high-skilled mothers are likely to access to better nonparental child care options. Moreover, the differences in the labor supply effect can be reconciled with heterogeneous preferences for child care activities, generating different patterns of time allocation between working, child care, and leisure time (Guryan *et al.*, 2008).

We further investigate these channels by comparing the investment in the child by maternal employment status and family income. The Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID) collects detailed information about a wide set of children's activities and parental investment for a representative sample of U.S. families. Results obtained with this data set highlight some evidence of differential investments as a response to the maternal employment status when low-income families are compared to high-income families.

Policymakers might obtain several suggestions from our results. First, by showing the trade-off between the income and substitution effect in terms of child development, this work speaks to the growing body of literature about the effect of conditional versus unconditional cash transfers. Many income subsidies worldwide base monetary transfers on work requirements. In this context, only looking at the effect of income on child development might lead to biased policy predictions. Our results support the idea that policies aimed at fostering maternal labor supply can be beneficial to child development if integrated with specific consideration about a minimum wage or the taxation of family income. Alternatively, policies that encourage maternal employment in low-income families should also consider how to guarantee alternative sources of child care to support child development.

The remainder of the paper is structured as follows. Section 4.2 introduces the empirical model and the identification strategy. The data used for the analysis are presented in Section 4.3, while the results are described in Section 4.4. Section 4.5 sheds light on the mechanism underlying the main findings of the work. Section 4.6 concludes.

Chapter 2

IDENTIFICATION AND ESTIMATION OF THE TECHNOLOGY OF CHILDREN'S SKILL FORMATION (WITH MATTHEW WISWALL)

2.1 Introduction

The wide dispersion of measured human capital in children and its strong correlation 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, 2014a). However, the empirical challenges in estimating the skill formation process, principally the technology of child development, is hampered by the likely imperfect measures of children's skills we have available. 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 and provide widely differing levels of informativeness about the underlying latent skills of the child. [1] 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.

In this paper, we develop a new method to estimate the skill formation process in children when skills are not observed directly but instead measured with error. Rather than assuming skills are measured perfectly by a particular measure, we accommodate the variety of skills measures used in practice and allow latent skills to be

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

measured with error using a system of arbitrarily located and scaled skill measures. In our framework, we treat the parameters of the measurement model as "nuisance" parameters and use transformations of moments of the measurement data to eliminate them, analogous to the transformations used to eliminate fixed effects with panel data. We show non-parametric identification of the primitive parameters of the production technology, without assuming any particular values for the measurement process parameters or "re-normalizing" latent skills each period.

The heart of our identification analysis is a characterization of the classes of production technologies which can be identified given different assumptions about the measurement process. We introduce the concept of production technologies that have a *known* location and scale, technologies which are implicitly restricted so that the location and scale is already known. These known location and scale (KLS) technologies include the CES production technologies considered in a number of previous papers (Cunha and Heckman, 2007; Cunha *et al.*, 2010; Cunha and Heckman, 2008; Pavan, 2015). Starting with this class of technologies, we show that standard measurement error assumptions non-parametrically identify the primitive production function parameters, up to a normalization on the initial conditions only. Importantly, identification is obtained without restrictions on the later skill measures as imposed in some previous papers, which can bias the production function estimates (see Agostinelli and Wiswall (2016b) for a discussion).

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 estimated already has a known location and scale, as is the case for the technology estimated by Cunha *et al.* (2010). We show that non-parametric 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)).

In an important extension of our baseline results, we develop additional restrictions on the measurement process which are sufficient for identification of more general production technologies, including those exhibiting Hicks neutral total factor productivity (TFP) dynamics and non-constant returns to scale. Using standard assumptions, these more general technologies cannot be identified because the location and scale of the technology cannot be separately identified from the location and scale of the measures. These more general aspects of the skill development formation process are nonetheless potentially important as restricting the technology can reduce the permissible skill dynamics and productivity of investments, substantially changing our inferences about the child development process and our evaluation of policy. Our paper provides the first identification results for these more general models. Our analysis makes clear the key identification tradeoff researchers face: identification of restricted KLS technologies is possible with standard measurement assumptions, but identification of more general technologies requires stronger assumptions. We evaluate the empirical relevance of these additional assumptions, and provide guidance to researchers to evaluate whether the measures available to them satisfy these assumptions.

In the second part of our paper, we estimate a flexible parametric version of our model using data from the US National Longitudinal Survey of Youth (NLSY). We examine the development of cognitive skills in children from age 5 to age 14, and estimate a model of cognitive skill development allowing for complementarties between parental investment and children's skills; endogenous parental investment responding to the stock of children's skills, maternal skills, and family income; Hicks neutral dynamics in TFP; non-constant returns to scale; and unobserved shocks to the investment process and skill production. Following Cunha *et al.* (2010), our empirical framework treats not only the child's cognitive skills as measured with error, but investment and maternal skills as well.

Constructively derived from our identification analysis, we form a method of moments estimator. Our estimator is not only relatively simple and tractable but also robust because it does not impose parametric distributional assumptions on the distribution of latent skills and measurement errors, as is commonly imposed in previous estimators. We jointly estimate the technology of skill formation, the process of parental investments in children, and the adult distribution of completed schooling and earnings, allowing the production technology and investment process to freely vary as the child ages. Our estimates of high TFP and increasing returns to scale at early ages indicate that investments are particularly productive early in the development period. We also find that the marginal productivity of early investments is substantially higher for children with *lower* existing skills, suggesting the optimal targeting of interventions to disadvantaged children.

Our estimates of the dynamic process of investment and skill development allow us to estimate the heterogeneous treatment effects of some simple policy interventions.

We show that even a modest transfer of family income to families at age 5 would substantially increase children's skills and completed schooling, with the effects larger for low income families. When we compare these estimates to those using models which restrict the technology or ignore measurement error, we estimate policy effects which are substantially smaller, indicating that the generalities we allow are important quantitatively to answering key policy questions.

The paper is organized as follows. In the next two sections, we develop the model of skill development and the measurement process. The next sections analyze the identification of this model, first under weak assumptions about the measurement process, and then under stronger assumptions about measurement which allows the identification of more general technology specifications, including those with TFP dynamics and non-constant returns to scale. The remainder of the paper develops our estimator and discusses our estimation results.

2.2 A Model of Skill Development in Children

In this section, we lay out our simple stylized model of skill development. In later sections, we develop a more detailed, and in many respects more general, empirical model which we take to the data.

Child development takes place over a discrete and finite period, $t = 0, 1, \ldots, T$, where $t = 0$ is the initial period (say birth) and $t = T$ is the final period of childhood (say age 18). There is a population of children and each child in the population is indexed i. For each period, each child is characterized by a stock of skills $\theta_{i,t}$, with $\theta_{i,t} > 0$ for all t and i, and a flow level of investments $I_{i,t}$, withe $I_{i,t} > 0$ for all i and t. For each child, the current stock of skills and current flow of investment produce next period's stock of skill according to the skill formation production technology:

$$\theta_{i,t+1} = f_t(\theta_{i,t}, I_{i,t}) \text{ for } t = 0, 1, \ldots, T-1 \tag{2.1}$$

where equation (2.1) can be viewed as dynamic state space model with $\theta_{i,t+1}$ the state variable for each child i. The production technology $f_t(\cdot)$ is indexed with t to emphasize that the technology can vary over the child development period. According to this technology, the sequence of investments and the initial stock of child skills $\theta_{i,0}$ produces the sequence of skill stocks for each child i: $\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,T}$.

There are several features of the technology which have particular relevance both to understanding the process of child development and in evaluating policy interventions to improve children's skills. We provide a more detailed analysis of policy interventions after the presentation of the full empirical model, but a few brief points are important to emphasize here. First, a key question is the productivity of investments at various child ages. At what ages are investments in children particularly productive in producing future skills ("critical periods") and, conversely, at what ages is it difficult to re-mediate deficits in skill? Second, how does heterogeneity in children's skills, at any given period, affect the productivity of new investments in children? Complementarity in the production technology between current skill stocks and investments implies heterogeneity in the productivity of investments across children. Third, how do investments in children persist over time and affect adult outcomes? Do early investments have a high return because they increase the productivity of later investments (dynamic complementarities) or do early investments "fade-out" over time as they are not reinforced by later investments? These features of the technology of skill development then directly inform the optimal *timing* of policy interventions – the optimal investment portfolio across early and late childhood – and the optimal *targeting* of policy – to which children should scarce resources be allocated to, with the goal of using childhood interventions to affect eventual adult outcomes.

2.3 Measurement

The focus of this paper is estimating the technology determining child skill development (2.1) while accommodating the reality that researchers have at hand various arbitrarily scaled and imperfect measures of children's skills. Our framework recognizes that children's skills are not directly measured by a single measure, but there exists multiple measures which we hypothesize can have some relationship to the unobserved latent skill stock θ_t.

2.3.1 Measurement Model

In our baseline case, we follow the literature and assume a commonly used (log) linear system of measures. In later sections,. we explore a variety of other measures and whether our identification results extend to these other types of measures. Each measure m for child i skills in period (age) t is given by

$$Z_{i,t,m} = \mu_{t,m} + \lambda_{t,m} \ln \theta_{i,t} + \epsilon_{i,t,m}, \tag{2.2}$$

For period t, we have $M_t \in \{1, 2, \dots\}$ measures for each child i skills ($\ln \theta_{i,t}$): $m = 1, 2, \dots, M_t$. $Z_{i,t,m}$ are the measures, $\mu_{t,m}$ are the measurement intercepts, and $\lambda_{t,m}$ are the measurement "factor loadings" or "scaling" parameters, with $\lambda_{t,m} > 0$ for all t and m. The $\mu_{t,m}$ and $\lambda_{t,m}$ measurement parameters allow the latent skills to be represented by arbitrarily located and scaled measures. Finally, $\epsilon_{i,t,m}$ are the individual measurement errors, with $E(\epsilon_{i,t,m}) = 0$ for all t, m (across children), which given the free intercept $\mu_{t,m}$, the assumption of mean zero $\epsilon_{t,m}$ errors is without loss of generality. To focus on the key identification issues, we assume investments I_t are observed without error. In the empirical model which we take to data, we allow for investments to also be measured with error and allow the investments to be endogenously determined by the existing skill stocks.

This measurement system has two important advantages over the alternative approach of using a single measure and assuming it perfectly measure skills, that is assuming $Z_{i,t,m} = \ln \theta_{i,t}$. First, the measurement system allows for noisy measures, in particular allowing measures to differ in their relative "noise" to "signal" ratio, $V(\epsilon_{i,t,m})/\lambda_{t,m}^2 V(\ln \theta_{i,t,m})$, thus allowing for the possibility that some measures have higher correlations to latent skills than others. Given this flexibility the researcher can then form estimators to take advantage of the greater signal some measures have available.

A second advantage is that the measurement parameters allow a kind of "arbitrariness" in the relationship between the measure and the latent skills. An ideal measurement system is one which can accommodate arbitrary changes in the location and scale of measures. Allowing the measures to have free measurement parameters $\mu_{t,m}$ and $\lambda_{t,m}$, which can vary by measure, allows the measurement model to capture the arbitrary location or scaling of particular measures. [2] We show below that the estimator of the primitive production function parameters we develop is robust to changes in the location and scale of the measures up to the initial normalization.

For the remainder of the paper, we omit the children's i subscript to reduce notational clutter. All expectations operations (E, Var, Cov, etc) are defined over the population of children (indexed i). For random variable $X_{i,t}$, we generically define $\kappa_t \equiv E(X_{i,t}) = \int X_{i,t} dF_t$, with F_t the distribution function for random variable $X_{i,t}$ in period t. For simplicity, we drop the i subscript and equivalently write this as $\kappa_t \equiv E(X_t)$.

[2] Measures, such as test scores, can be arbitrarily scaled and located in the sense that for any measure Z, we could create a new measures $Z' = a + bZ$, where a and $b > 0$ are some constants, and the new measure Z' therefore preserves at least the ordinal ranking of latent skills given by Z.

2.3.2 Normalization

Latent skill stocks θ_t have no natural scale and location. A normalization is then required to fix the scale and location of the latent skill stocks to a particular measure. We normalize the latent skill stock to one of the measures of initial period skills:

Normalization 1 *Initial period normalizations*

(i) $E(\ln \theta_0) = 0$

(ii) $\lambda_{0,1} = 1$

This normalization fixes the location and scale of latent skills θ_0 to a particular measure, $Z_{0,1}$, where the choice of the normalizing measure as measure $m = 1$ is arbitrary. For the normalizing measure, we then have the following:

$$Z_{0,1} = \mu_{0,1} + \ln \theta_0 + \epsilon_{0,1},$$

where $\mu_{0,1} = E(Z_{0,1})$ given the normalization $E(\ln \theta_0) = 0$. The latent skill stock θ_0 shares the scale of the normalizing measure in the sense that an 1 unit increase in log latent skills is equal to a 1 unit increase in the level of the normalized measure $Z_{0,1}$: $\frac{\partial Z_{0,1}}{\partial \ln \theta_0} = 1$, where, for intuition, we have treated the Z as a deterministic function. For symmetry with the latent skills, we also normalize log investment to be mean zero in the initial period $E(\ln I_0) = 0$. [3]

While the issue of model normalizations are typically trivial in most cases, in the case of dynamic models such as this, the type of assumed normalization is actually

[3] In practice, if investments are truly observed without error, this can be accomplished by simply de-meaning the investment data so that the sample mean of $\ln I_0$ is zero. In the more general model we estimate, we assume investment is also observed with error and there are multiple measures of latent investments. For now, given we assume investment is observed, this normalization is merely for convenience.

quite important. Our limited normalization for the initial period skills is quite different from the "re-normalization" approach used in much of the prior research (see Cunha and Heckman, 2007; Cunha *et al.*, 2010; Attanasio *et al.*, 2015a,b). In this approach, skills are re-normalized *every* period such that latent skills are assumed to be mean log stationery ($E(\ln \theta_t) = 0$ for all t) and latent skills "load" onto a different arbitrarily measure in each period ($\lambda_{t,1} = 1$ for all t). Agostinelli and Wiswall (2016b) analyze the implications of the re-normalization approach and find that in many standard cases these assumptions are not necessary for point identification and can bias the estimates of the production technology.

We argue that our limited normalization is appropriate for the dynamic setting of child development we analyze. With our normalization for the initial period only, latent skills in *all* periods share a common location and scale with respect to the one chosen normalizing measure. This approach is analogous to deflating a nominal price series to a particular base year; that is, "normalizing" prices to some chosen base year (e.g. 2012 US Dollars). [4] As in the price normalization context, the choice of normalizing skill measure does affect the interpretation of the production function parameters, and we return to this issue when interpreting our particular estimates.

2.3.3 Ignoring Measurement Error

Before analyzing the identification of the model, it is helpful to motivate our analysis by briefly pausing to consider the consequences if we were to ignore measurement error. Consider a simple regression estimator in which we regress a measure of skills

[4] Given the normalizing measure we use is for young children, as children develop, their stock of skill may increase to the extent that the implied measure of skill using the initial normalizing measure $Z_{0,1}$ exceeds the sample maximum level of the measure. This is not an issue for the identification of the model since the measurement system assumes no floor or ceiling to the measures. The measures, and the normalization we use, fixes only the location and scale of the skills, but not the maximum or minimum values. We briefly discuss the issues of measurement floor and ceilings in the Appendix. For an example of an alternative measurement system which respects the discreteness, floor, and ceiling of a particular skill measure, see Del Boca *et al.* (2014b).

in period $t+1$ on a measure of skills in period t:

$$Z_{t+1,m} = \beta_0 + \beta_1 Z_{t,m} + \eta_{t,m}$$

The Ordinary Least Squares (OLS) estimand is

$$\beta_1(OLS) = \frac{Cov(Z_{t+1,m}, Z_{t,m})}{V(Z_{t,m})}$$

Assuming the measurement system above (2.2) and that the measurement errors $\epsilon_{t,m}$ are uncorrelated with latent skills $\ln\theta_t$ for all t, m and uncorrelated across time, we have

$$\beta_1(OLS) = \frac{\lambda_{t+1,m}\lambda_{t,m}Cov(\ln\theta_{t+1}, \ln\theta_t)}{\lambda_{t,m}^2 V(\ln\theta_t) + V(\epsilon_{t,m})}$$

This expression makes clear several problems in naively using observed measures to uncover latent production function relationships given by $Cov(\ln\theta_{t+1}, \ln\theta_t)$. First, the standard issue of attenuation bias: as the "noise" in the measure $V(\epsilon_{t,m})$ increases the OLS estimand goes to 0, biasing the inference of the relationship in latent skills given by $Cov(\ln\theta_{t+1}, \ln\theta_t)$. Second, the OLS estimand $\beta_1(OLS)$ is a combination of model primitives (production technology parameters) and measurement parameters, but we cannot directly separately identify them from the data. One common solution is simply to set $\lambda_{t+1,m} = 1$ and $\lambda_{t,m} = 1$ (a "single measure" approach). If this assumption is incorrect, then the resulting inference about latent production function relationships are biased. [5] The problem is even more severe if we consider

[5] Other approaches include age standardizing the measures such that the measures have 0 mean and standard deviation 1 at each child age. However this approach does not imply $\lambda_{t+1,m} = 1$. Another approach is to re-normalize measures at all periods. This approach biases the resulting estimates. See Agostinelli and Wiswall (2016b) for more discussion of these issues. Similar issues arise if we to examine conditional expectations, $E(Z_{t+1,m}|Z_{t,m})$, instead of covariances. In this case, the intercept of the measurement equations, $\mu_{t,m}$ and $\mu_{t+1,m}$, would also come into play.

regressions including higher order terms (with the goal of identifying some curvature or complementarities in the skill production process):

$$Z_{t+1,m} = \beta_0 + \beta_1 Z_{t,m} + \beta_2 Z_{t,m}^2 + \eta_{t,m}$$

where

$$Z_{t,m}^2 = (\mu_{t,m} + \lambda_{t,m} \ln \theta_t + \epsilon_{t,m})^2$$

In this case, $\lambda_{t,m}$ (factor loadings), $\mu_{t,m}$ (measurement intercepts), and in general the $\epsilon_{t,m}$ distribution need to be identified to uncover structural relationships between latent skills.

2.4 Identification

This section provides our main identification results. These identification results are constructive in the sense that they form the basis of our estimator of the skill development technology.

Our identification analysis proceeds in two steps. First, we identify the distribution of latent skills and investments in the initial period $G_0(\theta_0, I_0)$. Our identification of the initial conditions follows standard arguments used in the current literature (e.g.: Cunha *et al.*, 2010), but for completeness we fully specify this first step of the identification analysis. The second step of our identification analysis is to identify the production technology. This identification analysis is new.

We consider identification under the following assumptions about the joint distribution of latent skills ($\{\theta_t\}_t$), investments $\{I_t\}_t$, and measurement errors ($\{\epsilon_{t,m}\}_{t,m}$):

Assumption 1 *Measurement model assumptions:*

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

(ii) $\epsilon_{t,m} \perp \epsilon_{t',m'}$ *for all* $t \neq t'$ *and all* m *and* m'

(iii) $\epsilon_{t,m} \perp I_{t'}$ *for all* t *and* t' *and all* m

(iv) $\epsilon_{t,m} \perp \theta_{t'}$ *for all* t *and* t' *and all* m

Assumption 1 (i) is that measurement errors are independent contemporaneously across measures. Assumption 1 (ii) is that measurement errors are independent over time. Assumption 1 (iii) and (iv) are that measurement errors in any period are independent of the latent stock of skills and parental investments in any period. While these assumptions are strong in some sense, they are common in the current literature. [6]

2.4.1 Identification of Initial Conditions

Under Normalization 1, Assumption 1, and with at least 3 measures in the first period, $M_0 \geq 3$, we identify the $\lambda_{0,2}, \lambda_{0,3}, \ldots, \lambda_{0,M_0}$ factor loadings from ratios of measurement covariances:

$$\lambda_{0,m} = \frac{Cov(Z_{0,m}, Z_{0,m'})}{Cov(Z_{0,1}, Z_{0,m'})}, \tag{2.3}$$

for $m \neq m'$, $m \neq 1$, $m' \neq 1$, where measure $m = 1$ is the normalizing measure.

Further, under the normalization that $E(\ln \theta_0) = 0$ (Normalization 1), we identify the $\mu_{0,1}, \mu_{0,2}, \ldots, \mu_{0,M_0}$ intercepts from

$$\mu_{0,m} = E(Z_{0,m}). \tag{2.4}$$

We then construct the following "residual" skill measures from the original raw measures:

[6] Our assumption of full independence is sufficient, but not necessary, for at least some of our identification analysis. Below, we point out instances where weaker assumptions, allowing for some forms of dependence among measures and among measures and latent variable, can be used for identification.

$$\widetilde{Z}_{0,m} = \frac{Z_{0,m} - \mu_{0,m}}{\lambda_{0,m}}, \tag{2.5}$$

where $\widetilde{Z}_{0,m}$ identifies the sum of the latent skill and a scaled version of the measurement error:

$$\widetilde{Z}_{0,m} = \ln \theta_0 + \frac{\epsilon_{0,m}}{\lambda_{0,m}}.$$

Applying the Kotalarski Theorem (Kotalarski 1964) to the $\{\widetilde{Z}_{0,m}\}_{m=1}^{M_0}$ residual measures, conditional on each level of investment I_0, we identify the distribution of θ_0 for any level of investment I_0. This then allows us to identify the joint distribution of latent skills and investment in the initial period $G_0(\theta_0, I_0)$, up to the normalizations given in Normalization 1. [7]

2.4.2 *Identification of the Production Technology*

With the initial distribution for latent skills and investments $G_0(\theta_0, I_0)$ identified in the first step, we next identify the process of child development given by the sequences of production technologies $f_0(\theta_0, I_0), \ldots, f_{T-1}(\theta_{T-1}, I_{T-1})$. Our identification analysis is sequential, and uses the production technology in period t, $f_t(\theta_t, I_t)$, to identify the distribution of latent skills (and investments) in the next period, $G_{t+1}(\theta_{t+1}, I_{t+1})$. We first establish a general identification result for any periods t and $t+1$. We then conclude this section by describing the sequence of identification steps starting from the initial period $t = 0$.

[7] The key necessary condition for the Kotalarski theorem to hold in this case is that at least two of the residual measures in the set of measures $\{\widetilde{Z}_{0,m}\}_{m=1}^{M_0}$ have full support conditional on I_0, that is $\widetilde{Z}_{0,m} \in \mathbb{R}$ conditional on I_0.

From Measures to Latent Relationships

Given the generalities we have allowed, in which we do not assume that skills are measured perfectly in data, identification of the production technology now poses considerable challenges. The production technology in some period t would in principle be identified by the relationship between output θ_{t+1} and inputs θ_t, I_t. We do not directly observe latent skills θ_{t+1} or θ_t in data. Instead, we observe relationships among measures $Z_{t+1,m}$ and $Z_{t,m}$. Under Assumption 1, we have the following relationship between measures and latent variables:

$$E(Z_{t+1,m}|Z_{t,m}, \ln I_t) = \mu_{t+1,m} + \lambda_{t+1,m}E(\ln\theta_{t+1}|Z_{t,m}, \ln I_t)$$

This expression shows that $E(Z_{t+1,m}|Z_{t,m}, \ln I_t)$ does not identify a production function relationship directly, but instead a combination of latent skill relationships and measurement parameters.

In the following Lemma, we first show that we can identify dynamic production function relationships, $E(\ln\theta_{t+1}|\ln\theta_t, \ln I_t)$, from measures of latent skills in periods t and $t+1$, $Z_{t,m}$ and $Z_{t+1,m}$, up to the measurement parameters for the $t+1$ measure, $\mu_{t+1,m}$ and $\lambda_{t+1,m}$. [8]

Lemma 1 *Given i) $G_t(\theta_t, I_t)$ is known, ii) a pair of measures $Z_{t,m}$ and $Z_{t+1,m}$ which satisfy Assumption 1, and iii) measurement parameters for $Z_{t,m}$ ($\mu_{t,m}$, and $\lambda_{t,m}$) are known, $E(Z_{t+1,m}|\ln\theta_t = a, \ln I_t = \ell)$ is identified for some $(a,\ell) \in \mathbb{R}_2$ and is equal to $\mu_{t+1,m} + \lambda_{t+1,m}E(\ln\theta_{t+1}|\ln\theta_t = a, \ln I_t = \ell)$.*

Proof. See Appendix. ■

[8] Note that the measure m for $t+1$, $Z_{t+1,m}$, which is used to measure $\ln\theta_{t+1}$, can be a completely different "kind" of measure from the measure $Z_{t,m}$ used to measure $\ln\theta_t$. The use of the same measure index m does not connote any relationship.

Lemma 1 establishes that while we cannot use realizations of measures $Z_{t,m} = z$ to identify particular values of the latent variable $\theta_t = p$, we can identify moments of the latent distribution.

Substituting the production technology $\theta_{t+1} = f_t(\theta_t, I_t)$, Lemma 1 shows that measures $Z_{t+1,m}$ and $Z_{t,m}$ identify the following:

$$E(Z_{t+1,m}| \ln\theta_t = a, \ln I_t = \ell) = \mu_{t+1,m} + \lambda_{t+1,m} \ln f_t(e^a, e^\ell) \quad (2.6)$$

Note that the left-hand side of (2.6) is not directly observed in data (given the unobservability of $\ln\theta_t$) but is identified from observed measures (Lemma 1). The right-hand side of (2.6) is a combination of production function relationships and measurement parameters $\mu_{t+1,m}$ and $\lambda_{t+1,m}$. [9] If we do not know the measurement parameters $\mu_{t+1,m}$, $\lambda_{t+1,m}$, we cannot directly use $E(Z_{t+1}| \ln\theta_t, \ln I_t)$ to identify the production technology. One simple but problematic solution to this problem is to assume values for the $\mu_{t+1,m}$ and $\lambda_{t+1,m}$ parameters, and identification is trivially obtained. [10] However, if the assumptions on the measurement parameters are incorrect, then estimation under these assumptions can be biased.

Transformations to Eliminate Measurement Parameters

Our solution to this problem is to treat the measurement parameters $\mu_{t,m}$ and $\lambda_{t,m}$ for all $t > 0$ as "nuisance" parameters and use transformations of the moments (2.6) to eliminate them. Using four pairs of $(\ln\theta_t, \ln I_t) = \{(a_1, l_1), (a_2, l_2), (a_3, l_3), (a_4, l_4)\}$,

[9] Note that we have used the fact that in our stylized model, there are no stochastic elements to the production process, hence θ_{t+1} given θ_t, I_t is a constant for all t. We return to the topic of how to identify a shock to the production technology below. In brief, adding a mean 0 (log) shock to the production technology does not change the main identification analysis. Re-write the technology as $\theta_{t+1} = f_t(\theta_t, I_t)\exp(\eta_t)$, where $E(\eta_t) = 0$ and η_t is independent of θ_t, I_t, and $\epsilon_{t,m}, \epsilon_{t+1,m}$. Log skills are then $\ln\theta_{t+1} = \ln f_t(\theta_t, I_t) + \eta_t$, and mean log skills are $\ln f_t(\theta_t, I_t)$ as before.

[10] For example, the researcher could assume the values $\mu_{t,m} = 0$ and $\lambda_{t,m} = 1$ for all t and m, as in the case when all measures are assumed to be "classical" in the sense that $Z_{t,m} = \ln\theta_t + \epsilon_{t,m}$, and $\epsilon_{t,m}$ is simply a mean zero measurement error.

we compute the following transformation of the conditional expectations:

$$\frac{E(Z_{t+1,m}|\ln\theta_t = a_1, \ln I_t = \ell_1) - E(Z_{t+1,m}|\ln\theta_t = a_2, \ln I_t = \ell_2)}{E(Z_{t+1,m}|\ln\theta_t = a_3, \ln I_t = \ell_3) - E(Z_{t+1,m}|\ln\theta_t = a_4, \ln I_t = \ell_4)}$$
$$= \frac{\ln f_t(e^{a_1}, e^{\ell_1}) - \ln f_t(e^{a_2}, e^{\ell_2})}{\ln f_t(e^{a_3}, e^{\ell_3}) - \ln f_t(e^{a_4}, e^{\ell_4})} \tag{2.7}$$

where the values $a_k, \ell_k, k = 1, 2, 3, 4$ are such that $E(Z_{t+1,m}|\ln\theta_t = a_3, \ln I_t = \ell_3) \neq E(Z_{t+1,m}|\ln\theta_t = a_4, \ln I_t = \ell_4)$.

The left-hand side of (2.7) is a transformation of moments which are identified directly from the measures of skills for periods $t+1$ and t (Lemma 1), and the right-hand side is the corresponding transformation of the technology. The transformation in (2.7) has eliminated the measurement parameters $\mu_{t+1,m}$ and $\lambda_{t+1,m}$ without making any assumption about their values. This transformation is analogous to the transformation used in panel data analysis where differences at the observation level are used to eliminate common fixed effects. [11] As in the panel data literature, we exploit the particular form of the measurement equations and Assumption 1 to find an appropriate transformation to eliminate the nuisance measurement parameters. Other transformations can accomplish the same goal, and for convenience in some examples, we work with ratios of covariances, which already implicitly eliminate dependence on the measurement intercepts $\mu_{t,m}$.

Location and Scale of the Production Technology

Much of our analysis centers on the classes of production technologies which can be identified given that some inputs (latent skills) are measured with error. Crucial to our analysis is whether the production technology has a known location and scale or whether the location or scale is unknown in the sense that it depends on free

[11] Consider the model $y_{i,t} = \mu_i + X'_{i,t}\beta + \epsilon_{i,t}$, and the within transformation of the data $y_{i,t+1} - y_{it}$ eliminates the μ_i fixed effects.

parameters which need to be estimated. This concept is new to the production function identification literature, as far as we know. This concept is key to our analysis because our results below show that we can identify the production technologies up to location and scale, and can therefore point identify production technologies which already have a known location and scale.

We first define the concept of a production function with "known location and scale":

Definition 1 *A production function* $f_t(\theta_t, I_t)$ *has* known location and scale (KLS) *if for two non-zero input vectors* (θ_t', I_t') *and* (θ_t'', I_t'')*, where the input vectors are distinct* ($\theta_t' \neq \theta_t''$ *or* $I_t' \neq I_t''$)*, the output* $f_t(\theta_t', I_t')$ *and* $f_t(\theta_t'', I_t'')$ *are both known (do not depend on unknown parameters), finite, and non-zero.*

A production technology with known location and scale implies that for a change in inputs from (θ_t', I_t') to (θ_t'', I_t''), the change in output $f_t(\theta_t', I_t') - f_t(\theta_t'', I_t'')$ is known. Other points in the production possibilities set may be unknown, i.e. depend on free parameters to be estimated.

For example, consider the class of Constant Elasticity of Substitution (CES) skill production technologies, the class of technologies estimated in a number of previous studies (e.g.: Cunha *et al.*, 2010). [12] The CES technology is

$$\theta_{t+1} = (\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t})^{1/\phi_t}. \tag{2.8}$$

with $\gamma_t \in (0,1)$ and $\phi_t \in (-\infty, 1]$, and $\phi_t \to -\infty$ (Leontif), $\phi_t = 1$ (linear), $\phi_t \to 0$ (log-linear, Cobb-Douglas). The elasticity of substitution is $1/(1-\phi_t)$.

[12] The functions estimated by Cunha *et al.* (2010) also include a mean zero production function shock. We consider identification of these functions below. In general, including a mean zero shock does not change the main results of the identification analysis, see Footnote 9.

The production technology (2.8) satisfies Definition 1 because for inputs $I_t = \theta_t = \alpha > 0$, $\theta_{t+1} = \alpha$. That is, for inputs which are known to be equal at value α, we also know the output is α as well. This property of known location and scale is related to constant returns to scale property of this function, but constant returns to scale is not necessarily a sufficient property to satisfy Definition 1, as shown below. While the scale and location of the production function (2.8) are known, other points in the production possibilities set are determined by the free parameters γ_t and ϕ_t. Identifying these remaining parameters is the subject of the section.

Another example of KLS production technologies are those based on the translog function, a generalization of the Cobb-Douglas production technology which does not restrict the elasticity of substitution to be constant:

$$\ln \theta_{t+1} = \gamma_{1t} \ln \theta_t + \gamma_{2t} \ln I_t + \gamma_{3t} (\ln \theta_t)(\ln I_t) \tag{2.9}$$

with $\sum_{j=1}^{3} \gamma_{jt} = 1$. Consider the points $(\theta_t, I_t) = (1, 1)$ and (e, e). For these points, the output of the production technology is known at $\ln \theta_{t+1} = 0$ and 1, respectively, and thus this function satisfies Definition 1.

In contrast, a class of technologies which does not satisfy the known location and scale property (Definition 1) is the following

$$\theta_{t+1} = A_t (\gamma_t \theta_t^{\phi_t} + (1 - \gamma_t) I_t^{\phi_t})^{1/\phi_t} \tag{2.10}$$

with $A_t > 0$ representing Total Factor Productivity (TFP). The previous case (2.8) is a special case of (2.10) with $A_t = 1$. In the more general case, the addition of the unknown TFP process term A_t implies that the scale of the function is unknown. For example, for $\theta_t = I_t = \alpha > 0$, we have $f_t(\alpha, \alpha) = A_t \alpha$, where A_t is a free parameter. This class of technologies has constant returns to scale but does not have a known location and scale. [13]

Another class of technologies which does not satisfy Definition 1 is CES technologies without constant returns to scale:

$$\theta_{t+1} = (\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t})^{\psi_t/\phi_t}. \tag{2.11}$$

where $\psi_t > 0$ is a returns to scale parameter, with $\psi_t = 1$ constant returns to scale, $\psi_t < 1$ decreasing returns to scale, and $\psi_t > 1$ increasing returns to scale. In this case, $f_t(\alpha, \alpha) = \alpha^{\psi_t}$. For this function, while we know the point $f_t(1,1) = 1$, and can identify the location of the function, we do not know a second point in the production possibilities set, and therefore cannot identify the scale of the function. Similarly, the translog function (2.9) with $\sum_{j=1}^{3} \gamma_{j,t}$ not equal to a known constant would not satisfy the KLS definition (Definition 1).

Per Period Identification of the Production Technology

We next proceed to the main identification result. We show that with the distribution of skills and investments in period t, $G_t(\theta_t, I_t)$ and the measurement parameters, $\mu_{t,m}$ and $\lambda_{t,m}$, known for period t, then a single measure of skills in period $t+1$, $Z_{t+1,m}$, with sufficient support, non-parametrically identifies a production technology $\theta_{t+1} = f_t(\theta_t, I_t)$ with known location and scale (satisfying Definition 1). The key aspect of the identification result is that we identify the production technology without knowledge of the period $t+1$ measurement parameters, $\mu_{t+1,m}$ and $\lambda_{t+1,m}$. We specify the exact conditions for identification in the following theorem.

[13] Similarly, a function where the factor share parameters did not sum to a known constant would also lack a known scale, for example $\theta_{t+1} = (\gamma_{1t}\theta_t^{\phi_t} + \gamma_{2t} I_t^{\phi_t})^{1/\phi_t}$, with $\gamma_{1t} + \gamma_{2t} \neq 1$. In this case, $\theta_t = \alpha$, $I_t = \alpha$, we have $f_t(\alpha, \alpha) = ((\gamma_{1t} + \gamma_{2t})\alpha)^{1/\psi_t}$.

Theorem 1 *If i) the distribution of skills $G_t(\theta_t, I_t)$ is known, ii) measurement parameters $\mu_{t,m}, \lambda_{t,m}$ are known, iii) there exists at least one measure $Z_{t+1,m}$ which satisfies Assumption 1, iv) the measure $Z_{t+1,m}$ has full support, $Z_{t+1,m} \in \mathbb{R}$, and v) the production technology $f_t(\theta_t, I_t)$ has known location and scale (Definition 1), then the production technology $f_t(\theta_t, I_t)$ is identified for all $(\theta_t, I_t) \in \mathbb{R}^2_+$.*

Proof. See Appendix. ■

Theorem 1 indicates that we can identify production technologies which have a known scale (Definition 1), such as the CES technologies (2.8) considered in much of the previous literature (see Cunha and Heckman, 2007; Cunha *et al.*, 2010). The limitation of Theorem 1 is that we cannot apply it to more general production technologies which do not satisfy the known location and scale property. In the next section, we propose stronger assumptions on the measurement process which could allow for identification of more general production technologies.

Sequential Production Function Identification

Theorem 1 shows identification of the production technology for period t, $\theta_{t+1} = f_t(\theta_t, I_t)$, given measures $Z_{t+1,m}$ and $Z_{t,m}$. We now apply these per-period results to show how we can sequentially identify the full sequence of production technologies, $f_0(\theta_0, I_0), \ldots, f_{T-1}(\theta_{T-1}, I_{T-1})$, and hence the distribution of the sequence of skill stocks $(\theta_1, \ldots, \theta_T)$. The minimal data we require are at least 3 measures of latent skills for the initial period $Z_{0,1}, Z_{0,2}, Z_{0,3}$, and a single measure m of latent skills in the following periods, $Z_{1,m}, Z_{2,m}, \ldots, Z_{T,m}$.

The sequential identification proceeds as follows. First, using the measures for the initial period, following the analysis above, we identify the initial distribution of skills and investments $G_0(\theta_0, I_0)$, and initial measurement parameters, $\mu_{0,m}$ and

$\lambda_{0,m}$, for some measure $Z_{0,m}$. Then, applying Theorem 1, we identify the production technology for period 0, $\theta_1 = f_0(\theta_0, I_0)$, where the technology is assumed to be of the known location and scale class (satisfying Definition 1). With the production function identified, we identify the distribution of period 1 skills from the production technology:

$$G_1(\theta_1|I_0) = pr(\theta_1 \leq \theta|I_1) = \int pr(f_0(\theta_0, I_0) \leq \theta) dG_0(\theta_0|I_0)$$

where $G_1(\theta_1|I_0)$ is the conditional distribution of latent skills, which given that investments are assumed observed, can then be used to identify the joint distribution of skills and investment.

We then proceed to identify the measurement parameters for measure $Z_{1,m}$ used to measure period 1 latent skills. The factor loadings can be identified from the across time correlation in measures of skills, $Cov(Z_{1,m}, Z_{0,m})$:

$$\lambda_{1,m} = \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m} Cov(\ln\theta_1, \ln\theta_0)}.$$

where $Cov(\ln\theta_1, \ln\theta_0)$ is identified from the production technology and the initial distribution of skills:

$$\begin{aligned} Cov(\ln\theta_1, \ln\theta_0) &= Cov(\ln f_0(\theta_0, I_0), \ln\theta_0). \\ &= \int (\ln f_0(\theta_0, I_0) \ln\theta_0) dG(\theta_0, I_0) \end{aligned}$$

The measurement intercept for period 1 is then identified from

$$\mu_{1,m} = E(Z_{1,m}) - \lambda_{1,m} E(\ln\theta_1),$$

where as above $E(\ln\theta_1)$ is identified from the production technology, as above:

$$E(\ln\theta_1) = \int \ln f_0(\theta_0, I_0) dG(\theta_0, I_0).$$

This shows the identification of the technology $f_0(\theta_0, I_0)$ and the measurement parameters for $\mu_{1,m}$ and $\lambda_{1,m}$. We can continue to follow these steps, applying Lemma 1 and Theorem 1 sequentially, to identify the technology in the next periods, $f_1(\theta_1, I_1), \ldots, f_{T-1}(\theta_{T-1}, I_{T-1})$.

2.4.3 Intuition

Before we continue with examples and extensions to our identification concept, we pause to consider some simple intuition for our idea in a general setting. Consider a general production technology $Y = f(X_1, X_2)$, where Y is some latent unobserved output and X_1 and X_2 are some observed inputs. We have measure Z of the output, and the measure has error of the form we consider above: $Z = \mu + \lambda \ln Y + \epsilon$, where ϵ is uncorrelated with $\ln Y$, X_1, and X_2, and μ and λ are measurement parameters. The ratio of covariances of the measure of the output Z with the two inputs X_1, X_2 is

$$\begin{aligned} \frac{Cov(Z, X_1)}{Cov(Z, X_2)} &= \frac{\lambda Cov(\ln Y, X_1)}{\lambda Cov(\ln Y, X_2)} \\ &= \frac{Cov(\ln Y, X_1)}{Cov(\ln Y, X_2)}. \end{aligned}$$

The ratio of covariances has eliminated the "nuisance" measurement parameter λ. Working with covariances, rather than conditional expectations, has already eliminated dependence on the measurement intercept μ. This expression makes clear that even with output mis-measured in data and with free unknown measurement parameters allowing for arbitrary scale and location, we can still learn something about the production technology. For example, the ratio in this example is related to the relative marginal product of the two inputs X_1, X_2. Considering ratios of higher order covariances, such as $Cov(Z, X_1^2)/Cov(Z, X_1)$ and $Cov(Z, X_1 X_2)/Cov(Z, X_1)$, can similarly provide information about the "curvature" of the production function and the degree of complementarities between inputs. Our results above show identification in models which generalize this simple example, allowing for a dynamic production technology and mis-measured inputs as well.

2.4.4 Examples

We next proceed to demonstrate the identification results using simple two period models and commonly used production technologies.

Example 1 *Log-Linear (Cobb-Douglas) Technology*

There are two periods $T = 2$. Skills in $t = 1$ are given by the following log-linear (Cobb-Douglas) production technology:

$$\ln \theta_1 = \gamma_0 \ln \theta_0 + (1 - \gamma_0) \ln I_0 \tag{2.12}$$

where $\gamma_0 \in (0, 1)$ is the unknown production function parameter we would like to identify. Like the more general CES class to which it belongs, this production function has a known location and scale (Definition 1).

We have three measures of initial period skills: $Z_{0,1}, Z_{0,2}, Z_{0,3}$. We have one measure of skills in period 1, $Z_{1,m}$. The measures satisfy Assumption 1.

We normalize initial period skills as $E(\ln \theta_0) = 0$ and initial investments $E(\ln I_0) = 0$. We normalize the factor loading for the first measure as $\lambda_{0,1} = 1$. Following the analysis above, we then identify the remaining measurement factor loadings $\lambda_{0,2}, \lambda_{0,3}$ and measurement intercepts $\mu_{0,1}, \mu_{0,2}, \mu_{0,3}$ for the initial period measures. We then identify the joint distribution of the latent skills and investments, $G_0(\theta_0, I_0)$. Applying

Lemma 1 identifies $E(Z_{1,m}|\ln\theta_0, \ln I_0)$ for values of $\ln\theta_0, \ln I_0$ from the measures $Z_{1,m}$ and $Z_{0,m}$ and the identified measurement parameters $\mu_{0,m}$ and $\lambda_{0,m}$.

Next we apply Theorem 1 to identify the production function parameter γ_0. The key to our analysis is that we identify the production function primitive without making any assumptions about the values of measurement parameters $\mu_{1,m}$ or $\lambda_{1,m}$. We compute the following transformations of conditional expectations (algebra is given in the Appendix):

$$\frac{E(Z_{1,m}|\ln\theta_0 = a, \ln I_0 = 0) - E(Z_{1,m}|\ln\theta_0 = 0, \ln I_0 = 0)}{E(Z_{1,m}|\ln\theta_0 = 1, \ln I_0 = 1) - E(Z_{1,m}|\ln\theta_0 = 0, \ln I_0 = 0)} = \frac{\gamma_0 a}{1}$$

Letting Δ be the left-hand side of the expression and solving for the production function parameter, we have

$$\gamma_0 = \frac{\Delta}{a}$$

This expression shows that the unknown parameter γ_0 of the production technology is identified from the transformation of the conditional expectations. Identification of γ_0 requires only a single measure of latent skills in period 1 and is invariant to the measurement parameters, $\mu_{1,m}$ and $\lambda_{1,m}$.

With the production technology $f_0(\theta_0, I_0)$ identified, we can now identify the measurement parameters for $Z_{1,m}$. $\lambda_{1,m}$ is identified from

$$\lambda_{1,m} = \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m}Cov(\ln\theta_1, \ln\theta_0)}$$

where we can use any of the three first period measures, $m = 1, 2, 3$ to form the right-hand side. Substituting for the production technology, we have

$$\lambda_{1,m} = \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m}Cov(\gamma_0\ln\theta_0 + (1-\gamma_0)\ln I_0, \ln\theta_0)}$$

$$= \frac{Cov(Z_{1,m}, Z_{0,m})}{\lambda_{0,m}(\gamma_0 V(\ln \theta_0) + (1-\gamma_0)Cov(\ln \theta_0, \ln I_0))}.$$

Given the identification of γ_0, and that we have already identified the initial joint distribution of θ_0, I_0 (and can compute $V(\ln \theta_0)$ and $Cov(\ln \theta_0, \ln I_0)$), we can compute the right-hand side.

$\mu_{1,m}$ is then identified from

$$\mu_{1,m} = E(Z_{1,m}) - \lambda_{1,m} E(\ln \theta_1),$$

where, for this particular production technology, we have $E(\ln \theta_1) = \gamma_0 E(\ln \theta_0) + (1-\gamma_0)E(\ln I_0) = 0$, given the normalization for the initial period ($E(\ln \theta_0) = 0$ and $(E(\ln I_0) = 0)$. As described in more detail below, the mean of log latent skills will in general not be 0 in periods after the initial period. For alternative production functions, $E(\ln \theta_1)$ can be computed from the identified production technology.

Example 2 *General CES Technology*

In our second example, we maintain the same setup as Example 1 but consider the general CES function (2.8):

$$\theta_1 = (\gamma_0 \theta_0^{\phi_0} + (1-\gamma_0) I_0^{\phi_0})^{1/\phi_0}.$$

with parameters defined as in (2.8). For this technology, there are two unknown production function parameters we wish to identify, γ_0 and ϕ_0. We have the same measures as in Example 1 and identify the initial condition as before.

As in Example 1, we compute the following:

$$\frac{E(Z_{1,m}| \ln \theta_0 = \ln a_1, \ln I_0 = \ln \ell_1) - E(Z_{1,m}| \ln \theta_0 = \ln a_2, \ln I_0 = \ln \ell_2)}{E(Z_{1,m}| \ln \theta_0 = \ln a_3, \ln I_0 = \ln \ell_3) - E(Z_{1,m}| \ln \theta_0 = \ln a_4, \ln I_0 = \ln \ell_4)} = \frac{\ln f_0(a_1, \ell_1) - \ln f_0(a_2, \ell_2)}{\ln f_0(a_3, \ell_3) - \ln f_0(a_4, \ell_4)}$$

Now define Δ_1 to be the left-hand side of the above equation and take values $a_1 \neq 0$, $a_3 \neq 0$, where $a_1 \neq a_3$, $a_2 = a_4 = \ell_2 = \ell_4 = 1$, $\ell_1 = 0$ and $a_3 = \ell_3 = e^1$. We have (see Appendix for omitted algebra):

$$\Delta_1 = \frac{\ln f_0(a_1, 0) - \ln f_0(1, 1)}{\ln f_0(e^1, e^1) - \ln f_0(1, 1)},$$

$$\Delta_1 = \frac{\ln(\gamma_0 a_1)}{1},$$

Solving for γ_0 , we have

$$\gamma_0 = \frac{e^{\Delta_1}}{a_1}$$

This expression identifies γ_0. With γ_0 identified, we form a second ratio:

$$\Delta_2 = \frac{\ln f_0(a_1, 1) - \ln f_0(1, 1)}{\ln f_0(a_3, 0) - \ln f_0(1, 1)},$$
$$= \frac{\ln(\gamma_0 a_1^{\phi_0} + 1 - \gamma_0)}{\ln(\gamma_0 a_3)},$$

Solving for ϕ_0 (see Appendix for omitted algebra), we have

$$\phi_0 = \frac{\ln\left(\frac{(\gamma_0 a_3)^{\Delta_2} - 1 + \gamma_0}{\gamma_0}\right)}{\ln(a_1)}$$

This analysis shows that a single measure of period 1 skills identifies the unknown production function parameters γ_0, ϕ_0 without imposing any restrictions on the values of the period 1 measurement parameters. We can follow the same analysis as in Example 1 to identify the measurement parameters.

2.4.5 Comparison to Cunha et al. *(2010)*

Our results show identification of a production function with known location and scale without imposing any particular values for the measurement parameters after the initial period. Cunha *et al.* (2010) provide identification results in which they not only normalize initial period latent skills, as we do here, but also "re-normalize" latent skills each period. In our notation, their re-normalization restriction is $E(\ln\theta_0) = E(\ln\theta_1) = \cdots = E(\ln\theta_T) = 0$ and $\lambda_{0,1} = \lambda_{1,1} = \cdots = \lambda_{T,1} = 1$ for the normalized measure $m = 1$.

Some normalization is necessary (and we impose a normalization on the initial period), but, as we prove here, the additional restrictions on later periods are not necessary for identification of a known location and scale technology. The function Cunha *et al.* (2010) estimate is a known location and scale CES technology of the form given by (2.8). Because this function is already restricted (as compared to the more general functions with non-constant returns to scale and TFP dynamics), the additional normalizations are unnecessary and over-identifying. Importantly these re-normalization restrictions are not cost free as these additional normalizations can bias the technology estimates toward the Cobb-Douglas technology and away from more general patterns of substitution (see Agostinelli and Wiswall, 2016b).

2.4.6 Errors-in-Variables Formulation

The KLS class of technologies can also be understood as a restriction in a traditional error-in-variables model (Chamberlain (1977a)). In this literature, identification is often achieved by proportionality restriction (linear regression parameters are assumed proportional to each other) within the context of a "reduced form" linear regression model. In our case, the restrictions we consider come from restrictions on the primitive production function, which is intuitively appealing because we can understand the consequences of these restrictions on the primitive production relationships.

Consider the Cobb-Douglas case (2.12). Using the normalizations on the initial period, we proceed as before and form measures for the initial period:

$$\widetilde{Z}_{0,m} = \frac{Z_{0,m} - \mu_{0,m}}{\lambda_{0,m}} = \ln\theta_0 + \epsilon_{0,m}.$$

We also have a single measure of period 1 skills θ_1 given by

$$Z_{1,m} = \mu_{1,m} + \lambda_{1,m}\ln\theta_1 + \epsilon_{1,m}$$

As in all of our analysis above, the measurement parameters $\mu_{1,m}$ and $\lambda_{1,m}$ are treated as free parameters.

Substituting the production technology into the period 1 measurement equation, we have

$$Z_{1,m} = \mu_{1,m} + \lambda_{1,m}[\gamma_0\ln\theta_0 + (1-\gamma_0)\ln I_0] + \epsilon_{1,m}$$

Substituting one of the measures for $\ln\theta_0$, say $\widetilde{Z}_{0,m}$, we have

$$Z_{1,m} = \mu_{1,m} + \lambda_{1,m}[\gamma_0(\widetilde{Z}_{0,m} - \widetilde{\epsilon}_{0,m}) + (1-\gamma_0)\ln I_0] + \epsilon_{1,m}$$

with $\widetilde{\epsilon}_{0,m} = \epsilon_{0,m}/\lambda_{0,m}$.

Re-arranging, we have

$$\begin{aligned} Z_{1,m} &= \mu_{1,m} + \lambda_{1,m}\gamma_0\widetilde{Z}_{0,m} + \lambda_{1,m}(1-\gamma_0)\ln I_0 + (\epsilon_{1,m} - \lambda_{1,m}\gamma_0\widetilde{\epsilon}_{0,m}) \\ &= \beta_0 + \beta_1\widetilde{Z}_{0,m} + \beta_2\ln I_0 + \pi_{1,m} \qquad (2.13) \end{aligned}$$

where $\beta_0 = \mu_{1,m}$, $\beta_1 = \lambda_{1,m}\gamma_0$, $\beta_2 = \lambda_{1,m}(1-\gamma_0)$, and $\pi_{1,m} = \epsilon_{1,m} - \lambda_{1,m}\gamma_0\widetilde{\epsilon}_{0,m}$. The "reduced form" equation (2.13) now has the standard errors-in-variables form: (2.13) is a linear regression of a measure of period 1 skills $Z_{1,m}$ on a measure for period 0 skills $\widetilde{Z}_{0,m}$. The β_1 and β_2 coefficients are combinations of the measurement factor loading $\lambda_{1,m}$ and the production function parameter γ_0.

Identification takes two steps. First, the standard error-in-variables problem is that the OLS regression estimands for β_1 and β_2 do not identify β_1 and β_2. We can solve this problem using any number of standard techniques. In this setting with multiple measures available satisfying independence assumptions, a second measure for period 0 skills, $\widetilde{Z}_{0,m'}$, can be used as an instrument for $\widetilde{Z}_{0,m}$, and we identify β_1 and β_2. Second, with β_1 and β_2 identified, we can then solve for the underlying primitive parameters γ_0 and $\lambda_{1,m}$:

$$\gamma_0 = \frac{\beta_1}{\beta_1+\beta_2}\ ,\ \lambda_{1,m} = \beta_1+\beta_2 \text{ and } \mu_{1,m} = \beta_0$$

The key to the identification here is that this commonly used production function (2.12) is already restricted (the factor shares sum to 1) and hence we can identify the production function parameters separately from the measurement parameters. Without this restriction on the production function, a function $\ln\theta_1 = \gamma_{0,\theta}\ln\theta_0 + \gamma_{0,I}\ln I_0$, where $\gamma_{0,\theta}$ and $\gamma_{0,1}$ are free parameters and do not sum to 1, point identification is not possible as there would be three unknown parameters $\gamma_{0,\theta}, \gamma_{0,I}$, and $\lambda_{1,m}$ and only two regression coefficients β_1, β_2.

2.4.7 Robustness to Alternative Types of Measures

One of the characteristics of the data used to study child development is the rich variety of skill measures. Here we considered identification where the skill measures are in a "raw" form: each measure is a linear function of the latent log skill. This measurement system, while commonly assumed in the prior literature, is in some respects a "best case." In the Appendix, we briefly discuss alternative forms of measures and re-examine whether we can identify the same types of production technologies using these alternative measures. We consider four classes of measures which are sometimes encountered empirically: (i) *age-standardized* measures where the raw measures are transformed ex post (in the sample) to have mean 0 and standard deviation 1; (ii) *relative* measures where the measures reflect not the level of a child's skill but the child's skill relative to the population mean (i.e. other children); (iii) *ordinal* measures which provide a discrete ranking of children's skills; and (iv) *censored* measures where the measures are truncated with a "floor" (finite minimum value) and/or a "ceiling" (finite maximum value).

For the age-standardized and relative measures, we find that our identification results continue to hold because these alternative measures can be expressed as alternative linear functions of the latent skills with particular measurement intercepts and factor loadings. Our identification results are invariant to these measurement parameters as the measurement parameters would be "transformed away," as described above (2.7). More generally, our identification results are robust to any linear increasing transformation of the original raw measures. On the other hand, without additional assumptions, the latter two classes of measures would appear to not allow non-parametric identification, at least globally, as these measures do not provide a one-to-one mapping between latent variables and measures (in expectation) as with the linear continuous measurement system we consider here.

2.5 Identification of General Technologies

The preceding analysis demonstrated that production functions with known location and scale (KLS, Definition 1) are non-parametrically identified using measures of latent skills that satisfy Assumption 1. This class of production technologies include the CES technologies analyzed in much of the previous work (see Cunha and Heckman, 2007; Cunha *et al.*, 2010). These types of production functions are restricted, and these restrictions can affect our inferences about the child development process and the effects of policy interventions, as we demonstrate empirically below. [14] We next consider classes of technologies which are more general and no longer have a known location and scale, and we analyze identification of these more general technologies under additional assumptions about the measurement error process. We conclude this section with a discussion of what empirical measures may justify these additional assumptions.

2.5.1 Identifying Production Technologies with Dynamics in TFP

Consider a general class of technologies which exhibit Hicks-neutral TFP growth:

$$\theta_{t+1} = A_t \tilde{f}_t(\theta_t, I_t) \tag{2.14}$$

where $A_t > 0$ is the TFP term and the $\tilde{f}_t(\theta_t, I_t)$ sub-function is a known location and scale (KLS) production technology. An example of this class of functions is the CES production technologies augmented with TFP dynamics:

$$\theta_{t+1} = A_t(\gamma_t \theta_t^{\phi_t} + (1 - \gamma_t) I_t^{\phi_t})^{1/\phi_t}$$

We first establish that our identification result for KLS production technologies fails in this case because we cannot separately identify the TFP parameter A_t from the measurement parameters. To see this, write the production technology in logs:

[14] See the Appendix for more discussion. For example, a CES technology with constant returns to scale implies that the elasticity of skill formation with respect to investment must be between 0 and 1, regardless of the data. This restriction can then bias downward the effects on skill formation of investment and interventions which increase investment.

$$\ln \theta_{t+1} = \ln A_t + \ln \tilde{f}_t(\theta_t, I_t)$$

Note that A_t is the scale of the production technology in levels, but $\ln A_t$ is the location of the production function in logs.

Next consider the following difference in the conditional expectations for a latent log skill measure m, $Z_{t+1,m}$:

$$E(Z_{t+1,m}|\ln \theta_t = a_1, \ln I_t = \ell_1) - E(Z_{t+1,m}|\ln \theta_t = a_2, \ln I_t = \ell_2)$$
$$= \mu_{t+1,m} + \lambda_{t+1,m}(\ln A_t + \ln \tilde{f}_t(e^{a_1}, e^{\ell_1})) - [\mu_{t+1,m} + \lambda_{t+1,m}(\ln A_t + \ln \tilde{f}_t(e^{a_2}, e^{\ell_2}))]$$
$$= \lambda_{t+1,m}(\ln \tilde{f}_t(e^{a_1}, e^{\ell_1}) - \ln \tilde{f}_t(e^{a_2}, e^{\ell_2}))$$

From this expression, it is clear that the TFP location $\ln A_t$ cannot be identified. Without further restrictions, the location of the production function (in logs) cannot be separately identified from the location of the measurement equations (which measure skills in logs) given by $\mu_{t,m}$ intercept.

Given the failure of identification for this more general technology, it is natural to ask what additional assumptions would be sufficient for identification. We show that if we have some auxiliary information on the relationship between measurement intercepts over time, then we can identify the A_t TFP terms in production functions of the form (2.14. We consider identification under the following assumption:

Assumption 2 *For some measures* $Z_{t+1,m} = \mu_{t+1,m} + \lambda_{t+1,m} \ln \theta_{t+1} + \epsilon_{t+1,m}$ *and* $Z_{t,m'} = \mu_{t,m'} + \lambda_{t,m'} \ln \theta_t + \epsilon_{t,m'}$*, we have* $\mu_{t+1,m} = g(\mu_{t,m'})$*, where* $g(\cdot)$ *is a known relationship.*

Whether Assumption 2 holds depends on the particular measures the researcher has available. We discuss the applicability of this assumption to our particular data and measures in our empirical application. This assumption could be justified if the measure in period $t+1$ and period t are *age-invariant* measures, as discussed below, where for example the measure is the same test given to children of different ages. In this case, it is plausible that the measures have the same location so that $\mu_{t+1,m} = \mu_{t,m}$. This assumption of age-invariant intercepts is of course sufficient but not necessary. And, to be clear, Assumption 2 does not require the researcher to assume any particular values for the measurement intercepts, but simply that they are related to each other in a known way.

We next present identification results which show that with Assumption 2, and the other assumptions previously used to prove Theorem 1, we can now identify skill development technologies of the form given in (2.14) which do not have a known location and scale:

Theorem 2 *Consider a production technology of the form* $f_t(\theta_t, I_t) = A_t\tilde{f}_t(\theta_t, I_t)$ *where* $\tilde{f}_t(\theta_t, I_t)$ *has known scale and location (Definition 1) and* $A_t \in \mathbb{R}_{++}$. *Under Assumption 1, Assumption 2, the full support assumption on some measure* $Z_{t+1,m}$ *and the conditions for Theorem 1, the technology* $\tilde{f}_t(\theta_t, I_t)$ *and* A_t *are separately identified.*

Proof.

The identification of $\tilde{f}_t(\theta_t, I_t)$ follows directly from Theorem 1, as the unknown A_t term is "differenced" away allowing identification of the $\tilde{f}_t(\theta_t, I_t)$ KLS sub-function. We identify the factor loading for the measure $\lambda_{t+1,m}$ as well because identification of this parameter does not depend on the A_t value. We then identify A_t from the mean of the measure of skills in period $t+1$, $E(Z_{t+1,m})$, and re-arranging for $\ln A_t$:

$$\ln A_t = \frac{E(Z_{t+1,m}) - (\mu_{t+1,m} + \lambda_{t+1,m} E(\ln \tilde{f}_t(\theta_t, I_t)))}{\lambda_{t+1,m}}$$

From Assumption 2, with $\mu_{t,m}$ known, we also identify the measurement intercept for $t+1$ from $\mu_{t+1,m} = g(\mu_{t,m})$.

■

From the proof we have some intuition for our result. As is common in the literature estimating TFP in a variety of contexts, TFP here is also identified by the *residual* growth in mean measured skills from period 0 to period 1 (scaled by $\lambda_{t+1,m}$ factor loading), netting out the growth due to period t inputs θ_t, I_t. Identification of the full sequence of production technologies then proceeds as above in a sequential fashion, and we identify the production function parameters, including the sequence of A_t TFP terms, for all periods. In the estimation sections below, we use this identification result constructively to develop an estimator for the TFP sequence.

2.5.2 Example

Next consider an example:

Example 3 *Log-Linear (Cobb-Douglas) Technology with TFP*

Return to the two period Cobb-Douglas example considered above (Example 1) but now add a scaling factor $A_0 > 0$:

$$\ln \theta_1 = \ln A_0 + (\gamma_0 \ln \theta_0 + (1 - \gamma_0) \ln I_0)$$

Assume the single period 1 measure $Z_{1,m}$ satisfies Assumption 2 and $\mu_{1,m} = g(\mu_{0,m})$ for some $m = 1, 2, 3$. We proceed as before to identify γ_0 from

$$\frac{E(Z_{1,m} | \ln \theta_0 = a, \ln I_0 = 0) - E(Z_{1,m} | \ln \theta_0 = 0, \ln I_0 = 0)}{E(Z_{1,m} | \ln \theta_0 = 1, \ln I_0 = 1) - E(Z_{1,m} | \ln \theta_0 = 0, \ln I_0 = 0)} = \frac{\ln A_0 + \gamma_0 a - \ln A_0}{\ln A_0 + 1 - \ln A_0}$$

where the $\ln A_0$ TFP terms drop out of the expression. As in Example 1, we can then solve for the γ_0 production function parameter.

The TFP term $\ln A_0$ is identified from

$$\ln A_0 = \frac{E(Z_{1,m}) - (\mu_{1,m} + \lambda_{1,m} E(\ln \tilde{f}_0(\theta_0, I_0)))}{\lambda_{1,m}}$$

$$= \frac{E(Z_{1,m}) - g(\mu_{0,m})}{\lambda_{1,m}}$$

because $E(\ln \tilde{f}_0(\theta_0, I_0)) = 0$ for this log-linear production function. $\ln A_0$ is identified from the growth in mean measured skills because $\mu_{0,m} = E(Z_{0,m})$ given the normalization of the initial conditions. Substituting, the TFP term is then

$$\ln A_0 = \frac{E(Z_{1,m}) - g(E(Z_{0,m}))}{\lambda_{1,m}}.$$

TFP is identified from the growth in mean skills between periods 0 and 1, scaled by the identified measurement factor loading for the period 1 measures, $\lambda_{1,m}$.

2.5.3 Identifying Production Technologies with Unknown Scale

We next consider a parallel problem to that of identifying the location (in logs) of the production technology considered above: identifying a production technology with an unknown scale. Consider the following production technology:

$$\theta_{t+1} = \widetilde{f}_t(\theta_t, I_t)^{\psi_t}, \tag{2.15}$$

where $\psi_t \in R^+$ is an unknown scaling parameter and $\widetilde{f}_t$ is a sub-function with known location and scale. Given the unknown scaling parameter, the technology described in (2.15) is not a known location and scale technology (Definition 1). An example of this type of production function is the following CES function with unknown scale:

$$\theta_{t+1} = (\gamma_t \theta_t^{\phi_t} + (1 - \gamma_t) I_t^{\phi_t})^{\psi_t / \phi_t} \tag{2.16}$$

As in the TFP case above, we cannot separately identify the scale parameter ψ_t from the measurement factor loading $\lambda_{t+1,m}$. We consider an auxiliary restriction on the factor loadings which would allow identification:

Assumption 3 *For some measures* $Z_{t+1,m} = \mu_{t+1,m} + \lambda_{t+1,m} \ln \theta_{t+1} + \epsilon_{t+1,m}$ *and* $Z_{t,m'} = \mu_{t,m'} + \lambda_{t,m'} \ln \theta_t + \epsilon_{t,m'}$*, we have* $\lambda_{t+1,m} = q(\lambda_{t,m'})$*, where* $q(\cdot)$ *is a known function.*

We show that with Assumption 3, together with the other assumptions previously used to prove Theorem 1, we can now identify skill development technologies of the form given in (2.15), which do not have a known scale:

Theorem 3 *Consider a production technology of the form* $f_t(\theta_t, I_t) = \tilde{f}_t(\theta_t, I_t)^{\psi_t}$*, where* $\tilde{f}_t(\theta_t, I_t)$ *has known scale and location (Definition 1) and* $\psi_t \in \mathbb{R}_{++}$*. Under Assumption 1, Assumption 3, the full support assumption on some measure* $Z_{t+1,m}$ *and the conditions for Theorem 1, the technology* $\tilde{f}_t(\theta_t, I_t)$ *and* ψ_t *are separately identified.*

Proof.

Identification of $\tilde{f}_t(\theta_t, I_t)$ follows directly from Theorem 1, as the unknown ψ_t term drops out allowing identification of the $\tilde{f}_t(\theta_t, I_t)$ KLS sub-function. To identify ψ_t, take the covariance between a measure of latent skills at age $t + 1$ and at age t:

$$\begin{aligned} Cov(Z_{t+1,m}, Z_{t,m}) &= \lambda_{t+1,m} \lambda_{t,m} Cov(\ln \theta_{t+1}, \ln \theta_t) \\ &= q(\lambda_{t,m}) \lambda_{t,m} \psi_t Cov(\ln \tilde{f}_t(\theta_t, I_t), \ln \theta_t) \end{aligned}$$

Given $\lambda_{t,m}$ is known and $Cov(\ln \tilde{f}_t(\theta_t, I_t), \ln \theta_t)$ can be computed from the identified sub-function $\tilde{f}_t(\theta_t, I_t)$, then we can re-arrange this expression to solve for ψ_t.

■

This proof mirrors the identification result for TFP. If we assume that the factor loading in the measurement equation (which provides the scale of the measure) has some known relationship with already identified factor loadings, then we can identify the scale of the production technology.

2.5.4 Age-Invariant Measures

We conclude this section with a discussion of measures which would satisfy these auxiliary assumptions. An extensive literature, principally in the field of psychometrics, is concerned with designing skill measures which can be "equated" across children of different ages so that the development of children can be tracked using a coherent single measure. These measures consist of tests which are designed to be applicable for children of various ages, and include a range of test items (questions) which show meaningful variation for both younger and older children. Tests such as the Peabody Individual Achievement Test (PIAT)and the Woodcock-Johnson tests are designed so that they include a range of questions of various difficulty levels. The simple raw scores on these tests, reflecting the total number of questions answered correctly, can then be interpreted as an *age-invariant* measure of skills. [15]

We formalize this notion of age-invariant measures in the following definition. A pair of measures are age-invariant if their measurement parameters are constant across child ages:

Definition 2 *A pair of measures $Z_{t,m}$ and $Z_{t+1,m}$ is* age-invariant *if $E(Z_{t,m}|\theta_t = p) = E(Z_{t+1,m}|\theta_{t+1} = p)$ for all $p \in \mathbb{R}_{++}$.*

[15] In practice, these types of age-invariant tests are often administered such that the questions are endogenously determined by the previous answers of the child. Therefore, while not all children are in fact answering the exact same test questions, their scores are determined in an age comparable way. The typical test includes a number of test items ranging from low difficulty to high difficulty questions. Testing begins by first establishing a baseline test item for each child. While the baseline is initially based on the child's age, the baseline adjusts downward (to less difficult questions) as the child is unable to answer questions correctly. Once the baseline is established, the test then progressively asks more difficult questions. Testing stops when the child makes a certain number of mistakes. The score is then determined as the number of correct answers before testing stops. Included in this number of correct answers are the lower difficulty test items prior to the baseline item because it is assumed the child would have answered these items correctly (given she was able to answer more the difficult items).

Age-invariant measures imply that two children of different ages t and $t+1$ would nonetheless have the same expected level of measured skill *if* the children have the same latent level of skill: $\theta_t = \theta_{t+1} = p$. [16] In this case, the younger child, aged t, could be considered "ahead" of her age group, and the older child, aged $t+1$, could be considered "behind" her age group. The age invariant measures $Z_{t,m}$ and $Z_{t+1,m}$ would report the same score (in expectation) for these two children. Definition 2 implies that for age-invariant measures both Assumption 2 and Assumption 3 hold, allowing identification of the technology with unknown TFP and unknown return to scale (see Theorem 3 (i)-(ii)). [17]

Finally note that whether a given pair of measures is age-invariant depends on the measures and must be evaluated on a case-by-case basis. Using pairs of unrelated measures, such as birth weight to measure cognitive skills at birth and SAT scores to measures skills at age 18, would not seem to constitute a set of age-invariant measures as there is no reason to believe these measures would have a common location and scale.

2.6 Estimation

In this section we discuss the empirical model we take to the data, the estimation algorithm we develop based on the identification analysis of the preceding sections, and briefly describe the data. Additional details about the data and sample are left for the Appendix.

[16] Age-invariant measures should not be confused with "age-standardized" measures, which are measures the researcher constructs to be mean 0 and standard deviation 1 at all ages for the particular sample at hand (See the Appendix). Our concept of age-invariant measures concerns the underlying primitive and unobserved parameters of the measurement equations. Age-standardized measures would in fact not represent any growth in average skills or changes in the dispersion of skills as children age.

[17] Age-invariance implies the following restrictions on measurement parameters: $\mu_{t+1} + \lambda_{t+1} \ln p = \mu_t + \lambda_t \ln p$ for all p. Re-arranging, we have $(\mu_{t+1} - \mu_t) = \ln p \, (\lambda_t - \lambda_{t+1})$ for all p. This is the case if and only if $\mu_t = \mu_{t+1}$ and $\lambda_t = \lambda_{t+1}$.

2.6.1 Empirical Model

There are five parts to the empirical model: 1) a model of skill development where skills in the next period are produced by the stocks of existing skills and parental investments; 2) a model of parental investment where investment depends on household characteristics and the existing stock of skills; 3) a distribution of initial conditions of household characteristics and child skills; 4) a model of the relationship between final childhood skills and adult outcomes (schooling and earnings); and 5) a measurement model relating each of the latent model elements to observed data measures. Besides specifying particular functional forms for the production technology, the major distinction between the empirical model and the preceding identification analysis is that we assume parental investment is also measured with error and allow parental investment to be endogenously related to the stock of existing children's skill.

The timing of the model is as follows. There are five biannual periods of child development: ages 5-6 ($t = 0$), 7-8 ($t = 1$), 9-10 ($t = 2$), 11-12 ($t = 3$), 13-14 ($t = 4$). While it would be ideal to extend the model to even earlier ages (to birth or even to pre-natal periods), we face the common tradeoff of assuming "too much" relative to the data we have available. We have chosen here to focus on the childhood period from age 5 to 14 where we have more skill measures, and plausibly age-invariant measures, and can judge the performance of the model and estimator in closer to ideal conditions.

Skill Production Technology

At each age t the current level of latent cognitive skills and investment produce the next period's $(t + 1)$ skills. The technology takes a stochastic translog specification:

$$\ln \theta_{t+1} = \ln A_t + \gamma_{1,t} \ln \theta_t + \gamma_{2,t} \ln I_t + \gamma_{3,t} \ln I_t \cdot \ln \theta_t + \eta_{\theta,t}, \tag{2.17}$$

where $\ln A_t$ is the TFP term, and $\eta_{\theta,t}$ is the stochastic production shock, which is assumed i.i.d. $\sim N(0, \sigma^2_{\theta,t})$ for all t and independent of the current stock of skills and investment. The translog specification is a generalization of the Cobb-Douglas specification, where the special case $\gamma_{3,t} = 0$ is the typical Cobb-Douglas specification (with the addition of a TFP term and a stochastic shock). We use the translog specification because of its flexibility relative to the Cobb-Douglas and other CES functions. The translog function allows a non-constant elasticity of substitution between inputs and can be expanded with the inclusion of additional terms to a close provide an approximation of any unknown production technology. The log-linear form of the function also facilitates convenient and fast closed form estimators, as detailed below. Our general translog function also allows non-constant returns to scale. With $\gamma_3 \neq 0$, the elasticity of skill production with respect to investment depends on the current level of children's skills:

$$\frac{\partial \ln \theta_{t+1}}{\partial \ln I_t} = \gamma_{2,t} + \gamma_{3,t} \ln \theta_t,$$

where $\gamma_{3,t} > 0$ implies a higher return to investment for children with currently high levels of skill than for children with low levels of skill, a dynamic complementarity where past skills (and past investments which produced those skills) affect the productivity of current investments. Moreover, $\gamma_{3,t} \neq 0$ implies that the elasticity of next period skills with respect to investment is a function of the child's stock of skills.

Parental Investment

We specify a parametric policy function for parental investment. Investment is endogenously determined by the current stock of the child's skills, mother's skills, and family income:

$$\ln I_t = \alpha_{1,t} \ln \theta_t + \alpha_{2,t} \ln \theta_{MC} + \alpha_{3,t} \ln \theta_{MN} + \alpha_{4,t} \ln Y_t + \eta_{I,t} \tag{2.18}$$

where $\sum_j \alpha_{j,t} = 1$ for all t, θ_{MC} is the mother's stock of cognitive skills, θ_{MN} is the mother's stock of non-cognitive skills, Y_t is household income, and $\eta_{I,t}$ is the investment shock, where $\eta_{I,t}$ i.i.d. $\sim N(0, \sigma^2_{I,t})$ for all t and independent of latent skills and income. Our concept of investment represents both quantity and quality aspects, where we use measures of investments which capture quantity aspects of investment (time parents spent reading to children) and quality aspects (whether children are "praised" by their parents).

This specification of investment is a kind of "reduced form" specification representing a policy function for parental investment which is not derived from an explicit economic model of the household behavior. This approach follows Cunha *et al.* (2010); Attanasio *et al.* (2015a,b). The advantages of this approach are twofold. First, this approach provides a simple and tractable model of the investment process which avoids the computational burden of solving and estimating a formal model of household behavior. Second, this approach has the potential to allow for some generality as our specification of the investment process can be consistent with multiple models of the households. Other work derives parental investment from explicit models of the household, including explicit representations of household preferences, decision making, beliefs, and constraints (see for example Del Boca *et al.*, 2014b, 2016; Cunha, 2013a; Cunha *et al.*, 2013; Bernal, 2008). The advantage of these latter approaches is that the counterfactual policy analysis incorporates well defined household responses to policy, see Del Boca *et al.* (2016) for some discussion.

Given the investment function does not derive from an explicit model, the interpretation of the parameters is in some sense speculative. $\alpha_{1,t}$ can be interpreted as reflecting whether parents "reinforce" existing skill stocks ($\alpha_{1,t} > 0$) or "compensate" for low skill stocks ($\alpha_{1,t} < 0$). $\alpha_{2,t}$ and $\alpha_{3,t}$ reflect the extent to which the mother's skills relate to the quantity and quality of her parental investment as in the case where more skilled mothers read to their children more or provide higher quality interactions. Finally, $\alpha_{4,t}$ reflects the influences that household resources have on the extent of parental investments, and reflects the combined effects of constraints the household faces (such as credit market constraints) and preferences the household has to invest scarce resources in children (see Caucutt *et al.*, 2015).

Finally, to close the investment model, we assume that log-family income ($\ln Y_t$) follows an AR(1) process which allows for life-cycle trends in income:

$$\ln Y_{t+1} = \mu_Y + \delta_Y \cdot t + \rho_Y \ln Y_t + \eta_{Y,t} \tag{2.19}$$

where the innovation is $\eta_{Y,t}$ i.i.d. $\sim N(0, \sigma_Y^2)$ and is assumed independent of all latent variables. Initial family income Y_0 is allowed to be correlated with mother's and children's initial skills, and hence our model captures important correlations between household resources and the skills of parents and children.

Initial Conditions

The initial conditions consist of the child's initial (at age 5-6) stock of skills $\theta_{C,0}$, the mother's cognitive and non-cognitive skills (θ_{MC} and θ_{MN}), which are assumed to be time invariant over the child development period, and the level of family income at birth (Y_0). Define the vector of initial conditions as

$$\Omega = (\ln \theta_0, \ln \theta_{MC}, \ln \theta_{MN}, \ln Y_0)$$

We assume a parametric distribution for the initial conditions:

$$\Omega \sim N(\mu_\Omega, \Sigma_\Omega)$$

where $\mu_\Omega = [0, 0, 0, 0, \mu_{0,\ln Y}]$. $\mu_{0,\ln Y}$ is the mean of the family log income when children are 5-6 years old. The means of the remaining variables are set to zero by Normalization 1. Σ_Ω is the variance-covariance matrix for the initial conditions.

Adult Outcome

In order to provide a more meaningful metric to evaluate policy interventions in our model, we relate adult outcomes to the stock of children's skills in the final period of the child development process (period $T = 4$ or age 13-14):

$$Q = \mu_Q + \alpha_Q \ln \theta_T + \eta_Q, \tag{2.20}$$

where η_Q is independent of $\ln \theta_T$. We use years of schooling measured at age 23 and log earnings at age 29 as adult outcomes. Schooling is an attractive adult outcome to use because it explains a large fraction of adult earnings and consumptions, is largely determined at an early point in adulthood and, unlike realized labor market earnings, does not suffer from a censoring issue due to endogenous labor supply.

Measurement

The final piece of our model is the model of measurement relating latent variables to observed data. Children's skills, parental investment, and mother's skills are all assumed to be measured with error. There are 4 latent variables: $\omega \in \{\theta, \theta_{MC}, \theta_{MN}, I\}$. There are in general multiple measures for each latent variable. Each measure is assumed to take the following form:

$$Z_{\omega,t,m} = \mu_{\omega,t,m} + \lambda_{\omega,t,m} \ln \omega_t + \epsilon_{\omega,t,m}$$

where m indexes the measures for each latent variable $\omega \in \{\theta, \theta_{MC}, \theta_{MN}, I\}$.

We assume a generalized version of Assumption 1 appropriate for this more general empirical model. All measurement errors are assumed independent of each other (across measures and over time), and all measurement errors are assumed independent of the latent variables, household income, and the "structural" shocks $(\eta_{I,t}, \eta_{\theta,t}, \eta_Q)$. This assumptions is strong, and weaker assumptions of mean-independence are sufficient for identification of the parametric model. While we assume strong independence assumption, we make no other restrictions on the distribution of measurement error (e.g. we do not assume $\epsilon_{\omega,t,m}$ is distributed Normal) as is common in previous approaches. Our sequential estimator, described below, is therefore robust to mis-specification of the marginal distributions of measurement errors.

2.6.2 Estimation Algorithm

Our estimation algorithm is formed from the identification results presented above, and in particular relies on the error-in-variable formulation from Section 2.4.6. Before describing the steps of the algorithm, consider several estimation options. One approach, a kind of "brute force" approach, is to simulate the full sequence of latent variables and measures from candidate primitive parameters and explicit assumptions about the distribution of measurement errors (e.g. assume they are Normally distributed) and compute a likelihood function or a set of moments to form the basis of an estimator. We do not prefer this approach because it requires additional assumptions about the distribution of measurement errors which are not required for identification. This approach may also involve a tremendous amount of computationally costly simulation given the non-linear nature of the model.

A second estimation approach is to use the measures directly to simulate the distribution of latent variables by assuming a particular distribution for the latent variables. One then could estimate the production function in a second step from the simulated distribution of latent variables. This is the approach of Cunha *et al.* (2010) and Attanasio *et al.* (2015a,b) in which both assume the latent variables are distributed according to a mixture of 2 Normal distributions. This approach too makes specific parametric assumptions which are not required.

Our estimation approach directly follows our identification approach in treating the measurement parameters as nuisance parameters which can be computed sequentially along with the primitive parameters of the model generating the latent variables. Following the estimation of the initial conditions using standard techniques, we sequentially estimate for each age the investment and production functions, followed by the measurement parameters for the measures used for that age. The sequential algorithm we develop has the advantage of tractability because our estimator does not require the simulation of the full model; the primitives of the production technology and investment functions can be estimated directly from data. In addition, another advantage of our approach over a joint estimation approach is by breaking the estimator into steps, we make the identification assumptions as transparent as possible. Of course, the disadvantage of our approach is a potential loss of efficiency from not estimating the parameters jointly and exploiting "cross-step" restrictions.

We present two versions of the estimation algorithm. The first version works with a unrestricted version of the technology:

Model 1 (General): $\ln A_t$ free and $\sum_{j=1}^{3} \gamma_{j,t}$ free. At least one measure is age-invariant.

The availability of this age-invariant measure allows us to identify the more general technology.

The second version of the model restricts the production technology (2.17) to have a known location and scale.

Model 2 (Restricted): $\ln A_t = 0$ for all t (no TFP dynamics) and $\sum_{j=1}^{3} \gamma_{j,t} = 1$ for all t (constant returns to scale).

Estimation of Model 2 (Restricted)

We begin with the estimator for the second version of the model, using the restricted technology. The estimator for the more general technology (Model 2) is below.

Step 0 (Estimate Initial Conditions and Initial Measurement Parameter)

First, we estimate the measurement parameters at the initial period (age 5-6), $\lambda_{\omega,0,m}, \mu_{\omega,0,m}$ for all measures m, for both children's and mother's skills. To estimate these measurement parameters, we use ratios of covariances and measurement means as outlined above (2.3) and (2.4). We choose one measure for children's cognitive skills, mother's cognitive skills, and mother's non-cognitive skills as the normalizing measure (which we label $m = 1$, without loss of generality) and normalize the factor loading for this measure to be 1: $\lambda_{\theta,0,1} = 1$, $\lambda_{MC,0,1} = 1$, $\lambda_{MN,0,1} = 1$. [18] We estimate the remaining factor loadings using the average of the covariances between all of the remaining measures, where each factor loading is computed from

$$\lambda_{\omega,0,m} = \frac{Cov(Z_{\omega,0,m}, Z_{\omega,0,m'})}{Cov(Z_{0,\omega,1}, Z_{\omega,0,m'})} \; \forall m \neq m' \text{ and } \forall \, \omega \in \{\theta, MC, MN\}.$$

Given the normalization that log skills are mean 0 in the initial period, we compute the initial measurement intercepts as

[18] Note that while investment is a latent variable as well, we do not need to normalize the scale and location of latent investment because investment already has a scale and location specified by the KLS investment equation (2.18).

$$\mu_{\omega,0,m} = E(Z_{\omega,0,m}) \; \forall m \text{ and } \forall \, \omega \in \{\theta, MC, MN\}$$

With the factor loading estimates in hand, we then estimate the initial period variance-covariance matrix Σ_Ω using variances and covariances in measures of skills and family income (assumed measured without error). This step provides estimates of the initial joint distribution of children's skills, mother's skills, and family income. In this initial step, we also estimate the parameters of the income process (2.19) using a regression of income on lagged income and a time trend.

Finally, given the estimates of the measurement parameters for children and mother skills, we form the following "residual" measures:

$$\widetilde{Z}_{\omega,0,m} = \frac{Z_{\omega,0,m} - \mu_{\omega,0,m}}{\lambda_{\omega,0,m}} \; \forall m \text{ and } \forall \, \omega \in \{\theta, MC, MN\}$$

We are now ready to estimate the investment function for period $t = 0$, where the investment in this first period depends on the initial child's skills and household characteristics (mother's skills and family income).

Step 1 (Estimate Investment Function Parameters):

Following the errors-in-variables formulation described above, substitute a "raw" measure for investment $Z_{I,0,m}$ and a "residual" measure for each of the latent skills ($\widetilde{Z}_{\theta,0,m}$, $\widetilde{Z}_{MC,0,m}$, $\widetilde{Z}_{MN,0,m}$) into the model of investment defined in terms of primitives (2.18):

$$\frac{Z_{I,0,m} - \mu_{I,0,m} - \epsilon_{I,0,m}}{\lambda_{I,0,m}} = \alpha_{1,0}(\widetilde{Z}_{\theta,0,m} - \widetilde{\epsilon}_{\theta,0,m}) + \alpha_{2,0}(\widetilde{Z}_{MC,m} - \widetilde{\epsilon}_{MC,m})$$
$$+\alpha_{3,0}(\widetilde{Z}_{MN,m} - \widetilde{\epsilon}_{MN,m}) + \alpha_{4,0} \ln Y_0 + \eta_{I,0}$$

Re-arranging, we have

$$
\begin{aligned}
Z_{I,0,m} &= \mu_{I,0,m} + \lambda_{I,0,m}\alpha_{1,0}\widetilde{Z}_{\theta,0,m} + \lambda_{I,0,m}\alpha_{2,0}\widetilde{Z}_{MC,m} + \lambda_{I,0,m}\alpha_{3,0}\widetilde{Z}_{MN,m} + \lambda_{I,0,m}\alpha_{4,0}\ln Y_0 \\
&+ \epsilon_{I,0,m} + \lambda_{I,0,m}(\eta_{I,0} - \widetilde{\epsilon}_{\theta,0,m} - \widetilde{\epsilon}_{MC,m} - \widetilde{\epsilon}_{MN,m}) \\
&= \beta_{0,0,m} + \beta_{1,0,m}\widetilde{Z}_{\theta,0,m} + \beta_{2,0,m}\widetilde{Z}_{MC,m} + \beta_{3,0,m}\widetilde{Z}_{MN,m} + \beta_{4,0,m}\ln Y_0 + \pi_{I,0,m} \quad (2.21)
\end{aligned}
$$

where $\beta_{j,0,m} = \lambda_{I,0,m}\alpha_{j,0}$ for all j and

$$\pi_{I,0,m} = \epsilon_{I,0,m} + \lambda_{I,0,m}(\eta_{I,0} - \alpha_{1,0}\widetilde{\epsilon}_{\theta,0,m} - \alpha_{2,0}\widetilde{\epsilon}_{MC,m} - \alpha_{3,0}\widetilde{\epsilon}_{MN,m}).$$

Estimation of (2.21) by OLS would yield inconsistent estimates of the $\beta_{j,0,m}$ coefficients because the measures are correlated with their measurement errors (included in the residual term $\pi_{I,0,m}$). Here the structure of the model affords the researcher several possible strategies to consistently estimate the $\beta_{j,0,m}$ coefficients. We use an instrumental variable estimator with the vector of excluded instruments composed of alternative measures of skills: $[Z_{\theta,0,m'}\ ,\ Z_{MC,0,m'}\ ,\ Z_{NC,0,m'}]$. Under Assumption 1, these instruments are valid because each of these alternative measures is uncorrelated with all of the components of $\pi_{I,0,m}$. Using this IV strategy, we obtain consistent estimators for the $\beta_{j,t,m}$ coefficients. The primitive parameters of the investment function are then recovered from

$$\alpha_{j,0} = \frac{\beta_{j,0,m}}{\sum_{j=1}^{4}\beta_{j,0,m}}$$

Step 2 (Compute Measurement Parameters for Latent Investment):

After estimating the primitive parameters of the investment function, we recover the scale and location for the investment equation without further re-normalizations on the measurement equation parameters. The intercept and factor loading for the investment measure are given by

$$\mu_{I,0,m} = \beta_{0,0,m}$$

and

$$\lambda_{I,0,m} = \sum_{j=1}^{4} \beta_{j,0,m}$$

With these consistent estimators for the measurement parameters for investment, we form the "residual" measures for investment in period $t = 0$:

$$\widetilde{Z}_{I,0,m} = \frac{Z_{I,0,m} - \mu_{I,0,m}}{\lambda_{I,0,m}}$$

Step 3 (Estimate Skill Production Technology)

Next, we use a similar technique to estimate the production technology. Substituting the residual measures into the production technology (2.17), we have

$$\frac{Z_{\theta,1,m} - \mu_{\theta,1,m} - \epsilon_{\theta,1,m}}{\lambda_{\theta,1,m}} = \gamma_{1,0}(\widetilde{Z}_{\theta,0,m} - \widetilde{\epsilon}_{\theta,0,m}) + \gamma_{2,0}(\widetilde{Z}_{I,0,m} - \widetilde{\epsilon}_{I,0,m}) + \gamma_{3,0}(\widetilde{Z}_{\theta,0,m} - \widetilde{\epsilon}_{\theta,0,m})(\widetilde{Z}_{I,0,m} - \widetilde{\epsilon}_{I,0,m}) + \eta_{\theta,0}$$

With some algebra, we can re-write this as:

$$Z_{\theta,1,m} = \delta_{0,0,m} + \delta_{1,0,m}\widetilde{Z}_{\theta,0,m} + \delta_{2,0,m}\widetilde{Z}_{I,0,m} + \delta_{3,0,m}\widetilde{Z}_{\theta,0,m} \cdot \widetilde{Z}_{I,0,m} + \pi_{\theta,0,m} \tag{2.22}$$

where $\delta_{0,0,m} = \mu_{\theta,0,m}$, $\delta_{j,0,m} = \lambda_{\theta,1,m}\gamma_{j,0}$ for $j = 1, 2, 3$ and

$$\pi_{\theta,0,m} = \epsilon_{\theta,1,m} + \lambda_{\theta,1,m}[\eta_{\theta,0} - \gamma_{1,0}\epsilon_{\theta,0,m} - \gamma_{2,0}\epsilon_{I,0,m} - \gamma_{3,0}(\widetilde{Z}_{\theta,0,m}\epsilon_{I,0,m} + \widetilde{Z}_{I,0,m}\epsilon_{\theta,0,m} - \epsilon_{\theta,0,m}\epsilon_{I,0,m})]$$

As with the investment function, estimation of 2.22 using OLS would lead to inconsistent estimates. We use the same IV approach as above using instruments formed from alternative measures $[Z_{\theta,0,m'}\ ,\ Z_{I,0,m'}\ ,\ Z_{\theta,0,m'} \cdot Z_{I,0,m'}]$. Under Assumption 1 these instruments are uncorrelated the residual error term $\pi_{\theta,0,m}$. [19] With consistent estimates of $\delta_{j,0,m}$s in hand, we can then recover the structural parameters and for the production technology as:

$$\gamma_{j,0} = \frac{\delta_{j,0,m}}{\sum_{j=1}^{3} \delta_{j,0,m}} \ \forall\ j \in \{1,2,3\}$$

Step 4 (Compute Measurement Parameters for Latent Skill):

The measurement parameters for the latent skill measure in period $t = 1$ ($Z_{\theta,1,m}$) can then be recovered from

$$\mu_{\theta,1,m} = \delta_{0,0,m},$$

$$\lambda_{\theta,1,m} = \sum_{j=1}^{3} \delta_{j,0,0}.$$

We then form the residual measure for latent skill as

$$\widetilde{Z}_{\theta,1,m} = \frac{Z_{\theta,I,m} - \mu_{\theta,1,m}}{\lambda_{\theta,1,m}}$$

Step 5 (Estimate variance of Investment and Production Function Shocks):

[19]Perhaps the less obvious terms are terms such as this $E(\widetilde{Z}_{\theta,0,m}\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'})$. Under the assumption of independence of the errors, we have

$$E(\widetilde{Z}_{\theta,0,m}\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'}) = E(\widetilde{Z}_{\theta,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'})E(\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'})$$

given $\epsilon_{I,0,m}$ is independent of $\widetilde{Z}_{\theta,0,m}$. Given the independence assumption, the latter term is $E(\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'}) = E(\epsilon_{I,0,m}) = 0$. Therefore, $E(\widetilde{Z}_{\theta,0,m}\epsilon_{I,0,m}|Z_{\theta,0,m'} \cdot Z_{I,0,m'}) = 0$.

The remaining parameters to be estimated for this period are the variances of the investment and production function shocks, $\sigma^2_{I,0}$ and $\sigma^2_{\theta,0}$. To estimate $\sigma_{I,0,}$ we use the covariance between the residual from (2.21), $\pi_{I,0,m}$ and an alternative residual measure of investment $\widetilde{Z}_{I,0,m'} = \ln I_0 + \epsilon_{I,0,m'}$:

$$Cov(\pi_{I,0,m}/\lambda_{I,0,m}, \widetilde{Z}_{I,0,m'}) = V(\eta_{I,0}) = \sigma^2_{I,0}$$

To compute the residual measure $\widetilde{Z}_{I,0,m}$ we need to compute the measurement parameters for this measure. We do this by repeating the estimation in Steps 2 and 3 replacing the left-hand side variable in (2.21) with the alternative measure $Z_{I,0,m'}$.

The variance of the production shock is estimated in the same way using an alternative measure of children's skills in period $t = 1$:

$$Cov(\pi_{\theta,1,m}/\lambda_{I,1,m}, \widetilde{Z}_{\theta,1,m'}) = V(\eta_{\theta,0}) = \sigma^2_{\theta,0}$$

Remaining Steps

We repeat Steps 1-5 for the remaining periods until the final period of child development T. This algorithm produces estimates of the parameters of the investment and production functions for all child ages.

Estimation of Model 1 (Unrestricted)

The preceding algorithm restricted the production technology to have no TFP dynamics and constant returns to scale (Model 2). Following Theorem 2 and Theorem 3, identification of the more general model can be accomplished with restrictions on the measurement parameters. We assume we have available at least one child skill measure which is age-invariant (Definition 2). Label the age-invariant measure to be measure m, and for this measure we have $\mu_{\theta,t,m} = \mu_{\theta,0,m}$ for all t and $\lambda_{\theta,t,m} = \lambda_{\theta,0,m}$ for all t.

With this age invariant measure, we repeat Step 3 (Estimate Production Technology). The "reduced form" equation (2.22) and estimation of the $\delta_{j,0,m}$ parameters remains the same. To allow for non-constant returns to scale we do not restrict the structural $\gamma_{j,0}$ parameters to sum to 1. The structural parameters are computed as

$$\gamma_{j,0} = \frac{\delta_{j,0,m}}{\lambda_{\theta,1,m}} \;\forall\, j \in \{1, 2, 3\}$$

With the inclusion of the TFP term $\ln A_0$, the $\delta_{0,0,m}$ intercept from the reduced form equation (2.22) is now

$$\delta_{0,0,m} = \mu_{\theta,1,m} + \lambda_{\theta,1,m} \ln A_0$$

Given the age-invariance assumption, we can consistently estimate $\mu_{\theta,1,m}$ and $\lambda_{\theta,1,m}$ and compute $\ln A_0$.

With the addition of these computations to Step 3, the other steps in the algorithm remain the same. We can use this extended to algorithm to compute the full sequence of parameters for the investment and production functions for all child ages.

Estimating the Adult Outcome Equation

Finally, after we have computed the full path of primitive parameters for the investment and production functions, we are able to estimate the adult outcome process (2.20). We focus on both final years of education at age 23 and log earnings at age 30. We use the same IV method as before to solve the measurement error issue. Substituting the measures for skills at age 13-14 ($t = 4$) in equation (2.20), we have:

$$Q = \mu_Q + \alpha_Q \widetilde{Z}_{\theta,4,m} + (\eta_Q - \alpha_Q \widetilde{\epsilon}_{\theta,4,m}) \tag{2.23}$$

We use a second measure for skills at age 13-14 as an IV to identify α_Q.

2.6.3 Data

We estimate the model using information about children and their families obtained from the National Longitudinal Study of Youth 1979 (NLSY). Descriptive statistics for the sample and additional data construction details are left for the Appendix.

The NLSY dataset is constructed by matching female respondents of the original dataset with their children who were part of the Children and Young Adults surveys, from 1986 to 2012. The dataset provides observations of the first period of the model (age 5-6) through adulthood. The total number of children in our sample is 11,509.

The NLSY dataset contains multiple measures of children's skills, mother's skills, and parental investments. The complete set of measures, their ranges and descriptive statistics for our sample are included in the Appendix. For children's skills we rely on different sub-scales of the Peabody Individual Achievement Test (PIAT) in Mathematics, Reading and Recognition, and the Peabody Picture Vocabulary Test (PPVT). Finally, we use information for children when they become young adults to link the children skills into a more meaningful metric to evaluate policy intervention: we use children's highest grade completed at age 23 or older and their earnings at age 29. The information about the educational attainment is measured as the highest grade completed as of date of last interview. We considered schooling information only for those young adults who were at least 23 years old or older in the last 2012 interview. Age 29 earnings is in real 2012 dollars.

For mother's cognitive skills we use sub-scales of the Armed Services Vocational Aptitude Battery (ASVAB), and for mother's non-cognitive skills we use the Rotter and Rosenberg indexes. For parental investments, we use the various HOME score measures from direct observation and interview with the mother. Family income includes all sources of income for the parents, including mother's and father's labor income, and any sources of non-labor income.

2.7 Results

In this section we discuss our parameter estimates, simulate the estimated model to describe the development of children's skills, and compute the effects of simple interventions to improve skills and adult outcomes. We begin by presenting estimates of the general model in which we allow for Total Factor Productivity (TFP) dynamics and non-constant returns to scale (Model 1). Because this general technology no longer has a known location and scale, we use an age-invariance restriction for identification. Given the structure of the PIAT tests, which administer the same test to children of various ages (given their ability level), we believe it is appropriate to assume the measurement intercepts and factor loadings for these measures of cognitive skills are age-invariant (Definition 2). Note that we do not assume any particular values for these measurement parameters, only the age invariance of them, and treat the measurement parameters as free parameters to be estimated.

We also consider results using alternative restricted models, and estimates which do not correct for measurement error and treat the measures as error free measures. We briefly discuss the policy predictions of these models below, but, for brevity, we report estimates of these several alternative models in the Appendix.

Another key issue involves interpreting the magnitude of the parameter estimates. Because the parameter estimates of the production technology and investment equations are relative to the initial skill normalizations, the magnitudes of many of the parameters estimates are not directly interpretable in isolation. We conclude this section with a series of policy counter-factual experiments using the estimated model.

These exercises provide necessary metrics to interpret the estimates with respect to adult outcomes, years of completed schooling at age 23 and earnings at age 29.

2.7.1 Parameter Estimates

Initial Conditions

Table 2.2 reports estimates of the initial conditions variance-covariance matrix Σ_Ω and the associated correlation matrix. We normalize children's cognitive skills to the PIAT-Mathematics test, mother's cognitive skills to the ASVAB2 (Arithmetics reasoning) and mother's non-cognitive skills to the Self-Esteem1 (Rosenberg Self-Esteem: "I am a person of worth") measure. The variances and covariances of the latent skills, and the investment and production function parameters, are interpreted relative to these normalizations. As expected, we estimate that children's skills, mother's cognitive and non-cognitive skills, and family income are all highly positively correlated. For space considerations, estimates of the dynamic family income process can be found in the Appendix.

Investment Function

Table 2.3 reports the estimates of the investment function specified in Section 2.6.1. At ages 5-6, we find that investment is increasing in children's skills, mother's skills, and family income. Because of the log-log form of the investment equation, we can interpret parameter estimates as elasticities. The parameter estimate of 0.230 on the log children's skills variable indicates that a 1 percent increase in children's skills raises investment by 0.23 percent, an inelastic response. The positive coefficient suggests that parents are "reinforcing" existing skills with further investments: children with higher skills are receiving even more investment than children with lower skills. Mother's cognitive skills and non-cognitive skills also increase investment at ages 5-

6, with non-cognitive skills of the mother estimated to have a substantially higher elasticity than cognitive skills. These coefficients indicate that mothers with higher skills are providing higher quantities and qualities of investments in children. Turning to the importance of income to parental investments, we find that a 1 percent increase in family income raises investment by 0.34 percent. The response of investment with respect to mother's skills and family income reflects the combination of parental preferences and household constraints, which we cannot unfortunately separately distinguish using this reduced form model of investment. Given that positive correlation between mother's initial skills, child's initial skills, and household income, taken together, these estimates of the investment function indicate that endogenous investment increases inequality in children's skills. The estimated variance on the investment shock reveals how much of the remaining variation in parental investments remains unexplained by this model, such as investments from schools, peers, and the child herself.

Comparing parameter estimates of the investment function over the development period reveals that the influence of the child's prior skills on investments becomes much smaller at later ages, indicating that parental investments are less reinforcing of existing skill stocks at older ages. As the child develops, we find that mother's non-cognitive skills becomes the dominant influence on investment. However, while the importance of family income falls somewhat from an elasticity of 0.34 at age 5-6 to 0.275 at age 11-12, income is still a significant and positive factor for parental investment even at later ages.

Production Function

Table 2.4 reports the parameter estimates for the technology of skill formation, as described in Section 2.6.1. We present measurement error corrected estimates of the two versions of the model: our preferred unrestricted Model 1 and the restricted Model 2. We turn first to the unrestricted Model 1 estimates.

At all ages, we find that skills are "self-productive" (next period's skills are increasing in existing skill stocks) and that skills are positively increasing in investment. For age 5-6 skill production, we estimate a statistically significant from 0 negative coefficient on the interaction term ($\ln \theta_t \ln I_t$) indicating that we reject the Cobb-Douglas special case.

The elasticities of skill production with respect to investment are heterogeneous, and we graph the skill elasticity for the age 5-6 production function in Figure 2.1 with respect to the existing stock of children's skill. The estimated negative coefficient on the interaction term indicates that the elasticity of skill production with respect to investment is decreasing in the child's current skill level. For low skill children, the elasticity approaches 1.4, indicating a that 1 percent increase in investment increases next period's skills by 1.4 percent. For already high skill children, the elasticity approaches 0.2, indicating that a 1 percent increase in investment raises future skills by only 0.2 percent. These heterogeneous investment elasticities suggest that targeting interventions to improve children's skills would have the largest effect on skill disadvantage children. This estimate stands in contrast to the estimates reported in Cunha *et al.* (2010). They estimate a CES technology which implies that the marginal productivity of investment is *higher* for high skill children given that current investments and the current stock of skills are complements. Note also that unlike the constant returns to scale CES case, our unrestricted model allows investment elasticities to be larger than 1, and we estimate, at least for some children, an elastic response of skill formation to investment.

The high TFP estimate for age 5-6 and the increasing returns to scale (indicated by the sum of the coefficients being greater than 1) indicate that existing skills and investments at this initial age are very productive relative to later ages. These estimates of high returns to early investment will underlie the policy experiment results we discuss next. As children age, Table 2.4 indicates that skills and investment become generally less productive and skills less "malleable." We graph the estimated TFP at each age in Figure 2.2. Our estimate of TFP at age 11-12 falls to 1/6 the level at age 5-6, indicating a dramatic slowdown in the productivity of existing skills and investments in producing new skills. This feature of the technology is largely consistent with the evidence that cognitive skills are difficult to change as children after age 10.

Comparing these estimates for the unrestricted Model 1 to the restricted Model 2 in Table 2.4 reveals that we clearly reject the restricted technology of Model 2. The estimated sum of the input coefficients far exceeds 1, with the estimated return to scale of 2.66 in the early period indicating increasing returns to scale. The estimated return to scale declines with the child's age to a value of 1.3 at older ages, revealing that even for older children we can reject constant returns to scale. In addition, the estimate of high positive TFP term also indicates that we clearly reject the assumption of a 0 log TFP in Model 2. As discussed below, these differences in production function estimates imply very different investment and policy effects, with the restricted Model 2 estimates implying a much smaller effect of an income transfer on children's skill development than in our preferred unrestricted model.

Adult Outcomes

Table 2.5 presents our estimates of the completed schooling outcome equation and log earnings equation. We estimate that a percentage change in children skills at age 13-14 leads to an increase of 0.086 years of school. We also find that a 1 percentage change in children skills leads to a 0.021 percentage change in earnings at age 29.

Below, we use these estimates to "anchor" our policy estimates to a meaningful adult outcome metric.

2.7.2 Estimated Child Development Path

We analyze the quantitative implications of the estimated model by simulating the dynamic model. Simulation of the model proceeds by drawing 100,000 children from the estimated initial conditions distribution and, for each child, forward simulating the path of income, investments, children's skills, and adult outcomes.

Figure 2.3 shows the estimated development path of mean log latent cognitive skills. Figures 2.4 and 2.5 show the dynamics in the distribution of latent skills. And, Figure 2.6 provides the estimated dynamics in the distribution of latent investment.

Perhaps not surprisingly, we find that children's mean latent skills grow substantially over this development period, from age 5 to 14, with the most rapid growth at early ages and growth slowing somewhat in the later period. In addition to growth in mean skills, we estimate that the latent distribution of cognitive skills becomes more dispersed as children age. Inequality rises substantially as there are different rates of skills growth for children at different percentiles of the initial skill distribution. Figure 2.5 shows that skills for high skill children at the 90^{th} percentile grow 20% from age 5-6 to age 9-10 and grow 9% during the rest of the childhood. For low initial skill children at the 5^{th} percentile, growth is slower, with a 6 % growth rate from age 5-6 to age 7-8 and a 3 % growth rate from age 11-12 to age 13-14.

2.7.3 Policy Experiments

In this section, we explore implications of the estimated model by using the estimated model to predict the effect of income transfers on childhood skill development and adult outcomes. While we do not have a fully developed model of household resource allocation to provide a more realistic setting to evaluate these policy, we argue that the experiments do at the very least provide a meaningful metric to understand the magnitude of the parameter estimates, and allow us to meaningfully compare the importance of various model features such as measurement error and the specification of general technologies.

Short and Long-Term Effects

Before we analyze the results for our particular parameter estimates, we first present a brief discussion of the effects of income transfers in our model. To allow for the possibility that an income transfer could have heterogeneous effects across households, we examine policy effects conditional on a vector of current state variables $\Omega_t = [\theta_t, \theta_{MC}, \theta_{MN}, Y_t]$, which includes the child's initial skills, the mother's skills, and initial family income. First, consider the expected *short-term* marginal effect of an increase in household income Y_t on the log of childhood skills in period $t+1$:

$$\begin{aligned}\Delta_{t+1,t}(\Omega_t) &= \frac{\partial \ln \theta_{t+1}}{\partial Y_t} \\ &= \frac{\partial \ln I_t}{\partial Y_t} \frac{\partial \ln \theta_{t+1}}{\partial \ln I_t},\end{aligned}$$

$\Delta_{t+1,t}(\Omega_t)$ is the product of the marginal change in parental investment and the marginal change in skill production. With our parametrization, this is given by

$$\Delta_{t+1,t}(\Omega_t) = \frac{\alpha_{4,t}}{Y_t}(\gamma_{2,t} + \gamma_{3,t} \ln \theta_t).$$

This short-term effect is heterogeneous by the level of family income and the existing stock of the child's skills. The marginal increase in investment is decreasing in the current level of income, as would be expected given the log form of the investment equation. The key parameter for the heterogeneity of the short-term effect is $\gamma_{3,t}$, with $\gamma_{3,t} > 0$ implying a higher return to investment for children with higher existing stocks of skills.

The dynamic model of skill development we estimate also allows us to consider the *long-term* effect of an income transfer at age t on outcomes beyond the immediate next period. The expected long-term effect of a marginal increase in income at period t on children's skills in period $t+2$ is given by

$$\begin{aligned}\Delta_{t+2,t}(\Omega_t) &= \frac{\partial \ln \theta_{t+2}}{\partial Y_t} \\ &= \Delta_{t+1,t}(\Omega_t)\frac{\partial \ln \theta_{t+2}}{\partial \ln \theta_{t+1}}(1+\frac{\partial \ln I_{t+1}}{\partial \ln \theta_{t+1}})\end{aligned}$$

Note that we are analyzing the long-term effect of a one-time change in income at period t; income remains at baseline levels for all subsequent periods. With our parametrization, the long-term effect becomes

$$\Delta_{t+2,t}(\Omega_t) = \frac{\alpha_{4,t}}{Y_t}(\gamma_{2,t} + \gamma_{3,t}\ln\theta_t)(\gamma_{1,t+1} + \gamma_{3,t}\ln I_t)(1+\alpha_{4,t+1}).$$

The short-term effect ($\Delta_{t+1,t}(\Omega_t)$) and the long-term effect ($\Delta_{t+2,t}(\Omega_t)$) can differ in general. Our model of skill and investment dynamics allows for the possibility that either short-term effects are higher than long-term effects (the effect of the policy "fades-out" as the child ages) or that long-term effects can exceed short-term effects (early interventions have a kind of "multiplier effect" on later skill development).

Effects on Final Skills

We first consider a simple exercise designed to assess the optimal timing of the income transfer. In Figure 2.7 we show the average percent change in the level of latent children's skills at age 13-14 by the different timing (age) of income transfer:

$$100 \cdot E(\frac{\theta'_T(a) - \theta_T}{\theta_T})$$

where $\theta'_T(a)$ is level of skill at age $t = T$ (age 13-14) with an income transfer of $1,000 dollars (in 2012 $) provided to the family at age a, and θ_T is level of skill at age 13-14 in the baseline model (no income transfer). The transfer is a one-time transfer and does not affect the future levels of income. The figure shows that a $1,000 transfer given at age 5-6 increase the average stock of age 13-14 skills by about 1 percent. Providing the same transfer later in the childhood period has a smaller average effect. Providing a $1,000 transfer at age 11-12 would increase the average skill stocks at age 13-14 by less than 0.4 percent. We estimate that providing transfers early in the development period would have a long-term effect that exceeds the short-term effect of providing a transfer in later childhood. This result reflects the high productivity of investment in the early periods and the high level of productivity of existing stocks of skill in producing future skills (limited fade-out).

Effects on Completed Schooling

Figure 2.8 displays the results of the same set of policy experiments as in Figure 2.7 but using completed schooling at age 23 as the outcome. In this Figure, we plot $E(S'(a) - S)$, where $S'(a)$ is the number of months of completed schooling at age 13-14 with an income transfer of $1,000 given at age a, and S is the number of months of completed schooling at age 13-14 in the baseline model (no income transfer). These estimates provide a meaningful metric to evaluate the magnitude of the policy effects. We find that a $1,000 transfer given at age 5-6 would increase the number of average months of completed schooling by about 1.80 months. Providing the same transfer at a later period would increase completed schooling by only 0.55 months.

Comparison with Dahl and Lochner (2012a)

Our estimated effects of a family income transfer on children outcomes are similar to some previous finding in the literature using different sources of identifying variation. Using changes in the Earned Income Tax Credit (EITC) to instrument for family income, Dahl and Lochner (2012a) find that a $1,000 dollars increase in family income implies an increase in PIAT score of about 4.5% of a standard deviation. [20] To directly compare our estimates to their reported effects, we compute the equivalence between their PIAT score outcome and years of schooling. We calculate that the Dahl and Lochner (2012a) estimates imply an average increase of about 0.54 months of schooling for a $1,000 transfer at age 11-12. [21] This results is quite similar to our main results (see Figure 2.8).

Heterogeneous Treatment Effects

The previous results showed the average effect of policies providing transfers at different stages of the development process. Our modeling framework allows potentially important sources of heterogeneity by the child's initial skills, mother's skills, and

[20] This result is based on the results reported in the correction dated March 2016 to the previous results (Table 4). In comparing our results to their results, it should be noted that the policy considered is different. Dahl and Lochner (2012a) consider a change in the EITC, which affects after-tax wage rates, parental labor supply, and hence parental time allocation, and we consider here a pure income transfer (where we do not distinguish between income from labor and other sources).

[21] As an outcome, Dahl and Lochner (2012a) use a combined PIAT test score (the average of the three separately age-standardized tests in Math, Reading Recognition, and Reading Comprehension). We rescale the PIAT scores in terms of schooling in the same way as we estimate the factor loadings for different skill measures. Define S to be years of schooling at age 23, Z_T to be the PIAT test score at age 13-14 (period T = 4), and Z_{T-1} the PIAT test score at age 11-12 (period 3). We write

$$S = \mu_S + \alpha_S \ln \theta_T + \eta_S$$
$$Z_T = \mu_T + \lambda_T \ln \theta_T + \epsilon_T$$
$$Z_{T-1} = \mu_{T-1} + \lambda_{T-1} \ln \theta_{T-1} + \epsilon_{T-1}$$

Under the assumption that error terms are uncorrelated, the following ratio of covariances provides the scaling of adult schooling with respect to the PIAT test score: $\frac{\alpha_S}{\lambda_T} = \frac{Cov(S, Z_{T-1})}{Cov(Z_{T-1}, Z_T)}$.

initial family income levels; all of which could affect the individual level treatment effect. The model estimates allow us to directly estimate this heterogeneity in the policy treatment effects.

Figure 2.9 plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income. This figure also plots the average treatment effect (ATE), the average effect over the income distribution; the same effect as reported above. While the ATE is about 1.8 months, the effect varies considerably depending on the child's initial level of income. For the children from poor households in the 9-10th income percentiles, the effect of the income transfer is to increase completed schooling by around 4 months, and for the children from the richest households, the effect is near 0. The large heterogeneous effects by family income stem from the estimated importance of family income in producing child investments and the estimated positive correlation of income with maternal skills and the child's initial skills. This heterogeneity in the effects by income mirrors the heterogeneity in income effects found in previous papers using alternative sources of identification (see Dahl and Lochner, 2012a; Loken *et al.*, 2012). Using the varied effects of the Norwegian oil boom to instrument for family income, Loken *et al.* (2012) report estimates on completed schooling which are smaller in magnitude than those reported here, but similar qualitatively in finding that the effects are substantially larger for low income Norwegian families

Figure 2.10 plots the heterogeneous effect of the same policy but by the level of the child's initial (age 5-6) skill. The ATE plotted in this Figure is the same as in the previous figure as it is simply the effect averaged over the initial skill distribution. In this Figure, we also find evidence of heterogeneous treatment effects with low initial skill children benefiting more (about 7 months of additional schooling) from the policy intervention than high initial skill children (near 0 effect). But the importance of heterogeneity by initial skill is substantially less than by family income. This suggest that it is better to target the policy to low income households than low skill households, but of course it cannot be worse to target based on both criteria.

2.7.4 Comparing Model Predictions: Quantifying the Importance of Model Generality and Measurement Error

Our results presented thus far have been focused on our preferred model estimates: estimates of the general unrestricted technology (Model 1) with measurement error correction. We next briefly discuss how the estimates of the primitive production technology would differ if we were to instead estimate the restricted model (Model 2) or ignore the measurement error issues. This analysis allows us to quantify how important measurement error and model generality are to our findings, using policy predictions on adult schooling as a meaningful metric for comparison.

Table 2.6 presents estimates for four versions of the model: Models 1 and 2, using both measurement error corrected and not corrected estimators. For each model and estimator, we re-estimate all parts of the model: the investment and technology process equations at each age and the final adult outcome equation. The estimates of the primitive parameters for these equations can be found in the Appendix; we present here only the implied policy effects.

In Panel A of Table 2.6, we present the average treatment effects (ATE) on adult schooling of the $1,000 income transfer at various ages. The first row repeats the estimates from the preferred model: using the unrestricted Model 1 and correcting for measurement error, we estimate that $1,000 income transfer at age 5-6 would increase average schooling by about 1.8 additional month. In comparison, using the restricted Model 2 (assuming constant returns to scale and no TFP dynamics) would imply an estimated increase in average schooling of about one-quarter this effect, at

0.40 additional months. This shows that restricting the model and ignoring possible TFP dynamics and non-constant returns to scale would severely bias downward the implied effects of income transfers on children's skill development.

The next panel of Table 2.6 presents the estimated ATE using the same models but not correcting for measurement error. Using these uncorrected estimates, we estimate policy effects less than half the size of the preferred measurement error corrected estimates of the most general model, Model 1. These substantially lower estimates of the effect of an income transfer are consistent with the standard attenuation bias in standard linear models, where classical measurement error biases coefficient estimates toward 0. Our models are dynamic, non-linear, and consist of inter-related multiple equations, so there is no clear theoretical prediction about the sign of the measurement error bias. But we estimate in this case that ignoring measurement error would substantially bias downward the estimates of the ATE of the income transfer policy.

Panel B of Table 2.6 repeats the analysis but focusing on the heterogeneity in the treatment effect at different parts of the family income distribution. Similar conclusions are evident here: restricting the model to have constant returns to scale and no TFP dynamics or ignoring measurement error would substantially reduce the estimated policy effect of the income transfer. We see that ignoring measurement error would bias the estimated policy effect on low income families at the 9-10th percentile from an effect size of about 4 months to only 1.4 - 1.8 months.

2.7.5 Cost-Benefit Analysis

We have thus far shown that the estimated model implies that a policy intervention of providing income transfers to family would produce modest but positive gains in children's skills, with larger effects for poorer households. Would these gains be justified given the cost? We next present a simple cost-benefit analysis to answer this question.

Table 2.7 shows the effects of the income transfer policy, by children's age, on the present value of earnings. The Table also provides the associated cost of that policy, including the cost of additional schooling. In this analysis, we consider a median earner worker. The expected present value of her lifetime earnings when she is age 5-6 is calculated to be approximately \$ 260,000 (in 2012 dollars). [22] The benefit of this policy is the comparison between the present value of worker's earnings with and without that policy during the childhood. In other words, we compute the counterfactual present value of earnings if the worker's family had received the income transfer when the worker was a child. The effect of the family income transfer to the growth in children earnings are computed using estimates in Table 2.5 under the assumption that the change in the growth rate due to the policy intervention is constant over the life-cycle. Table 2.7 suggests that, considering both the cost of the income transfer and the cost of additional education, the net benefit of the policy is positive for any age, and the effect is largest when implemented at age 5-6. The additional present value for the policy intervention at age 5-6 is slightly more than \$ 5,500 and the net benefit is around \$ 2,700.

2.8 Conclusion

This paper develops new identification concepts and associated estimators for the process of skill development in children. One of the key empirical challenges in this context is that the various measures of children's skills are in general imperfect and arbitrarily located and scaled. We introduce the concept of known location and scale production technologies, which are the type of technologies actually estimated in

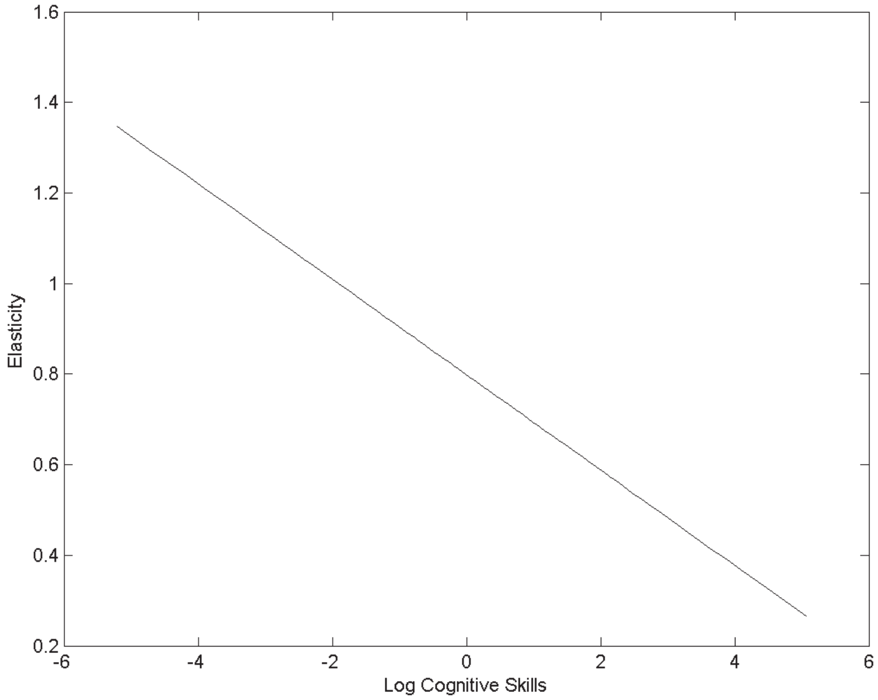
[22] The baseline present value of earnings is computed using data from the Bureau of Labor Statistics (BLS) for the fourth quarter of 2012 with a discount rate of 4 %.

many previous papers, and show that for these technologies, standard measurement assumptions non-parametrically identify the production technology, up to the normalization of initial period skills. Importantly, we show non-parametric identification for these cases without re-normalizing latent skills each period which can bias the production technology. For production functions which do not have a known location or scale, additional assumptions are necessary, and we provide empirically grounded assumptions which are sufficient for identification of these more general technologies. Our paper provides the first analysis of these crucial identification tradeoffs, and hopefully will serve as a useful guide for future work.

Based on our identification results, we develop a robust method of moments estimator and show that it can be implemented using a sequential algorithm. Our estimator does not require strong assumptions about the marginal distribution of measurement errors or the latent factors. We estimate the skill production process using data for the United States and a flexible parametric model of skill development allowing for non-constant returns to scale, dynamics in TFP, and for parental investment to endogenously depend on unobserved children's skills.

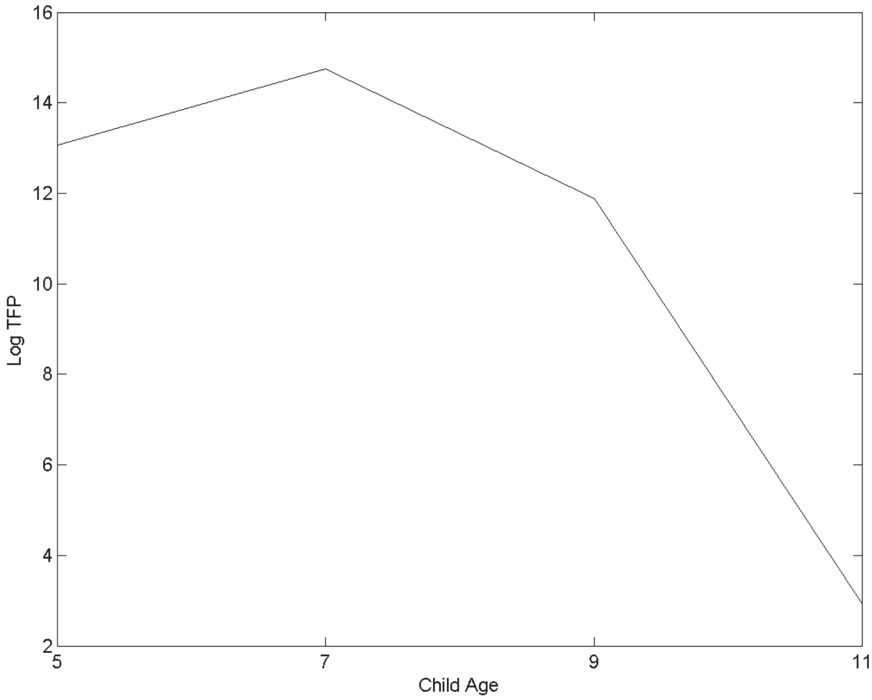
Our empirical results show a pattern of rapid skill development from age 5 to 14. We find that as children age, not only does their mean skill level increase, but the level of skill inequality also increases. Our parameter estimates reveal that investments are more productive at early ages and in particular for disadvantaged children. Our findings of a positive return to income transfers at early ages, especially for poorer households, is largely consistent with prior evidence of a positive effect of income on a number of child outcomes (see Dahl and Lochner, 2012a; Loken *et al.*, 2012) using different sources of identification. Our results suggest that family income is a better "target" than initial children's skills for children's skills. Lastly, our finding that that the estimated policy effects would be substantially smaller if one estimated a restricted technology or ignored measurement error demonstrates the critical importance of allowing for general technologies and correcting estimates for measurement error.

Figure 2.1: Estimates of Skill Production Elasticity with Respect to Investment at Age 5-6 (Model 1)



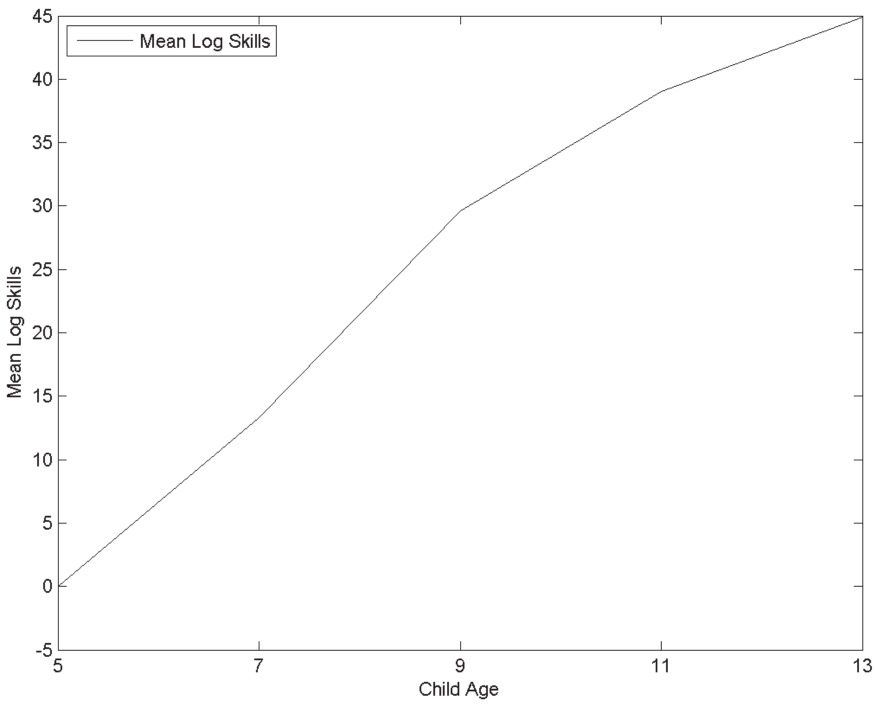
Notes: This figure shows the measurement error corrected estimates of the elasticity of children's skills at age 7-8 (θ_1) with respect to parental investments at age 5-6 (I_0) for Model 1: $\frac{\partial \ln \theta_1}{\partial \ln I_0} = \gamma_{2,0} + \gamma_{3,0} \ln \theta_0$.

Figure 2.2: Total Factor Productivity (TFP) Estimates (Model 1)



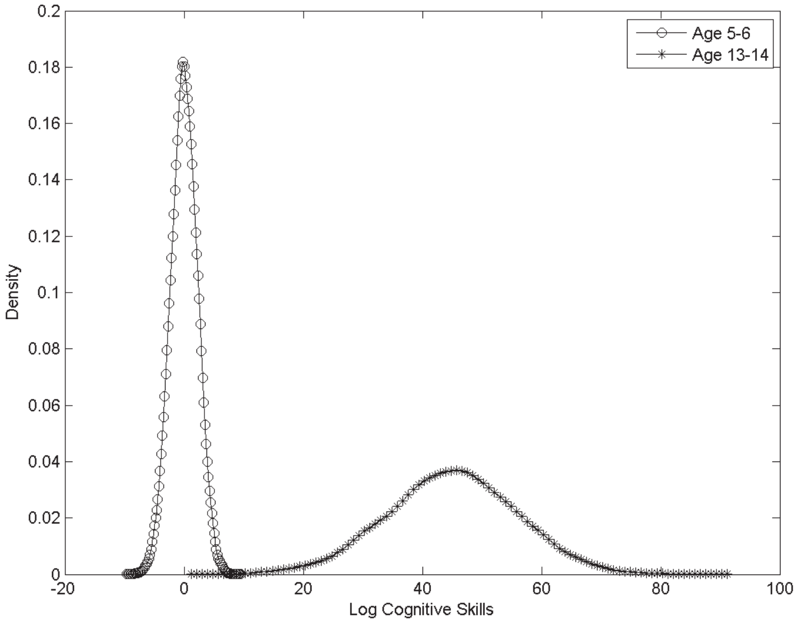
Notes: This figure shows the estimated log TFP (correcting for measurement error) for Model 1 (see Section 2.6.2). The x-axis shows children age. Child age of 5 is age 5-6, 7 is age 7-8, and so on.

Figure 2.3: Estimated Mean of Log Latent Skills (Model 1)



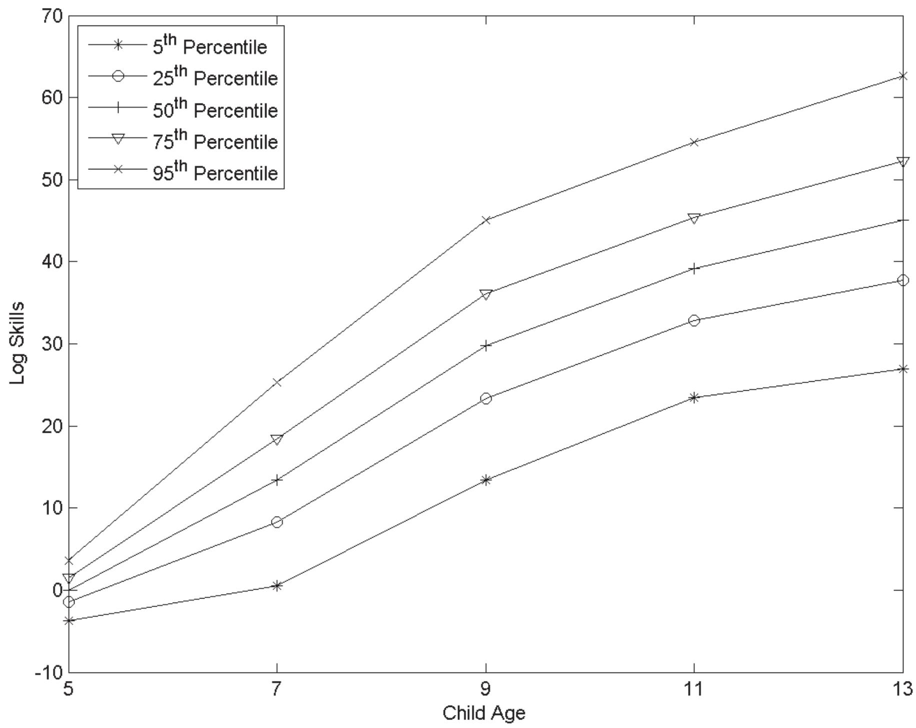
Notes: This figure provides the mean log latent skills ($E(\ln \theta_t)$) predicted by the estimated Model 1 (see Section 2.6.2), controlling for measurement error) . The x-axis shows children age. Child age of 5 is age 5-6, 7 is age 7-8, and so on. Log latent skills at age 5-6 are normalized to be mean 0.

Figure 2.4: Estimated Distribution of Log Cognitive Latent Skills at Age 5-6 and Age 13-14 (Model 1)



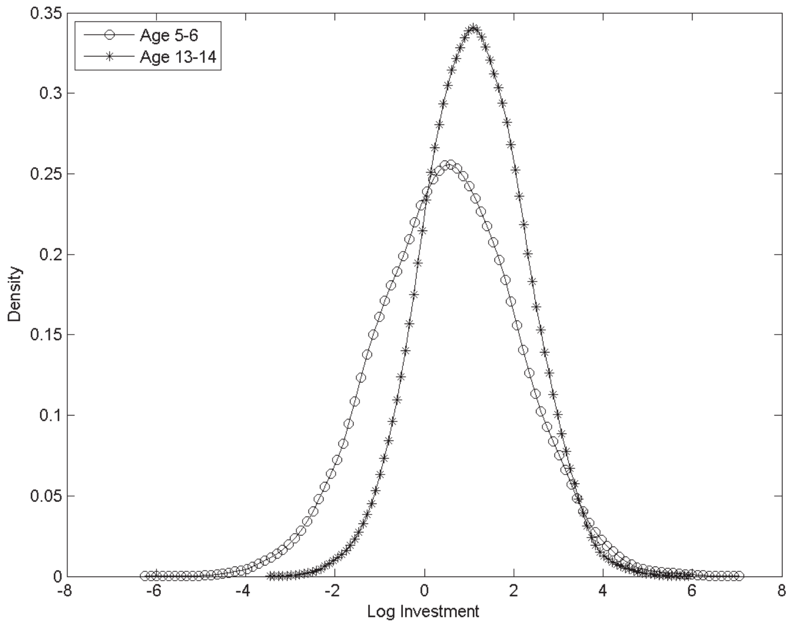
Notes: This figure shows the distribution of log latent skills at age 5-6 and at age 13-14 simulated from the estimated Model 1 (see Section 2.6.2), controlling for measurement error. Log latent skills at age 5-6 are normalized to be mean 0.

Figure 2.5: Estimated Dynamics in the Latent Skills Distribution (Model 1)



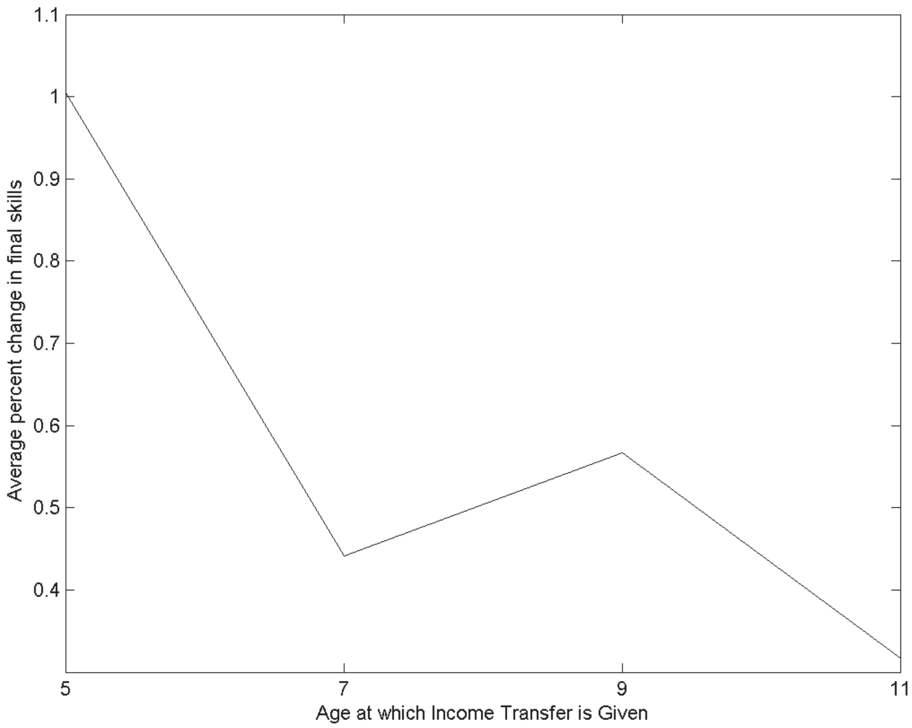
Notes: This figure shows the dynamics in the distribution of the log latent skill distribution for the estimated Model 1 (see Section 2.6.2), controlling for measurement error. Log latent skills at age 5-6 are normalized to be mean 0.

Figure 2.6: Estimated Distribution of Log Investments at Age 5-6 and Age 13-14 (Model 1)



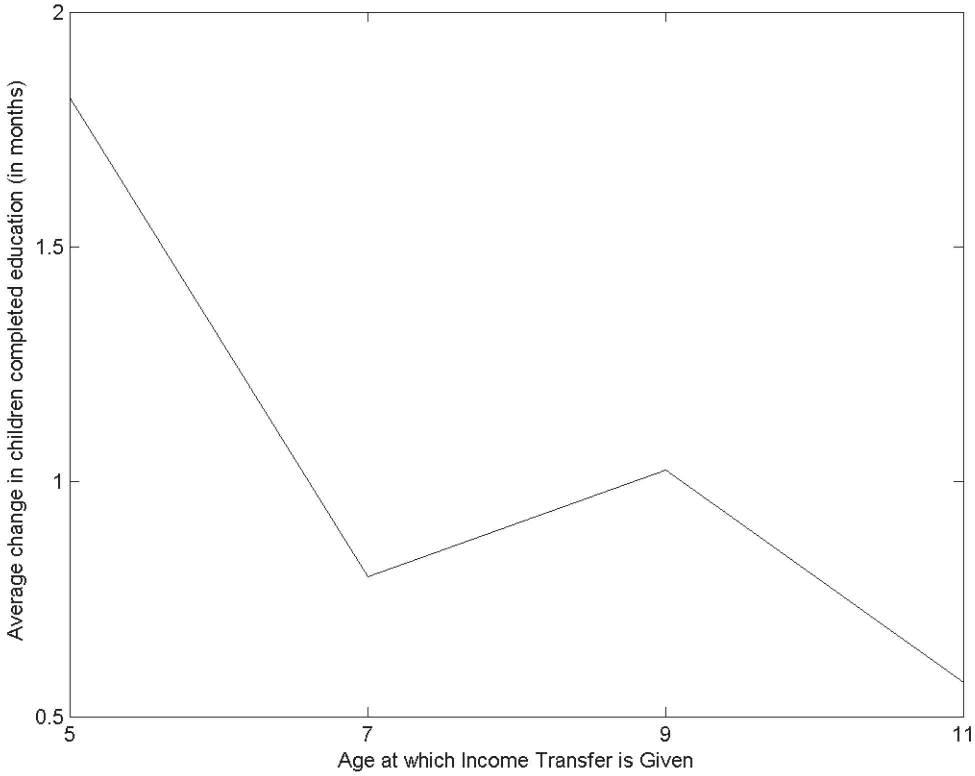
Notes: This figure shows the distribution of log latent investments at age 5-6 and at age 13-14 simulated from the estimated Model 1 (see Section 2.6.2), controlling for measurement error.

Figure 2.7: Average Effect of Income Transfer by Age of Transfer (Outcome: Final Period θ_T Skills)



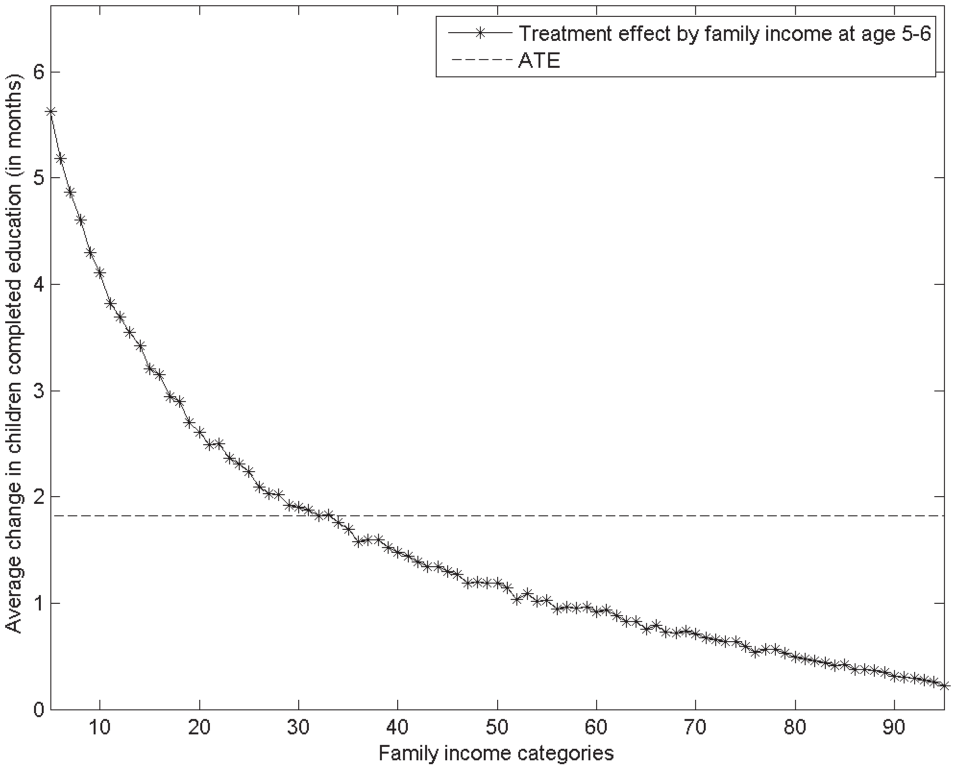
Notes: This figure shows the average percent change in the level of latent children's skills at age 13-14 by the different timing (age) of income transfer for the estimated Model 1 (see Section 2.6.2), controlling for measurement error. The transfer is \$1,000 in family income at some age t. We report $100 \cdot E(\frac{\theta'_T(a)-\theta_T}{\theta_T})$, where $\theta'_T(a)$ is level of skill at age 13-14 with an income transfer of \$1,000 dollars provide to the family at age a and θ_T is level of skill at age 13-14 in the baseline model (no income transfer).

Figure 2.8: Average Effect of an Income Transfer by Age of Transfer (Outcome: Schooling at Age 23)



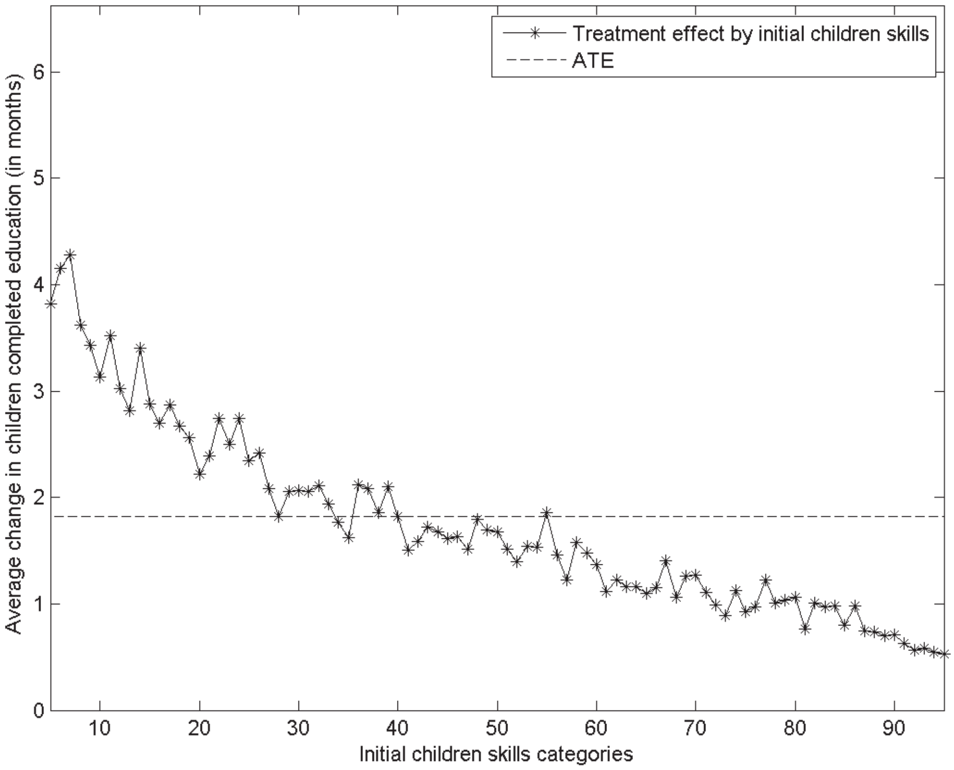
Notes: This figure shows the average change in the number of months of completed schooling at age 23 by different timing (age) of income transfer for the estimated Model 1 (see Section 2.6.2), controlling for measurement error. We report $E\left[S^{'}(a) - S\right]$, where $S^{'}(a)$ is the number of months of completed schooling at age 23 with an income transfer of \$1,000 given at age a while S is the number of months of completed schooling in baseline model (no income transfer). This figure reports the results of the same policy experiment as Figure 2.7 but with a different outcome measure.

Figure 2.9: Heterogeneity in Policy Effects by Age 5-6 Household Income (Outcome: Schooling at Age 23)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income for the estimated Model 1 (see Section 2.6.2), controlling for measurement error. Each income category is defined as the people contained between n^{th} and the $n-1^{th}$ of the percentiles of the income distribution. For example, Income category 10 in the graph means the people who belong between the 9^{th} and 10^{th} percentile of the income distribution. In the estimated income distribution for our sample, income categories 10, 50, and 90 contain families with about $14,000, $45,000, and $145,000 of annual family income. This figure also plots the average effect over the income distribution.

Figure 2.10: Heterogeneity in Policy Effects by Age 5-6 Children's Skills (Outcome: Schooling at Age 23)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of the child's initial (age 5-6) skill for the estimated Model 1 (see Section 2.6.2), controlling for measurement error. Each initial skills category includes the children contained between n^{th} and the $n-1^{th}$ of the percentiles of the skills distribution. For example, skill category 10 is the children between the 9^{th} and 10^{th} percentile of the initial skills distribution. This figure also plots the average effect over the initial skill distribution.

Table 2.1: Sample Descriptive Statistics

	Mean	Std
N Obs	19,070	
N of Mothers	3,199	
N of Children	4,941	
% Male Children	51.32	
% Female Children	48.68	
% Hispanic Children	21.44	
% Black Children	30.44	
% Other races	48.12	
Mom Education	12.59	2.63
Family Income	61,657.88	47,527.85
Children Final Years of Education	13.30	2.36

Notes: This table shows the main descriptive statistics of the CNLSY79 sample we use to estimate the model. Children's Completed Education is the child's completed years of education at age 23. The variable "other races" represents all children which are not black neither Hispanic (i.e. it includes white, non-Hispanic children). Income is in $2012 USD.

Table 2.2: Estimates for Initial Conditions

	Log Child Skills at age 5	Log Mother Cognitive Skills	Log Mother Noncognitive Skills	Log Family Income
		Variance-Covariance Matrix		
Log Child Skills at age 5	4.947	6.254	0.122	0.668
	(0.471)	(0.479)	(0.031)	(0.065)
Log Mother Cognitive Skills	6.254	30.190	0.593	2.588
	(0.479)	(1.032)	(0.137)	(0.099)
Log Mother Noncognitive Skills	0.122	0.593	0.046	0.058
	(0.031)	(0.137)	(0.017)	(0.012)
Log Family Income	0.668	2.588	0.058	0.780
	(0.065)	(0.099)	(0.012)	(0.018)
		Correlation Matrix		
Log Child Skills at age 5	1.000	0.512	0.256	0.340
	(-)	(0.026)	(0.029)	(0.027)
Log Mother Cognitive Skills	0.512	1.000	0.504	0.533
	(0.026)	(-)	(0.025)	(0.015)
Log Mother Noncognitive Skills	0.256	0.504	1.000	0.307
	(0.029)	(0.025)	(-)	(0.022)
Log Family Income	0.340	0.533	0.307	1.000
	(0.027)	(0.015)	(0.022)	(-)

Notes: This table shows the estimated variance-covariance matrix (Σ_Ω) and associate correlation matrix of the initial conditions at age 5-6. Standard errors in parenthesis are computed using a cluster bootstrap.

Table 2.3: Estimates for Investment (Model 1)

Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.230	0.027	0.020	0.018
	(0.059)	(0.009)	(0.009)	(0.009)
	[0.14, 0.33]	[0.01, 0.04]	[0.01, 0.04]	[0.01, 0.03]
Log Mother Cognitive Skills	0.071	0.004	0.012	-0.005
	(0.022)	(0.009)	(0.015)	(0.013)
	[0.04, 0.12]	[-0.01, 0.02]	[-0.01, 0.04]	[-0.02, 0.02]
Log Mother Noncognitive Skills	0.359	0.742	0.694	0.712
	(0.131)	(0.060)	(0.084)	(0.088)
	[0.11, 0.54]	[0.64, 0.82]	[0.52, 0.81]	[0.54, 0.82]
Log Family Income	0.341	0.227	0.274	0.275
	(0.076)	(0.056)	(0.076)	(0.087)
	[0.25, 0.48]	[0.15, 0.33]	[0.17, 0.43]	[0.17, 0.44]
Variance Shocks	1.186	1.019	0.868	1.087
	(0.232)	(0.148)	(0.236)	(0.296)
	[0.96, 1.53]	[0.83, 1.29]	[0.66, 1.33]	[0.82, 1.64]

Notes: This table shows the measurement error corrected estimates for the investment equation for Model 1 (see Section 2.6.2). Each column shows the coefficients of the investment equation at the given ages. The dependent variable is investment in period t which is determined by the covariates at time t . For example, the first column shows the coefficients at age 5-6 for both contemporaneous parental investments and contemporaneous child's skill and contemporaneous family income. Both standard errors in parenthesis and the 90% confidence interval in square brackets are computed using a cluster bootstrap.

Table 2.4: Estimates for Skill Technology (Model 1 and Model 2)

	Model 1 (Free Return to Scale Technology and TFP Dynamics)				Model 2 (Restricted Return to Scale Technology and No TFP Dynamics)			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	1.966	1.086	0.897	1.065	0.739	0.816	0.833	0.910
	(0.153)	(0.036)	(0.027)	(0.029)	(0.087)	(0.072)	(0.105)	(0.096)
	[1.69, 2.21]	[1.03, 1.15]	[0.84, 0.93]	[1.01, 1.11]	[0.61, 0.88]	[0.69, 0.93]	[0.71, 1.02]	[0.76, 1.07]
Log Investment	0.799	0.695	0.713	0.252	0.300	0.187	0.170	0.087
	(0.262)	(0.339)	(0.404)	(0.541)	(0.077)	(0.069)	(0.097)	(0.095)
	[0.41, 1.23]	[0.15, 1.24]	[-0.10, 1.25]	[-0.53, 1.20]	[0.18, 0.42]	[0.08, 0.32]	[-0.01, 0.30]	[-0.07, 0.23]
(Log Skills *	-0.105	-0.005	-0.003	0.003	-0.040	-0.004	-0.003	0.003
Log Investment)	(0.066)	(0.019)	(0.013)	(0.010)	(0.026)	(0.015)	(0.014)	(0.009)
	[-0.22,-0.03]	[-0.04, 0.03]	[-0.02, 0.02]	[-0.02, 0.02]	[-0.09,-0.01]	[-0.03, 0.02]	[-0.03, 0.02]	[-0.02, 0.01]
Return to scale	2.660	1.776	1.606	1.320	1.000	1.000	1.000	1.000
	(0.225)	(0.317)	(0.398)	(0.535)	(-)	(-)	(-)	(-)
	[2.30, 3.02]	[1.25, 2.31]	[0.79, 2.14]	[0.58, 2.25]	[-,-]	[-,-]	[-,-]	[-,-]
Variance shocks	5.612	4.519	3.585	4.019	2.110	1.279	0.944	0.903
	(0.174)	(0.184)	(0.181)	(0.247)	(0.178)	(0.144)	(0.163)	(0.165)
	[5.37, 5.93]	[4.27, 4.89]	[3.27, 3.88]	[3.70, 4.46]	[1.88, 2.44]	[1.09, 1.57]	[0.78, 1.32]	[0.74, 1.33]
Log TFP	13.067	14.747	11.881	2.927	0.000	0.000	0.000	0.000
	(0.295)	(0.367)	(0.541)	(0.957)	(-)	(-)	(-)	(-)
	[12.67,13.61]	[14.22,15.47]	[11.17,13.00]	[1.38, 4.65]	[-,-]	[-,-]	[-,-]	[-,-]

Notes: This table shows the measurement error corrected estimates for the technology of skills formation for both Model 1 and Model 2 (see Sections 2.6.2 and 2.6.2) . Each column shows the coefficients of the technology of skills formations at the given age. The dependent variable is log skills in the next period $t + 1$, and the covariates (inputs) are at time t. For example, the first column shows the coefficients for the skills inputs at age 5-6 which lead to log skills at age 7-8. Both standard errors in parenthesis and the 90% confidence interval in square brackets are computed using a cluster bootstrap.

Table 2.5: Estimates for Adult Outcome Equation (Model 1)

	Schooling	Log Wage
Constant	7.088	9.444
	(0.399)	(0.121)
	[6.56, 7.71]	[9.26, 9.64]
Log Children Skills	0.151	0.021
at age 13-14	(0.010)	(0.003)
	[0.14, 0.16]	[0.02, 0.03]
Variance Shock	4.333	0.246
	(0.143)	(0.012)
	[4.07, 4.56]	[0.22, 0.26]

Notes: This table shows the estimates for two adult outcome equation specifications: schooling and log earnings. In both cases the estimates are for Model 1 (see Section 2.6.2) and they are corrected for measurement error. The dependent variable is either the years of completed education for the child at age 23 or log earnings at age 29. Both standard errors in parenthesis and the 90% confidence interval in square brackets are computed using a cluster bootstrap.

Table 2.6: Estimated Policy Effects under Different Modeling Assumptions

Panel A: ATE by Age of Income Transfer

Measurement Error Corrected

Model	Age of Income Transfer ($ 1000) Age 5-6	Age 7-8	Age 9-10	Age 11-12
Model 1	1.818	0.799	1.025	0.574
	[0.93, 2.56]	[0.29, 1.33]	[-0.05, 2.15]	[-0.39, 1.74]
Model 2	0.404	0.179	0.229	0.128
	[0.22, 0.64]	[0.07, 0.32]	[-0.02, 0.42]	[-0.10, 0.36]

Not Corrected for Measurement Error

Model	Age of Income Transfer ($ 1000) Age 5-6	Age 7-8	Age 9-10	Age 11-12
Model 1	0.687	0.220	0.210	0.251
	[0.48, 0.90]	[0.09, 0.36]	[0.07, 0.34]	[0.06, 0.47]
Model 2	0.846	0.271	0.259	0.309
	[0.62, 1.06]	[0.12, 0.44]	[0.09, 0.41]	[0.08, 0.55]

Panel B: ATE at age 5-6 by Family Income

Measurement Error Corrected

Low Income Families (10^{th} Income Percentile)		High Income Families (90^{th} Income Percentile)	
Model 1	4.11	Model 1	0.313
Model 2	0.91	Model 2	0.070

Not Corrected for Measurement Error

Low Income Families (10^{th} Income Percentile)		High Income Families (90^{th} Income Percentile)	
Model 1	1.465	Model 1	0.158
Model 2	1.806	Model 2	0.194

Notes: Panel A shows the average treatment effects on additional months of completed education by age of policy intervention ($ 1000 income transfer) for different model specifications (Model 1 vs Model 2, see Sections 2.6.2 and 2.6.2) and different estimators (controlling for measurement error or not). The 90% confidence interval in square brackets are computed using a cluster bootstrap. Panel B shows the ATE respect to family income for the different model specifications and different estimators.

Table 2.7: Average Effect of an Income Transfer by Age of Transfer (Outcome: PV of Earnings)

Panel A: Benefit-Cost Analysis by Age

Age of Intervention	Benefit on PV Earnings ($)	Direct Cost (Income Transfer) ($)	Cost of Education ($)	Net Benefit ($)
Age 5-6	5549	1000	1818	2730
Age 7-8	2437	1000	799	638
Age 9-10	3128	1000	1025	1103
Age 11-12	1750	1000	574	177

Notes: This table shows the benefit-cost analysis for a 1000 dollars transfer to family of a future median earner workers with 12 years of completed education. The benefit on the PV of earnings is the difference between the present value of earnings with and without that transfer when worker was age 5-6. The effect of family income transfer on earning growth is computed adjusting for the increased earning growth implied by estimates in Table 2.5. The cost of that policy takes into account both the direct transfer and the discounted cost of additional education that the policy induces. We use a yearly cost of school of 12,000 dollars as approximately estimated from the National Center for Education Statistics.

Chapter 3

SKILLS, PARENTAL INVESTMENTS AND PEER EFFECTS

3.1 Introduction

This paper analyzes the effect of social interactions on skill formation in children. In particular, I build and estimate a model of child development, where children grow up in different *environments*, which are defined by: peers' composition, neighborhood quality and school quality. The dynamics of skills is governed by a technology of skill formation, which depends upon parental investments, the current child's skills and the environment-specific inputs. In this framework, I shed light on the importance of the dynamic effects of children's endogenous social interactions and the parental investment decisions in explaining developmental differences between different environments. A growing consensus in the literature emphasizes the importance of neighborhoods in shaping children's opportunities later in life (Chetty and Hendren, 2016a,b; Chetty *et al.*, 2016a,b). However, despite extensive research, the mechanisms behind these results remain unexplained. This paper reconciles the previous findings of childhood exposure to neighborhood with the role of children's social interactions in child development.

This project advances the current literature of child development by building and estimating a dynamic equilibrium model of children's skill formation with two innovative empirically grounded features. First, within different environments, children endogenously select their peer groups based on their preferences for their peers' characteristics. Social interactions can exhibit the tendency of children to become friends with others who share similar characteristics: a phenomenon called homophily bias.

Second, parental investments respond to changes in peer groups. Decisions regarding parental investments depend upon a child's current peers, as well as on expectations about future peer groups. Equilibrium effects arise from the socially determined aspects of parental investments. In this framework, parental investments not only directly affect a child's skills, but also affect the development of the child's peers through social interactions. Consequently, the individual return on investing in children is affected by the equilibrium parental engagement within each environment.

Skills are formed dynamically through a technology of skill formation, which defines the complementarities between parental investments and the other inputs of child development in producing a child's skills: the current endowment of skills, the skills of peers, the school quality and the neighborhood quality. In this framework, there are two main channels through which peers affect parental behavior. First, contemporaneous changes in current peers and parental investments are related to the *static* complementarity between the two inputs. Second, permanent changes in peer composition affect parental behavior through the *dynamic* complementarity in skill formation. In other words, a permanent change in peer composition affect the return of parental investments through the dynamic aspect of skill formation.

The model is estimated using data on U.S. adolescents from the National Longitudinal Study of Adolescent Health (Add Health). Add Health provides information about friendships within each school, which is key for analyzing the formation of peer groups. Moreover, information about child achievements and parental investments are available.

The identification of the model comes with two main challenges: (i) unobserved heterogeneity in how peer groups are endogenously formed; and (ii) children's skills and parental investments are unobserved. Ignoring these issues by using correlational relationships would cause the model's estimates and subsequent quantitative analysis to be biased.

The first challenge presents itself from the fact that peer groups may be formed based on additional unobserved heterogeneity, which can cause correlation between peer groups' realization and the residual unexplained variation in skill formation. To address this concern, I implement a standard instrumental variable (IV) approach in the literature. This identification strategy exploits random variations in cohort composition within school / across cohorts. The idea behind this identification strategy is simple: random changes in cohort compositions affect the opportunities for friendships between children. These shifts in the formation of peer groups affect the return of parental investments and the subsequent parental decisions. [1]

In addressing the second challenge, Cunha *et al.* (2010) illustrate that even the classical measurement error in measuring a child's skills can cause important biases in estimating the technology of children's skill formation. Following the approach in Cunha *et al.* (2010) and Agostinelli and Wiswall (2016a), I implement a dynamic latent factor model, which allows me to identify the joint distribution of latent skills and investments by exploiting multiple measurements in the data.

I estimate the model via simulated method of moments (SMM). I find that parental investments and peers are substitute inputs in producing children's skills. At the same time, I find a strong dynamic complementarity between parental investments and future expected peers. As a result of these two findings, a permanent change in peer composition has two opposing effects on parental investments. On one hand, "better" peers generate contemporaneous substitution effects in investment decisions due to the high substitutability in the production function. On the other hand, higher

[1] For previous use of similar source of identifying variation, see Hoxby (2000); Hanushek *et al.* (2003); Ammermueller and Pischke (2009); Lavy and Schlosser (2011); Lavy *et al.* (2012); Bifulco *et al.* (2011); Burke and Sass (2013); Card and Giuliano (2016); Carrell *et al.* (2016); Olivetti *et al.* (2016); Patacchini and Zenou (2016)

expected future skills for peers produce an "income" effect through the dynamic complementarity of skill formation. Parents have the incentive to invest more in their children because a higher-skilled child benefits more from higher-skilled peers in the future.

Furthermore, my estimates suggest that the formation of peer groups displays an extensive degree of homophily bias. I show evidence of homophily bias with respect to a child's race and level of latent skills. A child who is in the lower quartile of the skill distribution and belongs to a minority group is four times more likely to befriend a same-race child than a different-race child. In addition, the same child is two times more likely to befriend a same-skill and same-race child than a same-race child in the upper quartile of skill distribution.

I first use the estimated model to analyze the extent to which growing up in different environments accounts for the variation in children's outcomes. I find sizable effects for children moving to better environments. The effects are in proportion to the exposure time. The earlier children are moved, the higher the effect. A child who is moved at age 12 to an environment where children have 1 percentile higher skills at age 16 exhibits, on average, an improvement in her skills rank at age 16 by 0.63 percentiles. The average effect is 0.48 percentiles if the child is moved at age 15. As model validation, I show that my findings track (out-of-sample) the quasi-experimental findings of childhood exposure effects of neighborhoods for the U.S. from Chetty and Hendren (2016a). In addition, my model allows me to decompose these effects. I find that *peers* account for more than half of the exposure effects.

The relative importance of peers for the exposure effects underlines the role of policies that change peers' composition and promote socioeconomic integration in environments, as a way to improve outcomes for disadvantaged children. I find that by moving the most disadvantaged children (in the lower quartile of skill distribution) from a low-income environment to a high-income environment generates important dynamic equilibrium effects, with heterogeneous treatment effects for both the moved and receiving children. I first consider a large-scale policy, i.e. a policy that moves a sizable fraction of disadvantaged children into a higher-income environment (approximately 5% of the population of the receiving cohort). I find that the policy increases the skills of the moved population of 16-year-old children, on average, by approximately 0.40 standard deviations. On average, I do not find any adverse effect for receiving children. On the contrary, when the fraction of moved population increases to 30%, I find that the policy generates *winners* and *losers*. First, I find that the policy increases the skills of the moved population of 16-year-old children on average by 0.22 standard deviations. In contrast, there is an adverse effect for receiving children, with the skills of 16-year-old children decreasing, on average, by 0.15 standard deviations. Additionally, I find that children who remained in the sending environment benefit from the outflow of the most disadvantaged companions, with an average increase in skills at age 16 of 0.17 standard deviations.

I find that large-scale changes in peers' composition generate important equilibrium feedback effects, and as a result amplify the policy effects. Ignoring equilibrium effects would lead to large biases in counterfactual policy predictions for children's final skills. In the case of the larger policy, I find that the policy predictions for the children's skills in the receiving environment would be approximately seven times smaller. Part of the bias is due to the dynamic-equilibrium feedback effects on parental investments. In fact, in the absence of dynamic-equilibrium feedback effects, the *static* complementarity between parents and peers dominates the dynamic effects of the policy.

I find that policy effects for receiving and remaining children reduce in magnitude as the fraction of moved children decreases. An increase of inflow of the most disadvantaged children from the low-income environment to the high-income environment increases the probability of the receiving children becoming friends with the new companions. For the same reason, an increase of the outflow of the moved population benefits children who remain in the sending environment. For children who were moved, the opposite is true. The higher the outflow of disadvantaged companions, the higher the chances that the moved children remain friends with each other in the new environment.

My structural model allows me to analyze the distributional policy effects. I find that large-scale changes in peers' composition exhibit heterogeneous treatment effects as a result of the endogenous formation of new peer groups. Children with lower skills (in the first quartile of the skills distribution in each subpopulation): (i) benefit the most in leaving disadvantaged social environments; (ii) benefit the most amongst the children who remained in the sending environment; (iii) are the ones who are more adversely affected in receiving the new peers. Furthermore, I find stronger policy effects for minorities, with detrimental effects in black and Hispanic children living in the receiving environment. This is explained by the fact that most of the moved children are minorities, and as a result, the minority children from the receiving environment are more likely to interact with the new companions because of the race effects in the endogenous formation of peer groups. In line with this result, previous empirical studies pointed out that peer effects seem to be stronger *intra*-race and for minorities (see Hoxby, 2000; Angrist and Lang, 2004; Imberman *et al.*, 2012)

The paper will be presented as follows. In Section 3.2, I discuss the related literature. In Section 3.3, I present the data used for the empirical work and preliminary empirical results. In Section 3.4 and 3.5, I present the model. In Section 3.6, I describe the identification strategy. Section 3.7 contains a discussion of the structural estimation and results. Section 3.8 and Section 3.9 discuss the quantitative analysis and the model validation. Section 3.10 concludes.

3.2 Related Literature

This paper builds upon two important areas of the literature: child development and social interaction. There is extensive evidence in the literature on parental and public investment in children that highlights the important role of play inside and outside the household on the development of children's skills (see Todd and Wolpin, 2003, 2007; Del Boca *et al.*, 2014b). Cunha and Heckman (2008) and Cunha *et al.* (2010) estimate a dynamic latent factor model of cognitive and non-cognitive skill formation, allowing for unobservability of both inputs and outputs, endogeneity of inputs and unobserved child-specific heterogeneity. They find that investments made early in life are more effective in remediations for low-skilled children. Agostinelli and Wiswall (2016a) follow the framework considered in Cunha *et al.* (2010) and develop a new identification strategy for the technology of skill formation with unknown total factor productivity and unknown return to scale. Their empirical results show a pattern of rapid skill development from age 5 to 14. They find that as children age, skill inequality increases. Estimates reveal that investments are more productive at early ages and in particular for disadvantaged children. This paper is the first work in the literature that sheds light upon the dynamic equilibrium effects of children's social interactions and the parental investment decisions in explaining developmental differences in children.

A wide set of previous work analyzed peer effects in various outcomes. Manski (1993) points out the challenges in identifying peer effects by considering three observable equivalent specifications in a model of social interactions: peer effects (endogenous effects), selection into peer groups (correlated effects) and common exogenous (contextual) effects. One part of the literature tried to overcome this challenge in identification of peer effects by exploiting exogenous variation in peer-group composition. [2] For example, in Abdulkadiroglu *et al.* (2014), the identification of peer effects at school is based on test-score discontinuity in admission criteria. In the context of college students, De Giorgi *et al.* (2010) and Sacerdote (2001), respectively, exploited random assignments of peers at college. Arcidiacono *et al.* (2012) develop a new algorithm for estimating peer effects using panel data and which controls for peer selection and unobserved heterogeneity using University of Maryland transcript data. Finally, Sacerdote (2001) highlights that the literature's findings on peer effects are quantitatively and statistically larger when considering non-linear models of peer effects. However, results are often context specific, a potential limitation for policy analysis. This paper is exploiting quasi-experimental variation in cohort composition within school / across cohorts, to identify the degree of complementarity between parental investments and peers in the technology of skill formation. The set of policy-invariant parameters of the model are used to evaluate policies that have not previously been implemented.

A parallel literature started to consider identification and estimation of peer effects within micro-funded models of behavior and social interactions (see Brock and Durlauf, 2001b,a, 2007; Blume *et al.*, 2011, 2015). Calvó-Armengol *et al.* (2009) estimate a model of adolescent effort choice within a social network. The authors define peers as the set of nominated children in Add Health data. However, the authors do not consider any model of network formation and peer selection, while they control for peer selection through network-specific fixed effects. This is equivalent to assuming that peers select themselves into groups but friendship formation within each group is independent of observable and unobservable characteristics of people in the group.

[2] For a complete literature review on the identification of peer effects through experiments, see Sacerdote (2014)

They find that the level of an adolescent's connectivity (position within the network measured by Katz–Bonacich centrality) is an important predictor for their school performance. Finally, Fu and Mehat (2016) estimate a model of student achievements and a class-tracking regime with endogenous parental effort using ECLS-K. They find that accounting for endogenous parental responses to class-quality changes is key to evaluating class-tracking policies. These works explicitly focus on the contemporaneous peer effects on children's outcomes. My paper contributes in this literature by building a new structural model of child development and peer effects, and by highlighting the importance of dynamic peer effects in shaping the developmental trajectories of children.

Analysis of endogenous network formation became popular among both theoretical (see for example Jackson and Wolinsky, 1996; Bala and Goyal, 2000; Dutta *et al.*, 2005; Mele, 2010) and econometric studies (see Christakis *et al.*, 2010; Sheng, 2014; Auerbach, 2016; Graham, 2016, 2017). Carrell *et al.* (2013) estimates peer effects on academic performance at the United States Air Force Academy. Using an assignment algorithm designed to foster the academic achievement of the lowest-ability students, the authors find a negative treatment effect for the targeted group. The authors provide evidence that this finding is the result of endogenous peer-group formation, which displays the tendency of students to generate homogeneous subgroups. This result underlines the importance of accounting for endogenous peer-group formation once considering policies that manipulate peer composition.

Important progress has been made in Badev (2016), where the author develops a model of individual behavior and endogenous peer selection. He estimates the model using Add Health data on smoking decisions and friendship nominations. He finds that neglecting the endogeneity of the networks leads to important biases on policy evaluations. However, Badev (2016) focuses on the contemporaneous effects of peers, and through this paper, I will emphasize how the dynamic aspect of peer effects is key in understanding the role of social interactions in child development. Additionally, the empirical analysis in Badev (2016) looks at a specific outcome for adolescents (smoking decisions), while my empirical analysis will be based on a dynamic latent factor model. This allows me to consistently study peer effects on children's skills and to avoid relying on arbitrary variables as measures for children's outcomes.

Finally, my work sheds light on the mechanisms behind the recent research on the effects of neighborhood exposure on children. Chetty and Hendren (2016a) find sizable childhood exposure effects of neighborhood. Their results show that the return for children's outcomes of moving to a better neighborhood is in proportion to the amount of time spent in that neighborhood, with a rate of 4% decline for each additional year of exposure to the origin area. In a companion paper, Chetty and Hendren (2016b) find strong correlation between the childhood exposure effects and specific characteristics of neighborhoods, like racial segregation, income inequality, school quality, and social capital. My estimated model replicates (out-of-sample) the findings in Chetty and Hendren (2016a), and it decomposes the causal effects of neighborhoods in different policy-relevant mechanisms, like the effect of peer composition, school quality and neighborhood quality in child development.

3.3 Data and Empirical Evidence

3.3.1 The National Longitudinal Study of Adolescent to Adult Health (Add Health)

This paper uses the National Longitudinal Study of Adolescent to Adult Health (Add Health). [3] The Add Health original sample comprises students among 132

[3] For additional information about the dataset, see Appendix B.1 or visit http://www.cpc.unc.edu/projects/addhealth

representative schools in the United States. There are 90,118 students, ranging between grade 7 and grade 12 in the 1994–1995 school year (Wave I). A subsample of students (20,745) is selected for having an additional home interview (in-home). The home interview includes new questions for the children and a questionnaire for one of their parents. The dataset includes specific information on family background, students' school grades and their scores in the Add Health Picture Vocabulary Test (AHPVT – a revised version of the Peabody Picture Vocabulary Test [PPVT]) , as well as information about children's peers.

A main source of information that makes the Add Health dataset particularly attractive for achieving the objective of this project is the friendship nomination. During the first two waves, children were asked, both during the in-home and in-school interviews, to nominate their best five male and best five female friends. This detailed information helps me to reconstruct the structure of friendship for every child in the sample by simply matching their identifier. Additionally, during the in-home interview, children are asked about their relationship with their parents. Respondents provide information regarding whether, during the last four weeks, they were involved in specific activities with their parents. The activities include: going shopping, sport activities, going to a movie/museum/concert or sport event, talking about personal problems or school, or working on a project for school. I use all the activities as measures of parental investments.

One important challenge in the empirical analysis of peer effects using Add Health comes from the fact that children are able to nominate up to five friends for each gender. This feature of the data can lead to a potential censoring and mismeasurement of peer groups (see for example Chandrasekhar and Lewis, 2016; Griffith, 2017). In the sample I use for my empirical analysis approximately 11% of children showing a full list of 10 best friends within the school roster. [4] To address this concern, I construct the peer-group information for each child from both the individual child's list of friends as well as from the unilateral friendship nominations coming from the other children who are not nominated. In other words, if child i does not nominate child j as a friend, but child j nominates child i, then I consider them as friends in the data. In a case where child i's list of friends is binding, I am able to recover additional friends in his peer group, alleviating the truncation problem. Furthermore, through my empirical analysis of parental investments and peer effects, I implement an IV estimation analysis to deal with the endogeneity of the network formation. Given that my instrument is unrelated to the network structure of friends, this approach is also effective in dealing with mismeasurement of peer effects. [5]

3.3.2 Descriptive Statistics

Table 3.1 reports descriptive statistics for the sample I use in the estimation of the model. The average age is 15.65. In terms of racial composition, 16% of the children are black and 17% are Hispanic, while the remaining 67% are white (or other races). On average, children report 4.48 friends out of the maximum number of 10 possible nominations. [6] The average PPVT raw score is 64.26, while the average grades for

[4]By gender, respectively, 28% of male and 32% of female children report a list of five best same-gender friends within the school roster

[5]A common instrument in the analysis of peer effects is constructed with exogenous characteristics of the friends of friends (see for example Bramoullé *et al.*, 2009; Calvó-Armengol *et al.*, 2009; Patacchini and Zenou, 2012). The validity of this instrument requires the correct measurement of the network structure, at least until the second degree of separation between links.

[6]Here, I report the average number of nominated friends. Later, in the empirical section, I will consider the total number of friends, which is constructed by also including the unilateral friendship nominations of other children. The average number of friends is approximately seven.

English, math, history and science vary from 2.72 to 2.86. [7], [8] The average family income is $42,884 (in 1994), while the average number of years of schooling for the child's mother is 13.13. [9]

Table 3.1 also provides descriptive statistics on measures of a parent's (mother's) level of engagement, which I use to identify latent parental investments in my empirical analysis. The most frequent activities performed by children with their mothers in the four weeks preceding the in-home interview are: shopping (72%), talking about school work (63%) and school activities (54%). On average, from one-third to approximatively a half of children had a conversation about personal issues or an argument with their mothers (47% of them talked with their mothers about someone they are dating or about a party they attended, 39% talked about a personal problem and 33% had a serious argument). A quarter of children went with their mother at least once to a movie theater, museum, concert or sport events, while 38% went to a religious service. Finally, approximately 10% of children either played a sport or worked on a school project with their mother.

3.3.3 Empirical Findings on the Endogeneity of Network Formation

In this section, I provide some empirical evidence that friendships are not formed at random, but instead children display the tendency to become friends with others who are similar to themselves: a phenomenon called homophily bias. One important concern in any empirical analysis of peer effects is related to the endogeneity of the network formation (see Carrell *et al.*, 2013). In particular, the tendency of chil-

[7] The national GPA in 1994 for U.S. high schools was 2.44 in math, 2.50 in science and 2.63 in English. Source: the National Center for Education Statistics, The Nation's Report Card

[8] Only 34 out of 19,713 children achieved the maximum PPVT score. Only five children scored the minimum.

[9] Using the Current Population Survey, I find that mothers of children with similar age, as in Add Health, have an average family income for 1994 of $42,759. Their average number of years of education is 12.63.

dren with a similar background (both observable and unobservable characteristics) to socialize together can be an important challenge for identification. Furthermore, evaluation of many policies which can change cohort composition in specific social environments (e.g. school vouchers, housing vouchers, classroom tracking, etc.) is required to predict new policy-induced children's networks to account for social interactions and to predict the effects on the dynamics of children's skills.

Following the method in Currarini *et al.* (2010), I test for homophily bias in the formation of peer groups by looking at the homophily bias index (hereafter referred to as HBI) developed in Coleman (1958). The intuition behind the HBI is straightforward: the index captures the tendency of friendships to be biased towards other children of the same type, adjusting for the relative frequency of that specific type in the overall population. In detail, letting $f_{x,s}$ be the average fraction of friends who are of the same type x at school s, and $q_{x,s}$ being the total fraction of children of type x in a given school s, then we have:

$$HBI_{x,s} = \frac{f_{x,s}}{q_{x,s}} \tag{3.1}$$

A value of one for the HBI index in (3.1) suggests that friendships are formed at random, preserving the frequencies of school composition in peer-group composition. On the other hand, the homophily bias in network formation generates an HBI index greater than one: the fraction of same-type friends consistently exceeding the fraction of that type of children at school. Figure 3.1 shows graphically the HBI for race. The x-axis represents the values for the fraction of same-race children in the school ($q_{x,s}$). The y-axis displays the fraction of same-race friends ($f_{x,s}$). Each point in Figure 3.1 represents the average fraction of friends of the same race for individuals of a specific race in a specific school. Figure 3.1 shows that children of different races tend to form friendships with same-race children at a higher frequency than the frequencies of racial composition at school. Figure 3.2 is the analogue of the previous figure with respect to a child's skills. Each of the sub-figures in Figure 3.2 uses a different criterion for defining "same-skills children" relative to the standard deviation of the skills distribution. Each of the four specifications exhibits the tendency of children to become friends with other children with the same level of skills. [10]

3.3.4 *Empirical Findings on Parental Investments and Peers' Skills*

In this section, I show the extent to which changes in peers' skills induce changes in parental investment behavior. The results provide the support for the framework of my model of child development and peer effects described in the next section. In addition, these empirical findings are used in the structural estimation of my model as identifying moments. Consider the following empirical model of investment decision:

$$I_{i,s,t} = \beta_0 + \beta_1 \ln h_{i,s,t} + \beta_2 \ln \overline{H}_{i,s,t} + X_i'\beta_3 + \beta_s + u_{i,s,t} \ , \tag{3.2}$$

where $I_{i,s,t}$ is the parental investment (as a fraction of time) for parent of child i, in school s when she is t years old, which is recovered through a latent factor model (see Section 3.5.1) using data on parental engagement described in the previous section. The child's skills are defined as $h_{i,s,t}$, while $\overline{H}_{i,s,t}$ is the mean of her peers' skills. X_i is a vector of the child's and parents' exogenous characteristics and β_s is the school fixed effects. The coefficient β_2 represents the effect of peers' skills on investment decisions. The equation in (3.2) is similar to the investment decision function estimated in the previous literature, where I additionally include peers' skills as an explanatory variable (see Cunha *et al.*, 2010; Agostinelli and Wiswall, 2016a; Attanasio *et al.*, 2017a,b,c).

I first estimate the model in (3.2) using school fixed effects. Column (1) of Table

[10] The null hypothesis of random formation of peer groups with respect to race and skills is rejected in both cases at a 1% significance level.

3.2 shows the results. [11] The effect of peers on parental investment is negative and statistically significant at the 5% level. The estimate of -1.44 indicates that doubling peers' skills is associated with a decrease in investments of 1.44 percentage points. On the other hand, the effect of a child's skills on parental investment behavior is positive and statistically significant at the 1% level. The point estimate of +2.66 suggests that doubling a child's skills induces parental investments to increase by 2.66 percentage points. Overall, the school fixed effects estimates suggest that parental investments respond in opposite direction to changes in their own child's skills in comparison to changes in skills of their child's peers.

Following the analysis of endogenous network formation in the previous section, I address the endogeneity of peers' skills in (3.2), implementing a within-school instrumental variable (IV) estimator. I use variation in the racial compositions of different cohorts within the same school to analyze the effect of changes in peers' skills on parental investments, where cohorts are defined by children's ages (for previous use of a similar source of within-school/across-cohorts identifying variation for peer effects, see Hoxby, 2000; Hanushek *et al.*, 2003; Ammermueller and Pischke, 2009; Lavy and Schlosser, 2011; Lavy *et al.*, 2012; Bifulco *et al.*, 2011; Burke and Sass, 2013; Card and Giuliano, 2016; Carrell *et al.*, 2016; Olivetti *et al.*, 2016; Patacchini and Zenou, 2016). The idea behind this identification strategy is simple: the differences in cohort composition define the choice set for the children's network formations, shifting the peer-group realizations and identifying the causal effect of peers' skills on investment decisions. [12]

Evidence of homophily bias in friendships in Section 3.3.3 underlines that the

[11] All results in Table 3.2 are adjusted for measurement error through the latent factor model explained in Section 3.5.1.

[12] This identification strategy does not require that friendships are only formed within a unique cohort. It only requires that changes in cohort composition alter the peer-choice set between cohorts. Empirically, most of the friendship nominations are within the same cohort.

heterogeneous effects in the formation of peer groups in children depend on the individual characteristics of the child. Differences in the fraction of children from a minority group between cohorts would asymmetrically impact the friendship realizations that depend on race. Likewise, differences in the fraction of low-skilled children between cohorts would predict dissimilar changes in peers' skills for low-skilled versus high-skilled children. For this reason, I implement an IV specification which allows for different effects of cohort racial composition based on the individual child's own race and skills.

I construct two different instrumental variables for whether the child is part of a minority group or not. In the case of white children, the instrument corresponds to the interactions between the individual child's log-skills ($\ln h_{i,t}$) and the fraction of white children in that cohort. In the case of children from a minority group, the children's skills are interacted with the fraction of black children in that cohort. [13] Allowing for the heterogeneous effects of cohort compositions is important in terms of predicting power on the formation of peer groups, and consequently for the *relevance* of the instrumental variables.

The validity of the instruments relies on (i) the conditional independence and (ii) the exclusion restriction. The first condition requires that differences in racial composition between cohorts be uncorrelated with the unobserved heterogeneity in investment decisions. This assumption would be consistent with a sorting model of neighborhood/school choice through which parents decide where to permanently move according to their expectations about the school's composition. Random differences between the ex-ante expectations and the actual realization of the new cohort's composition would generate exogenous shifts in the set of potential peers. Addition-

[13] The regression is estimated with a within-school IV estimator, and the instrumental variables are all transformed with a within-school transformation.

ally, conditional independence is valid under the assumption that the latent factor model fully captures the child-specific unobserved heterogeneity, which is a common assumption in studies that estimate the technologies of skill formation (see Cunha and Heckman, 2007, 2008; Cunha *et al.*, 2010; Agostinelli and Wiswall, 2016a). [14] The exclusion restriction requires that the differences in racial composition between cohorts within the same school affect parental investment decisions only through the peer-effects channel, and not directly in any other way.

Figure 3.3 shows graphically the first-stage coefficients of the two instruments. In the background of the two figures is a histogram that shows the distribution of the two instruments (after controlling for the explanatory variables in (3.2)), revealing the identifying sources of the variation. Figures 3.3a-3.3b show the different predictions for peers' skills due to changes in compositional effects by race: an increase in the fraction of children with the same race within the same cohort predicts a decrease in the expected level of peers' skills in the case of a child from a minority group, while it predicts an increase in the expected level of peers' skills for a white child in that cohort. Specifically, for the average child, an increase of 20% of same-race children within the same cohort induces an increase of peers' skills of approximately 1.7% if the child is white, while it induces a decline of peers' skills of 2.2% if the child belongs to a minority group [15] . Results are stronger for the high-skilled children. For children in the 95^{th} percentile of the skills distribution, a 20% increase of same-race children within the same cohort induces an increase of peers' skills of approximately 6% if the child is white, while it induces a decline of peers' skills of 7% if the child belongs to

[14] An exception to this can be found in Cunha *et al.* (2010), where the authors consider additional model specifications that allow the factors to be correlated with unobserved heterogeneity in investment decisions. However, in order to identify the model, the authors need additional exclusion restrictions. One possible exclusion restriction is that during the first period of development, skills are uncorrelated with the unobserved heterogeneity.

[15] The average for children's skills is approximately 1.1

a minority group. Moreover, I formally test the relevance of the two instrumental variables in Panel B of Table 3.2. I find that the two instruments are relevant, with an F-statistic of the test of joint significance equal to 11.78. [16]

Column (2) in Table 3.2 reports the IV estimates. Peer effects on parental investments are both statistically and quantitatively different from the estimate of the school fixed effects estimator. In fact, using shifts induced from within-school/across-cohort changes in cohort composition, I find that the causal effect of peers' skills on parental investment is positive. The estimate suggests that doubling the average of skills of a child's peers is associated with an increase of parental investment of 0.72 percentage points.

Anticipating the next discussions regarding the identification of the model of investment decisions and the formation of peer groups, there are three main findings from the above estimates which are extremely informative for identification of the model: (i) point estimates of fixed effect estimator; (ii) the permanent nature of shift-induced changes in peers by the instrumental variables and the associated causal findings; (iii) The relative bias of the school fixed effects estimates relative to the IV estimates. Through the lens of the structural model, these three pieces of information from the empirical analysis will directly map into three specific features of the model of investment decisions and formation of peer groups: (i) *static* complementarity between parents and peers in producing skills; (ii) *dynamic* complementarity between parents and peers in producing skills; (iii) endogenous network formation and selection into peer groups on unobservables.

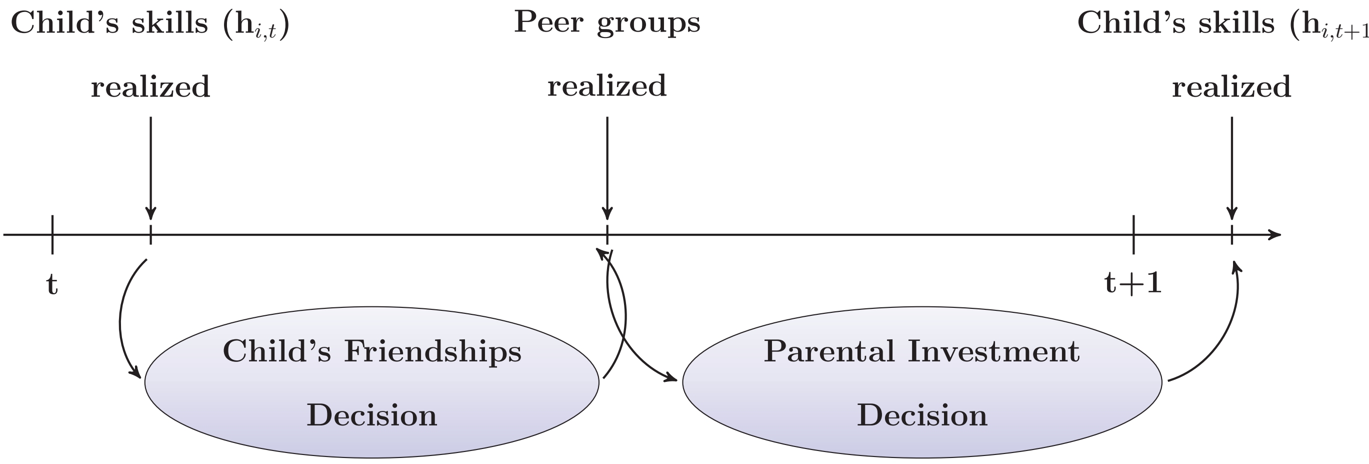
[16] Stock and Yogo provide critical values to test weak IV conditions based on the F-stat of excluded instruments. Those critical values can be interpreted as a test, with a 5 % significance level, of the hypothesis that the maximum relative bias (with respect to the OLS estimates) is 10% or at least 15%. In this case, Stock and Yogo's critical values for the F-stat of the excluded instruments are 19.93 (10%) and 11.59 (15%)

3.4 An Equilibrium Model of Parental Investments and Endogenous Network Formation

The social environment in which children live predicts their success later in life, defining important room for policy interventions in fostering children's skills. However, analysis of any policy which changes the composition of a specific social context (e.g. school vouchers, housing vouchers, classroom tracking, etc.) requires knowledge of the endogenous policy-induced changes in parental investments and peer groups in order to predict the overall policy effects. This is particularly relevant considering the empirical evidence on endogenous network formation and peer effects on parental investments shown in the previous section. For this reason, in this section, I develop the model that serves as the basis of my empirical and quantitative analysis.

This model represents a network economy populated by a finite number of families, each formed by one parent (mother) and one child. There are several environments $e \in \{1, \ldots, E\}$, each populated by N_e number of families. Children can form social networks only within each environment e. Children from different environments are isolated from each other and they cannot socially interact. The model has four periods (T= 4), each consisting of one year. The first period ($t = 1$) is when children are 13 years old, while the last period ($t = T$) is when children are 16 years old. Since I observe a negligible percentage of people changing school during the considered period (probably because children are enrolled in high school), I simplify the model by assuming that the parent cannot decide to change their environment during the model's period of study.

Parents and children solve different problems. Mothers altruistically invest time to foster children's skills. Children endogenously decide their peers according to their skills and other exogenous characteristics, forming a potentially large social



network within the environment. Potential skills spillovers for child development take place between peers within a children's social network. Peers' skills spillovers affect parental investment productivity. In equilibrium, parental investments form the within-environment children's skills distribution, determining the children's social interactions. This mechanism generates an equilibrium-feedback effect on parental behavior caused by peer effects through the formed social network.

3.4.1 The initial conditions

The initial period of the model ($t = 1$) is fixed when children are 13 years old. At the beginning of this period, I assume each family i draws the vector of individual initial conditions composed of the initial skills for both mother and child (m_i, $h_{i,1}$), its exogenous characteristics X_i, and the neighborhood quality d and school quality s. The peers' composition is defined by the set of children who share the same neighborhood quality and school quality. The combination of peers' composition, neighborhood quality and school quality defines the environments where children live. I do not allow mobility of a family between different environments during the period of consideration (when the child is between 13 and 16 years old). In Add

Health, I observe a negligible percentage of people changing school neighborhood during the considered period. This finding is probably related to the fact that parents do not tend to move once their child is already enrolled in high school. Hence, I use this assumption in the model since it significantly simplifies the model. However, different realizations of initial conditions generate the sorting of people into different environments in terms of parents' skills, income and child's skills, which is a key mechanism to analyze children's social interactions and the consequences of social isolation in generating inequality in children's outcomes. Perhaps, in an economy where selection into environments is based on a family's skills and income, children raised in high-skilled and high-income families will tend to interact with peers also raised in high-skilled and high-income families. The characterization of these social-interaction patterns are due to the sorting into environments.

3.4.2 Skill Formation

At each period t, children's skills ($h_{i,t}$) evolve dynamically through a technology of skill formation. The children's skills in the next period are produced by the current stock of children's skills, parental investment, peer effects, school quality and neighborhood quality. The first two inputs are generally considered in the literature of child development (see for example Cunha and Heckman, 2007, 2008; Cunha *et al.*, 2010; Heckman and Masterov, 2007; Del Boca *et al.*, 2014b, 2016). Here, peer effects ($\overline{H}_{i,t}$) are captured by the average of peers' skills:

$$\overline{H}_{i,t} = \frac{1}{\sum_{j \in N_e} L_{i,j,t}} \sum_{j \in N_e} L_{i,j,t} \cdot h_{j,t} \quad , \tag{3.3}$$

where $L_{i,j,t}$ is an indicator function, which equals one if child i and child j are friends, and zero otherwise. The formation of peer groups is endogenous in the model and is defined by the decision of children (see next section).

The choice of the average effect of peers' skills approach is in line with previous literature on peer effects and social networks, where the mean effect (unweighted or weighted average) is considered a first-order approximation of the peers' externality (see Brock and Durlauf, 2001b,a, 2007; Blume *et al.*, 2011, 2015; Calvó-Armengol *et al.*, 2009; Patacchini and Zenou, 2012; Patacchini *et al.*, 2012). However, in this framework, peer effects can be potentially highly non-linear, depending on the technology specification. [17] I allow the dynamics of children's skills to be affected also by the parental investments ($I_{i,t}$), some individual specific neighborhood/school effects ($A_{i,d,s,t}$) and total factor productivity (TFP). The technology of skill formation which defines children's skills in the next period looks as follows:

$$h_{i,t+1} = {h_{i,t}}^{\alpha_1} \cdot \left[\alpha_2 \left(I_{i,t}\right)^{\alpha_3} + (1-\alpha_2)\left(\overline{H}_{i,t}\right)^{\alpha_3}\right]^{\frac{\alpha_4}{\alpha_3}} \cdot A_{i,d,s,t} \cdot exp(\xi_{i,t+1})\,, \quad (3.4)$$

where I assume that $(\alpha_1,\alpha_2,\alpha_4) \in (0,1)$ and $\alpha_4 \in (-\infty, 1]$. The stochastic component $\xi_{i,t+1}$ represents the production function shocks. It is unrealized at time t and it affects children's skill dynamics. It represents the variation of skills dynamics unexplained by the specified technology in (3.4). I allow $\xi_{i,t+1}$ to be correlated with the unobservable heterogeneity in the formation of peer groups process $\nu_{i,j,t}$, which represents the unexplained variation in friendship realizations between children from the model in (3.5). Specifically, I define the shock for the formation of peer groups to be $\nu_{i,j,t} = \widetilde{\nu}_{i,j,t} + \zeta_{i,j,t}$, where $\zeta_{i,j,t} \sim \mathrm{N}(0, \sigma^2_\zeta)$, and it is potentially correlated with the production shock $\xi_{i,t+1}$. The correlation between production function shock $\xi_{i,t+1}$ and friendship shock $\nu_{i,j,t}$ effectively allow the possibility of selection into peer groups on unobservables.

The specification in (3.4) allows parents and peers to vary from being perfect

[17]See Sacerdote (2001) for the importance of non-linearity in empirical analysis of peer effects

complements to being perfect substitute inputs. Additionally, technology in (3.4) is consistent with the idea of dynamic complementarity of skills evolution, where higher skills today induce higher skills tomorrow (see Cunha and Heckman, 2007). Equation (3.4) allows me to have a flexible specification for the analysis of peer effects in children's skills accumulation. The level of *static* complementarity/substitutability between parents and peers is defined by α_3, while the *dynamic* complementarity between investment and future peers comes from the self-productivity of skills to beget skills, i.e. the complementarity of future peers with future skills.

3.4.3 The Child's Problem

At the beginning of every period t, each child i decides to become friends with another child j, independently of their parents. I define the process of children's network formation as a function of the children's skills, their exogenous characteristics (X_i, X_j) and a vector of environment-specific characteristics (O_e) such as race composition and population size [18] The network-formation process takes place only within the same environment, generating social isolation between children of different areas. Empirically, this is consistent with the fact that I observe friendships only within the same school. At the same time, within the same environment, the children's meeting process is "frictionless", meaning that each child meets the other children in that social context. However, friendships are endogenously formed by the joint decision of children.

Following a similar specification as in Christakis *et al.* (2010), Goldsmith-Pinkham and Imbens (2013) or Graham (2017), child i's utility to become friends with child j at time t is:

[18]Figure B.1 shows that the probability of forming a friendship is a function of the population size of the school.

$$u^C_{i,j,t} = \delta_1 + \delta_2 \ln h_{j,t} + \delta_3 X_j + \delta_4\, \mathbb{1}(X_i = X_j) + \delta_5\, (\ln h_{i,t} - \ln h_{j,t})^2 + \delta_6 O_e + \delta_7\, t - \nu_{i,j,t}\,, \tag{3.5}$$

where $\nu_{i,j,t}$ is a utility shock for the formation of peer groups and δs are the parameters associated with each variable affecting the friendship decision. This utility function has a similar representation to the one used in the demand for products in the literature of industrial organization, where individuals may have direct preferences over the attributes of the potential partners. This part is captured by δ_2 and δ_3. The difference from that literature comes from the other component of the utility function $\delta_4\, \mathbb{1}(X_i = X_j) + \delta_5\, (\ln h_{i,t} - \ln h_{j,t})^2$, which captures the propensity of children to interact with children who are alike both in terms of skills and other individual characteristics. This phenomenon is called homophily bias in the network literature (see Jackson, 2008; Christakis and Fowler, 2009). A specific age (t) effect in the formation of peer groups is captured by δ_7. Hence, each child i solves the following problem for each potential future peer j at each period t:

$$V^C(h_{i,t}, X_i, h_{j,t}, X_j, O_e) = \max \left\{0\,,\ u^C_{i,j,t}\right\} \tag{3.6}$$

where I normalize to zero the value to have no friend. [19] Child i and child j become friends if both children find the friendship beneficial, i.e.:

$$L^*_{i,j,t} = \begin{cases} 1 \quad \text{if} \quad u^C_{i,j,t} > 0 \quad \& \quad u^C_{j,i,t} > 0 \\ 0 \qquad \text{otherwise} \end{cases} . \tag{3.7}$$

The model in (3.6) does not consider a potential decrease of marginal returns to additional friendships as the number of friends increases. Perhaps an alternative way

[19]The underlying assumption is that the outside option is common for different types of children and for different meetings.

of modeling friendship formation could consider children with a limited endowment of time who optimally allocate their time in interacting with other children. In this case, the social interactions would be limited by this time constraint and children would have to coordinate relative to their own time constraints. While the latter model seems more realistic, the lack of data on time allocation between peers as well as the additional computational burden in the model directed me to the model in (3.6). Model (3.6) represents a simple and flexible way of capturing the endogeneity of peer groups and the main driving forces affecting friendship formation. For convenience, let us define W as the set of variables of the utility function in equation (3.5):

$$W_{i,t} = \left[1,\ h_{j,t},\ X_j,\ \mathbb{1}(X_i = X_j),\ (\ln h_{i,t} - \ln h_{j,t})^2,\ O_e, t\right]\ .$$

The conditional probability for child i and child j to become friends, under the independence assumption between utility shocks is then:

$$Pr(L^*_{i,j,t} = 1) \equiv P_{i,j,t}\left(h_{i,t}, X_i, h_{j,t}, X_j, O_e\right) = Pr(\nu_{i,j,t} \leq W'_{i,t}\delta) \cdot Pr(\nu_{j,i,t} \leq W'_{j,t}\delta) \tag{3.8}$$

where the probability of two children connecting together can be higher (lower) in a case where the two children have the same characteristics (δ_4). Also, if δ_5 is negative, a higher difference in skills will reduce the probability of the two children deciding to connect, while a positive coefficient will increase it. In other words, the sign of the coefficient will reflect a positive or negative assortative matching between children with respect to their skill development. All children compute the utility of forming a friendship with other children within the social network, and the entire social network graph is determined. The set of probabilities between different children as in (3.8) forms the probabilities of the possible networks in each environment e.

In this framework, children do not directly make investments in themselves (e.g. a study effort decision problem), which perhaps could depend on their own skills as well as the effort their peers make. This hypothetical modeling choice would consider a specific aspect of social interactions either emerging from a conformity effect or from strategic complementarities in skill formation between children (see Blume *et al.*, 2015). In the next section, I will describe the technology of skill formation and how children's skills evolve over time. The dynamics of skills depend on both a child's own level of skills as well as the level of peers' skills, capturing the potential effects of studying effort, allowing for other more general peer effects and reducing the computational burden of the model.

3.4.4 The Parenting Problem

Preferences

I assume that each parent in family i at any period t has preferences over their own consumption ($c_{i,t}$) and over the skills of their children ($h_{i,t}$), while they do not receive any direct utility from time spent with their children ($I_{i,t}$). I additionally assume that preferences are stable over time. Parental investments are made dynamically to foster children's skills over time. This specification is in line with the recent literature in child development (see Cunha, 2013b; Del Boca *et al.*, 2014b, 2016; Mullins, 2016; Gayle *et al.*, 2015, 2016; Caucutt and Lochner, 2017). There is only one decision maker regarding parental investments, as I assume that the mother–father interactions occur at a prior stage . At any period, each parent is endowed with τ units of time and decides how to allocate this endowment between working ($\tau - I_{i,t}$) and parenting ($I_{i,t}$). Finally, I assume that the utility function for parents i at period t is as follows:

$$u^P(c_{i,t}, h_{i,t}) = \frac{c_{i,t}^{1-\gamma_1} - 1}{1-\gamma_1} + \gamma_2 \frac{h_{i,t}^{1-\gamma_3} - 1}{1-\gamma_3}\,, \tag{3.9}$$

where γ_2 ¿ 0. [20] The specification in (3.9) underlines the main parent's trade-off: the benefit of higher children's skills at the cost of their own foregone consumption. Another model choice I could use would be to define parental investments in terms of the effort parents need to make in order to invest in their children's skills, and the associated utility cost of that effort. In this case, the trade-off would be between the altruistic benefit of fostering children's skills and the disutility of the required effort. The two specifications are isomorphic. [21] Finally, family income is defined by the mother's labor and non-labor income. I assume that both the mother's hourly wage ($w_{i,t}$) and non-labor income ($y_{i,t}$) are a function of her skills (m_i). The exact wage and non-labor income specifications are described in Section 3.5, where the relationship between a mother's skills and her non-labor income aims to capture the potential effect of her skills in assortative mating and family formation.

Terminal value

I assume that the parent' s problem ends when children reach 16 years of age. This assumption can be read as the fact that children leave the household at 16 years old or that after that age parental investments become unproductive. I think of the child's final skills at 16 as an initial condition of another developmental process which I am

[20]Preferences over consumption or a child's outcomes are logarithmic functions if, respectively, γ_1 = 1 or γ_3 = 1.

[21] In this specification, I abstract from the labor–leisure decision margin. The main reasons are related to the lack of data on Add Health about parent's leisure choices, which would allow me to identify the elasticity of leisure choice with respect to changes in peers' skills. Additionally, adding another endogenous variable to the model would increment its computational burden. For this reason, my policy counterfactual experiments will only focus on policy-induced change in peers' skills and the associated change in the return of parenting, while I will abstract from any welfare analysis and/or changes in family resources, which would need additional important predictions on the family time allocation between working, leisure and parenting

not modeling here, such as finishing high school, starting a job or going to college. Hence, I allow a possible change in parental preferences over the skills of the final childhood period. I am defining the terminal value for the parent i with respect to children's skills as follows:

$$V_4^P(h_{i,4}) = \gamma_4 \cdot \frac{h_{i,4}^{1-\gamma_5} - 1}{1-\gamma_5} \;, \tag{3.10}$$

where both γ_3 and γ_4 are free parameters that potentially differ from the altruistic parameter specified in (3.9).

The recursive representation of the parent's problem

The two endogenous-state variables of the problem are, respectively, the child's skills ($h_{i,t}$, individual-state variable) and peers' skills ($\overline{H}_{i,t}$, aggregate-state variable). The dynamics of the network state within each environment are taken as given from the parent' s perspective. Parents form expectations with respect to the next period's average skills of peers. Different types of peers in the next period affect the return of investment in their offspring today through the dynamics of the child's skills. Anticipating the discussion on the equilibrium in Section 3.4.5, the consistency condition in this economy is that the expectations about the next period's peers' skills will be consistent with the transition probabilities generated by the endogenous network formation from the child's problem (see Section 3.4.3). [22] The parent's problem can be represented as follows:

$$V_t^P(h_{i,t}, \overline{H}_{i,t}) = \max_{I_{i,t} \in [0\,,\tau]} \; u^P(c_{i,t}, h_{i,t}) + \beta\, E\left[V_{t+1}^P(h_{i,t+1}, \overline{H}_{i,t+1}) \;\middle|\; h_{i,t} \right] \tag{3.11}$$

[22] The consistency condition between the individual behavior of parents and the aggregate distribution of skills in the network is the analogue of the consistency condition used to solve recursive competitive equilibrium in macroeconomic models with aggregate externalities.

$$s.t. \quad c_{i,t} = (\tau - I_{i,t}) \cdot w_{i,t} + y_{i,t}$$

where $\beta \in (0,1)$ is the discount factor, while the consumption $c_{i,t}$ is a function of earnings (through the labor supply $\tau - I_{i,t}$) and the non-labor income $y_{i,t}$. Parents are uncertain about the production shock as well as their child's peer group in the next period, which will affect their future investment productivity. The law of motion for the next period's child's skills (or technology) is defined in (3.4), while the law of motion for peer effects ($\overline{H}_{i,t+1}$) is as follows:

$$Pr\left(\overline{H}_{i,t+1} = \frac{1}{\sum_{j \in N_e} L_{i,j,t+1}} \sum_{j \in N_e} L_{i,j,t+1} \cdot h_{j,t+1}\right) = \prod_{i=1}^{N_e} \pi_{i,j,t+1}^{L_{i,j,t+1}} (1 - \pi_{i,j,t+1})^{1-L_{i,j,t+1}} \tag{3.12}$$

where $L_{i,j,t}$ is an indicator function equal to one if child i and child j are friends, and zero otherwise. Given the conditional independence assumption about the formation of peer groups, the stochastic law of motion in (3.12) represents the probability distribution of N_e independently and differently distributed Bernoulli random variables (friendships), where $\pi_{i,j,t+1}$ is the relative probability of that friendship happening.

3.4.5 Equilibrium of the Network Economy

In this section, I describe the equilibrium of the economy. For computational reasons, I restrict my attention only to the (short-memory) Markovian equilibrium, where the parent's and child's policy functions depend only on the current realization of state variables during each period. Nevertheless, a desirable property of the Markovian equilibrium is that, in this framework, it generates non-ergodic skill dynamics (i.e. the property of skill formation depending on the history of developmental inputs throughout childhood), a key mechanism in explaining diverging patterns in outcome inequality in children. In fact, as I will explain later in the policy analysis, moving children at age 13 to a different environment predicts *persistent* effects in the dynamics of children's skills. Skills beget skills through many mechanisms: self-production, better peers and higher investments.

Alternative classes of equilibrium concepts consist of longer-memory equilibria where parents' and children's behavior is explained both by the realization of the current states as well as by the equilibrium path history that led to that state. This would lead to even stronger dynamic equilibrium spillover effects of skills, because it would strengthen the role of social interactions in explaining children's developmental differences through the determinants of the equilibrium path of a child's development.

Parents and children have two different and separate problems. In particular, parents observe the current realization of their offspring's peer groups and then form expectations about the next period's peer groups when deciding on today's investment. Parents take as a given both the dynamics of network structure as well as the distribution of children's skills within the social network. At any point in time, children decide about their friends, generating the network of friendships. Then, parents decide how much time to invest in their offspring, forming expectations with respect to the next period's distribution of peers' skills. Given that the next period's distribution of peers' skills is an endogenous object in the model, the equilibrium characterization will take into account the consistency condition between parents' expectations and both skills and network equilibrium realizations.

Definition 3 *A Markovian equilibrium of the network economy is a set of functions* $\{I_t(\cdot),\ \mathbb{1}_t(\cdot)\}_{t=13}^{16}$ *such that:*

1. $\nVdash^*(\cdot)$ *solves the child's problem in (3.6), for every period* t*,*

2. $I_t^*(\cdot)$ *solves the parent's problem in (3.11) , for every period* t*,*

3. The probability for the formation of peer groups is consistent with the skills dynamics generated by the parental optimal behavior:

$$\pi_{i,j,t+1} = P_{i,j,t+1}\,(h^*_{i,t+1}, X_i, h^*_{j,t+1}, X_j, O_e) \quad \textit{for all}\, i,j,t$$

where

$$h^*_{i,t+1} = {h^*_{i,t}}^{\alpha_1} \cdot \left[\alpha_2\,(I^*_{i,t}(\cdot))^{\alpha_3} + (1-\alpha_2)\left(\frac{1}{\sum_{j\in N_e} L^*_{i,j,t}} \sum_{j\in N_e} L^*_{i,j,t}\cdot h^*_{j,t}\right)^{\alpha_3}\right]^{\frac{\alpha_4}{\alpha_3}} \cdot A_{i,d,s,t}\cdot exp(\xi_{i,t+1}),$$

for any production shock realization $\xi_{i,t+1}$. [23]

Definition 3.4.5 provides that both parents and children maximize their utility at each point in time. The last equilibrium condition, the consistency condition, closes the model. In fact, condition (3) implies that the endogenous stochastic network structure, which depends on the skill dynamics, is determined simultaneously in equilibrium from both the parents' and the children's optimal behavior.

Theorem 4 *In this economy, a Markovian equilibrium exists.*

See proof in Appendix B.4.

Theorem 4 formalizes the existence of equilibrium of the model and is the theoretical base of the algorithm used in my simulation-based estimation procedure.

A common feature of any model of social interactions and spillover effects is the potential existence of multiple equilibria. Multiplicity can arise from the presence of

[23] The same condition can be rewritten in terms of expected next period child's skills

$$\pi_{i,j,t+1} = \int P_{i,j,t+1}\,(h^*_{i,t+1}, X_i, h^*_{j,t+1}(\xi_{i,t+1}), X_j, O_e)\,dF(\xi_{i,t+1}) \quad \text{for all}\, i,j,t \;\;,$$

where $F(\xi_{i,t+1})$ is the distribution of the production shocks.

"strong" peer externalities. In this framework, this translates into a strong complementarity between parental investments and peers' skills, which is reflected directly in a low value for the CES complementarity parameter α_3. The possibility of multiple equilibria creates a challenge for the use of standard econometric methods through the presence of an indeterminacy condition in the map from the observed data to the structural parameters. In this case, the estimation procedure would require implementing additional steps to recover the parameters of the model.

Three possible solutions can be considered. First, a common approach in the literature is to assume that the data is generated from a specific equilibrium selection. Generally, the equilibrium selection rule considers the equilibrium with the highest welfare amongst all the possible equilibria (see for example Lazzati, 2015; Fu and Gregory, 2017).

A second approach consists of partially identifying the model. In this case, the econometrician does not need to make any assumptions about the equilibrium selection. A set of moment inequalities arises from the different equilibria and can be used to create bounds on the structural parameters of the model (using, for example, the moment inequalities estimator in Chernozhukov *et al.*, 2007; Andrews and Soares, 2010; Pakes *et al.*, 2015)

A third approach, which is the one I use here, is to determine (if possible) which specific parameter (or set of parameters) is responsible for the presence of multiple equilibria, and for what specific threshold value of that parameter multiplicity arises.

In my model, the key parameter which determines whether the model generates multiple equilibria is the CES parameter of complementarity between parental investments and peers' skills (α_3). A high level of complementarity between parents and peers can generate multiplicity: within each environment, the parental decisions of other parents affect the individual decisions of everybody else, creating possible extreme equilibria where no parents invest at all or all the parents invest the majority of their time in child development. This statement on how peer externalities affect parental investment decisions is a testable prediction. This means that the previous empirical results in Section 3.3 can be considered as a pre-test for multiplicity in this model. Specifically, the fact that by using within-school cross-sectional variation in peers' skills I find a negative effect in parental investment decisions suggests that the two inputs cannot be too complementary; to let the model reproduce that cross-sectional negative relationship between investments and peers' skills, the complementarity between parents and peers should be less than in the Cobb–Douglas case (so the CES parameter α_3 should be bigger than 0). [24] Given this low level of complementarity, I am going to implement a "guess and verify" method, in which I assume that the equilibrium is locally unique for values of $\alpha_3 \in (0, 1]$, and will then computationally verify if the assumption is correct by implementing a monotone method for equilibria computations. This method requires calculation of the two possible extremal equilibria of this economy using the algorithm in Topkis (1979), and then simply comparing them. For the data-driven model's parametrization of $\alpha_3 \in (0, 1]$, I find no evidence of multiple equilibria.

3.5 Econometric Specification

3.5.1 The Latent Factor Models

In line with the recent literature on child development (see Cunha *et al.*, 2010; Agostinelli and Wiswall, 2016a; Attanasio *et al.*, 2017a,b,c), I implement a dynamic latent factor model to map the key unobserved variables of the model into data. The

[24] A recent work of Datta *et al.* (2017) shows that in a similar environment, a macro growth model with externalities, the unique equilibrium is proved in a case where externality is not big, like in the case of constant return to scale in the production technologies.

factor model overcomes the main problem in the analysis of skill formation: mismeasurement of skills and the arbitrariness of test-score scales relative to the scale of skills. For both mothers' and children's skills, I follow the latent factor model implemented in Agostinelli and Wiswall (2016a) as follows:

$$Z^h_{i,t,k} = \mu^h_{t,k} + \lambda^h_{t,k} \cdot \ln h_{i,t} + \epsilon^h_{i,t,k}$$
$$Z^m_{i,k} = \mu^m_{t,k} + \lambda^m_k \cdot \ln m_i + \epsilon^m_{i,k} \tag{3.13}$$

The index k is for indexing each of the multiple measurements (proxies) Z for each latent factor. Because the location and scale of skills can differ from the arbitrary location and scale of the proxies I use, I implement the factor model in (3.13) with free measurement parameters (μ and λ). Finally, the noises in (3.13) have a mean of zero in any period and for any measure. [25] I assume the common independence conditions about the measurement error to hold. These conditions include the independence of measurement noises with both a child's and mother's latent skills and between different measures of skills, as well as between different children and over time (for more details, see Appendix B.3).

Mapping data to the distribution of latent investment is challenging due to the nature of the proxies included in Add Health. Add Health asks children whether they have been engaged in specific activities with their mothers in the last four weeks. Examples of activities are "gone shopping," "played a sport," "gone to a movie, play, museum, concert, or sports event" or "had a talk about a personal problem." Each question requires a "yes" or "no" answer, generating a set of binary proxies for investments defined by $Z^I_{i,k} \in \{0,1\}$, where i and k indexes are relative to the child and the specific question. These measures can be considered indicators as to

[25] Given the intercept $\mu_{t,m}$, the assumption of a mean of zero $\epsilon_{t,m}$ errors is without loss of generality.

whether parents spent some time with their children or not. Hence, each measure of investment can be thought of as a Bernoulli random variable with probability $p_k(I_{i,t})$, a function of the latent investment. I adopt a similar approach as in Del Boca *et al.* (2014b), and I consider a specific parametric distribution for $p_k(I_{i,t})$, which is a Beta distribution with parameters Beta($\alpha + Z^I_{i,t,k}$, $1 + \beta$ - $Z^I_{i,t,k}$). [26] I can now draw p_k from this distribution to recover, for each measure k, the latent distribution of parental investments $F_k(I_{i,t})$. Let $\widehat{p}_{i,t,k}$ be the draw from the parametric distribution for some observation i at time t, and I can impute the level of investment by inverting the probability function at $\widehat{p}_{i,t,k}$ (assuming the inverse exists):

$$I^k_{i,t} = p_k^{-1}(\widehat{p}_{i,t,k}). \tag{3.14}$$

where α and β define the location and scale for the latent investments. [27] To assure that imputed levels of investments are constrained between 0 and τ (the max time in the model), I map each specific probability into the fraction of time spent with children ($\frac{I_{i,t}}{\tau}$). Each probability of observing a measure of investment equal to one increases with respect to the fraction of parental investments (higher parental investments lead to a higher probability of observing, in data, children involved in activity "k" with their parents). Moreover, a desirable property for the probability function is that $\lim_{I_{i,t}\to 0} p_k(I_{i,t}) = 0$, $\lim_{I_{i,t}\to \tau} p_k(I_{i,t}) = 1$. This means that once the fraction of invested time goes to zero or to one, the probability of observing a parent involved in that specific activity goes to zero or to one. For all these reasons, I choose the following simple functional form, which respects all the required properties:

[26] Del Boca *et al.* (2014b) use the same approach to address the issue of measuring continuous skills with a discrete test score as well as the possibility of measurement error of the measured score. Additionally, this method overcomes the problem of a measurement floor and ceiling of test scores

[27] In contrast to the skills case, latent time investments have a well-known location and scale, which is in terms of time units or fractions of total time. I will measure this information by looking at the time allocation of parents on the American Time Use Survey (ATUS).

$$p_k(I_{i,t}) = \left(\frac{I_{i,t}}{\tau}\right)^{\lambda^I_{t,k}} \quad \lambda^I_{t,k} > 0, \tag{3.15}$$

where the parameter $\lambda^I_{t,k}$ is the loading factor for each activity m.

3.5.2 Parametric Assumptions of the Model

In this section, I illustrate the assumptions I am making in order to parametrically estimate the model. I consider three different types of neighborhood quality according to the income distribution in the data. For each school I have in Add Health, I compute the within-school mean family income and then assign each family to the low-, medium- or high-income neighborhood type according to the median school income. [28] This distinction is made based on the terciles of income-distribution (33^{th} percentiles, between the 33^{th} and 66^{th} percentile and above the 66^{th} percentile). [29] Families first draw their race (I allow race to be either black, Hispanic or other) and their neighborhood type from a joint distribution P(d,r), which I directly estimate from the data. [30] After race and neighborhood are drawn, I assume that the initial distribution of a family's log-skills (mother and child) are drawn from a conditional bivariate distribution:

$$(\ln h_{i,1}, \ln m_i) \sim N(\eta_{r,d}, \Sigma_{r,d}) \tag{3.16}$$

where I allow mean and variance of latent skills to vary by race and neighborhood. The specification allows me to have a flexible framework to capture potential sorting of families within specific environments. Taking into consideration this kind

[28] The results do not change if I consider the median family income within the Census Tract where families live

[29] For further details on the descriptive statistics for each of the neighborhood types, see Table B.1.

[30] This empirical distribution represents the probability for each of the nine possible combinations of race and neighborhoods and is directly observed and estimable from the data

of assortative pattern is fundamental in order to properly identify the peer effects in child development.

I also make some parametric assumptions about the measurement error equations in (3.13). I assume that the measurement noises for each measure are mean-zero normally distributed $\epsilon_k \sim$ N$(0,\sigma^2_{\epsilon,k})$ for any measure k at any age t.

The TFP term in the technology of skill formation is composed of two parts: neighborhood-quality effects and school-quality effects. For each neighborhood type d, I assume there is a distribution of school-quality families drawn together with their neighborhood realization: $A_s \sim$ N$(\eta_{s,d},\sigma^2_{s,d})$. To generalize the school effects in skill dynamics, the parametric functional form for the technology of skill formation I bring to data is:

$$h_{i,t+1} = {h_{i,t}}^{\alpha_1} \cdot \left[\alpha_2 \left(I_{i,t}\right)^{\alpha_3} + (1-\alpha_2)\left(\overline{H}_{i,t}\right)^{\alpha_3} + \alpha_5 \, A_s^{\alpha_3}\right]^{\frac{\alpha_4}{\alpha_3}} \cdot A_{d,t} \cdot exp(\xi_{i,t+1}) \,, \tag{3.17}$$

where the share parameter of A_s (α_5) is normalized to 1 given that the variance of the latent school fixed effects is a free parameter estimated directly in the data ($\sigma^2_{s,d}$). Technology in (3.17) generalizes the school effects to have individual specific elasticities of skill production with respect to school quality. The environment-specific TFP term is assumed to be a function of the neighborhood quality and child's age as follow: $A_{d,t}$ = exp($\gamma_{0,tfp}$ + $\gamma_{1,tfp}$ · d + $\gamma_{2,tfp}$ · t). Additionally, I am assuming some parametric form for both shocks of skill production and preference for children's peer groups. Production shocks are assumed to be mean-zero normally distributed $\xi_{i,t+1}$ $\sim$ N$(0,\sigma^2_{\xi})$, while a child's preference shocks for friendships, defined in eq. (3.5), are distributed as a standard logistic . This implies that the probability that child i and child j become friends at any time t is:

$$P_{i,j,t} \equiv Pr(L^*_{i,j,t} = 1) = \frac{exp(W'_{i,t}\delta)}{1 + exp(W'_{i,t}\delta)} \cdot \frac{exp(W'_{j,t}\delta)}{1 + exp(W'_{j,t}\delta)} \quad (3.18)$$

where $W_{i,t}$ and $W_{j,t}$ are the set of variables affecting the decision and are defined in eq. (3.4.3), while δ are the common utility parameters.

Finally, I assume that both the mother's hourly wage ($w_{i,t}$) and non-labor income ($y_{i,t}$), at any period t, are defined as a function of the mother's skills in a linear fashion, as in the classic Mincer wage equation (see Mincer, 1958):

$$\begin{bmatrix} \ln w_{i,t} \\ \ln y_{i,t} \end{bmatrix} = \begin{bmatrix} \kappa_{1,0} \\ \kappa_{2,0} \end{bmatrix} + \ln m_i \cdot \begin{bmatrix} \kappa_{1,1} \\ \kappa_{2,1} \end{bmatrix} + \begin{bmatrix} \epsilon^w_{i,t} \\ \epsilon^y_{i,t} \end{bmatrix}$$

where $\epsilon^w_{i,t}$ and $\epsilon^y_{i,t}$ are measurement noises.

3.6 Identification of the Model

As explained above, both mother's and child's skills as well as parental investments are assumed to be unobserved. With this in mind, the goal is to identify peer effects on parental investment decisions and skill formation, dealing with the endogeneity of the peers network formation. In this section I describe how I approach this task. The identification of the wage and non-labor income process is standard: in both cases, once the scale and location for the latent mother's skills are identified (see Section 3.6.1), it is a simple linear errors-in-variables models. In the next sections, I focus more on the more challenging part of the model for identification. I am following the logical order from the model: (i) initial conditions; (ii) network formation process; (iii) static and dynamic complementarity between parents and peers.

3.6.1 Identification of the Initial Conditions

The main challenge in the identification of the family skills distribution comes from the fact that skills are unobserved and they have not a natural scale and location. Recalling that the initial conditions are assumed to be normally distributed,

$$(\ln h_{i,1},\, \ln m_i) \sim N\left(\eta_{r,d},\, \Sigma_{r,d}\right),$$

where $\eta_{r,d} = \begin{bmatrix} \eta^h_{r,d} \\ \eta^m_{r,d} \end{bmatrix}$ and $\Sigma_{r,d} = \begin{bmatrix} {\sigma^h_{r,d}}^2 & \sigma^{h,m}_{r,d} \\ \sigma^{h,m}_{r,d} & {\sigma^m_{r,d}}^2 \end{bmatrix}$ are, respectively, the vector of means and the variance–covariance matrix for skills for a specific neighborhood type (d) and race (r). I normalize the scale and location of skills for a subgroup of the population. Without loss of generality, I normalize the distribution of skills for white (r=w) families living in the lowest-income neighborhood ($d = 1$) as follows:

Normalization 2 *Initial period normalization:*

- $\eta_{w,1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$
- $\Sigma_{w,1} = \begin{bmatrix} 1 & \sigma^{h,m}_{w,1} \\ \sigma^{h,m}_{w,1} & 1 \end{bmatrix}$

Under normalization 2, I am able to identify the remaining measurement equation parameters and the initial joint distribution of skills for each race and each neighborhood type. The initial period loading factors (λ) and the location parameters (μ) are identified as follows:

$$[\lambda^h_{1,k},\ \lambda^m_k] = \left[\frac{Cov(Z^h_{i,1,k}, Z^h_{i,1,j})}{Cov(Z^h_{i,1,j}, Z^h_{i,1,1})},\quad \frac{Cov(Z^m_{i,k}, Z^m_{i,j})}{Cov(Z^m_{i,j}, Z^m_{i,1})}\right] \text{ for } k \neq j \text{ and for } k, j \neq 1 \tag{3.19}$$

$$[\mu^h_{1,k},\ \mu^m_{1,k}] = \left[E[Z^h_{i,1,k}|r = w, d = 1],\quad E[Z^m_{i,k}|r = w, d = 1]\right] \text{ for all } k. \tag{3.20}$$

Using both (3.19) and (3.20), it is possible to identify the means and the variance-covariance matrix of the latent skills for all the rest of race and neighborhood type combinations in the following way:

$$\eta_{r,d} = \left[\frac{E[Z^h_{i,1,k}|r, d] - \mu^h_{1,k}}{\lambda^h_{1,k}}, \frac{E[Z^m_{i,k}|r, d] - \mu^m_k}{\lambda^m_k}\right] \text{ for all } (r, d) \neq (w, 1) \tag{3.21}$$

$$\Sigma_{r,d} = \begin{bmatrix} \dfrac{Cov(Z^h_{i,1,k}\,,\ Z^h_{i,1,j}\ |r, d)}{\lambda^h_{1,k}\ \lambda^h_{1,j}} & \dfrac{Cov(Z^h_{i,1,k}\,,\ Z^m_{i,j}\ |r, d)}{\lambda^h_{1,k}\ \lambda^m_j} \\ \dfrac{Cov(Z^h_{i,1,k}\,,\ Z^m_{i,j}\ |r, d)}{\lambda^h_{1,k}\ \lambda^m_j} & \dfrac{Cov(Z^m_{i,k}\,,\ Z^m_{i,j}\ |r, d)}{\lambda^m_k\ \lambda^m_j} \end{bmatrix} \text{ for all } (r, d). \tag{3.22}$$

Repeating the identification of mean and variance–covariance matrix for all the neighborhood types, and for each race, I am able to identify the distribution of initial conditions in the economy. [31] Finally, I identify the variance of the measurement errors for each measure:

$$\sigma^2_{\epsilon^m,k} = Var(Z^m_{i,k}) - (\lambda^m_k)^2 Var(\ln m_i)$$

$$\sigma^2_{\epsilon^h,t,k} = Var(Z^h_{i,t,k}) - (\lambda^h_{t,k})^2 Var(\ln h_{i,t}) \text{ for all } t \tag{3.23}$$

where each variable on the right-hand side of the equations in (3.23) is already identified. The repeated cross-sectional dimension of the Add Health data allows

[31]Notice that the identifying assumption here is that the measurement parameters are common for all the children. This assumption is common in the literature of latent factor models.

me to observe the same test score administered to children at different ages. Hence, following the approach in Agostinelli and Wiswall (2016a), I assume the PPVT to be an age-invariant measure – that is, the latent skills load into the PPVT in the same manner through the ages of consideration (13 to 16). [32] As shown in Agostinelli and Wiswall (2016a), this assumption is sufficient to identify an unknown TFP term and an unknown return to scale in the CES technology case. [33]

3.6.2 Identification of the Formation of Peer Groups

Using the measurement system defined in (3.13), I can rescale the observed measures for skills to be:

$$\widetilde{Z}^h_{i,t,k} \equiv \frac{Z^h_{i,t,k} - \mu^h_{t,k}}{\lambda^h_{t,k}} = \ln h_{i,t} + \frac{\epsilon^h_{i,t,k}}{\lambda^h_{t,k}} \tag{3.24}$$

where $\widetilde{Z}^h_{i,t,k}$ represents a monotone transformation of raw test scores that permit consideration of the data on a comparable scale as the latent skills. Each transformed measure is composed by two elements: the factor and the noise. For this reason, in order to use this information to identify the peer-group formations, we need to deal with the presence of this measurement error. In particular, I exploit the identified parametric distribution of the measurement errors to integrate out the noise from the data.

Following the notation in (3.4.3), let us define W to be the set of variables of the utility function of child i to become friends with child j as in equation (3.5) and $\widetilde{W}$ to be the analogue measured in the data:

[32] Agostinelli and Wiswall (2016a) defines an age-invariant measure to be a repeated achievement test, the assessment of which would not depend on a child's age, but on their cognitive development

[33] It also allows me to identify the measurement parameters for all remaining proxies for skills through all the periods of the model

$$W_i = \left[1,\ h_{j,t},\ r_j,\ \mathbb{1}(r_i = r_j),\ (h_{i,t} - h_{j,t})^2,\ O_e, t\right],$$

$$\widetilde{W}_i(\epsilon^h_{i,t,k}, \epsilon^h_{j,t,k}) = \left[1,\ (\widetilde{Z}^h_{j,t,k} - \frac{\epsilon^h_{j,t,k}}{\lambda^h_k}),\ r_j,\ \mathbb{1}(r_i = r_j),\ \left((\widetilde{Z}^h_{i,t,k} - \frac{\epsilon^h_{i,t,k}}{\lambda^h_k}) - (\widetilde{Z}^h_{j,t,k} - \frac{\epsilon^h_{j,t,k}}{\lambda^h_k})\right)^2,\ O_e, t\right]. \tag{3.25}$$

The conditional probability of child i and child j becoming friends, given their skills and race, is as follows:

$$\begin{aligned} P_{i,j,t} &= Pr(\nu_{i,j,t} \leq W_i'\delta) \cdot Pr(\nu_{j,i,t} \leq W_j'\delta) \\ &= \int Pr\left(\nu_{i,j,t} \leq \widetilde{W}_i(\epsilon^h_{i,t,k}, \epsilon^h_{j,t,k})'\delta\right) \cdot Pr\left(\nu_{j,i,t} \leq \widetilde{W}_j(\epsilon^h_{j,t,k}, \epsilon^h_{i,t,k})'\delta\right)\ d\Phi(\epsilon^h_{i,t,k}, \epsilon^h_{j,t,k}; \Sigma_{\epsilon,t}) \end{aligned} \tag{3.26}$$

where $\Phi(\cdot; \Sigma_{\epsilon,t})$ is the bivariate normal CDF of the measurement errors with zero mean and a variance–covariance matrix $\Sigma_{\epsilon,t} = \begin{bmatrix} \sigma^2_{\epsilon^h,t,k} & 0 \\ 0 & \sigma^2_{\epsilon^h,t,k} \end{bmatrix}$, which is already identified in (3.23).

Equation (3.26) shows that I can identify the parameters of the model of peer-group formation by mapping the noisy measures of children's skills available in data to information about the children's latent skills and their effects on children's friendship decisions.

3.6.3 Identifying Peer Effects in Child Development

Given the complexity of the above dynamic equilibrium model, it is important to develop an easy intuition behind the source of variation which will lead to identifying peer effects on child development. In particular, one helpful source of identifying variation comes from the endogenous response of parental investments to peer quality, allowing us to infer the degrees of both *static* and *dynamic* complementarity between parents and peers in producing a child's skills. These parent response elasticities are key for identification because they map directly into the deep structural parameters of the technology of skill formation and are the key margin of interest for policy predictions. The main object of interest for identification is the full joint distribution of inputs of development $\Psi(\Omega)=\left\{\left\{I^*_{i,t}, h_{i,t}, \overline{H}^*_{i,t}\right\}_i\right\}_t$, where Ω is the set of structural parameters left to identify, the peers' skills spillover determined in the equilibrium social network is defined as $\overline{H}^*_{i,t}$, and the star defines the equilibrium realization.

Static Complementarity

A useful set of moments to identify the static complementarity between parents and peers in skill formation is the partial correlations of parental investments and peers' skills. In fact, an intuitive prediction of my model is a positive (negative) cross-sectional relationship between parental investments and peers in the presence of a high degree of *static* complementarity (substitutability) between parents and peers in child development. Once peer quality increases, the return on investing in child development increases as well, making parents more productive in parenting. On the other hand, parents have an incentive to decrease their costly investments once an alternative substitute input (peers) increases. Hence, the *static* complementarity/substitutability between parents and peers is identified by the cross-sectional variation between investments and peers' skills.

A specific but helpful representation of these moments is given by the linear regression in (3.2), which provides useful information about the variation of investment decisions with respect to the type of peers. Specifically, the coefficient β_2, which represents the effect of peers' skills on investment decisions, is informative about the degree of *static* complementarity/substitutability (α_3) between parental investment and peers in the production of children's skills.

Dynamic Complementarity

The *dynamic* complementarity between parents and peers is identified by looking at the variation in parental investments induced by changed expectations for future peers' skills. For example, as a thought experiment, consider two children (and their relative families) who are alike in all dimensions including their peer groups, but only one of them is assigned *permanently* (treatment) to a different (better) peer group, while the other child (control) has no change in peer composition. Given the permanent nature of the experiment, the two parents now observe today's peers and have different expectations about tomorrow's peers. Again, the model's prediction is helpful in fixing ideas: an increased expectation of tomorrow's peers' skills would change the return on today's investments. Current investments would foster the child's skills of tomorrow, thus benefiting more from better future peers. Hence, the higher the dynamic complementarity, the higher the difference in parental investments induced by the policy between the control and the treatment of children: the treatment induces higher investments because of the effects of expectations of future peers.

My empirical identification strategy is based on the instrumental variable approach explained in Section 3.3.4, which represents the quasi-experimental analogue of the identifying variation considered in the above thought experiment. Specifically, in order to estimate the causal effects of permanent peers changing investment decisions, I use the random realization of different cohort composition of children within the same school to analyze how permanent changes in peers affect parental investments. The idea behind this identification strategy is simple. The cohort realization *permanently* defines the choice set for the children's network formation. Hence, different cohort compositions shift parents' expectations about future peer groups, allowing us to identify the causal effect of a permanent change in peer groups on investment decisions (β_2) and, consequently, the dynamic complementarity of investments and peers. [34]

3.7 Estimation

I implement a two-step estimation algorithm to alleviate the estimation burden of this model. In the first step, I estimate the initial conditions, the wage and the family income process, as well as the distribution of school quality for low/medium/high neighborhood types. [35], [36] The second step of estimation is focused on the rest of the model: preferences, technology and parameters of the network formation. I use a simulation-based estimation technique which allows me to simulate the distributions of investments and skills over the entire childhood and replicate the statistics (M) I observe in Add Health data. In greater detail, I simulate the equilibrium dynamics of investments, skills and networks and use the realized simulated data set to compute the analogous statistics (M_S) observed in the data. Hence, the second-step estimator is a simulated method of moments (SMM) estimator:

$$\widehat{\Omega} = \arg\min_{\Omega} \left(M - M_S(\Omega)\right) W \left(M - M_S(\Omega)\right) \tag{3.27}$$

where Ω is the set of structural parameters previously described and W represents

[34] This identification strategy does not require that friendships are only formed within a unique cohort. It only requires that changes in cohort composition also change the peer-choice set between cohorts. Empirically, most of the friendship nominations are within the same cohort.

[35] The school quality distributions are identified estimating the school fixed-effects distribution from a value-added model of child development. This method is simple and alleviates the computational burden of estimating different school quality parameters in the second step of the estimation. See Appendix B.5 for further details.

[36] Identification of latent skills requires at least three proxies for each latent variable. In the case of a mother's skills, Add Health provides only information about a mother's years of education. I use the NLSY79 sample to recover the loading factor and location parameter for this proxy.

the weighting matrix. [37] In the estimation procedure, I set the weighting matrix to be the inverse of the diagonal variance–covariance matrix of moments computed by bootstrapping the data. The selected moments include: (i) first- and second-order moments for the conditional distribution of skills by race and neighborhood type; (ii) the set of auxiliary coefficients from the two auxiliary regressions in (3.2), as well as auxiliary coefficients for the technology of skill formation (in an *indirect inference* fashion); (iii) the set of moments about network formation, including the homophily bias index for skills and race for each neighborhood type, which I described during the previous empirical analysis in Section 3.3.3.

3.7.1 Structural Estimates

Network formation parameters

Table 3.3 shows estimates for the formation of peer groups. The unconditional effects of age, race or skill level on the formation of peer groups are negligible (the unconditional numbers of friends by race and by skills are similar). Race and skills play a role through homophily bias in friendship formation, affecting the composition (not the quantity) of friends. Qualitatively, coefficients in Table 3.3 are consistent with the empirical evidence of homophily bias both in race and in skills. Specifically, I find that for any race, the fact that the other child is of the same race is highly predictive about the friendship realization between the two children. The coefficients associated with being of the same race are, respectively, 0.76, 0.70 and 0.60 for black, Hispanic

[37] The choice of the simulated method of moments with respect to a likelihood-based method is due to three reasons: first of all, the SMM approach overcomes the additional source of computational burden which arises from the multi-dimension integration problem associated with the maximum-likelihood estimator of this model. Secondly, because of the data structure, I observe each child only for two consecutive waves (with a temporal distance of one year) in Add Health, making the SMM a more flexible estimator in combining the information of skills dynamics from the dataset. Finally, the SMM does not require any assumption about the cross-sectional distribution of the children's skills over childhood.

and white children. The higher coefficient is for black children. The probability for the formation of peer groups depends on whether children have similar skills. In particular, the higher the difference in children's skills, the lower the probability of the two children becoming friends, with a coefficient of -0.038. These homophily bias effects get bigger in context with a lower fraction of minorities, which is shown by how the two coefficients of homophily bias interact with the total fraction of black and white children (0.042 and -0.063, respectively). I find that the correlation of unobservable heterogeneity in the formation of peer groups with the production shocks is -0.40, while its standard deviation is 0.11. I do not find relevant effects of the specific race or skills level in the probability level (only through homophily bias).

To better interpret the estimates, I plot the marginal probabilities of two children becoming friends over the spectrum of children's skills and for different races of children living in the poorest environment (see Figure 3.4). For a black low-skilled child (within the first quintile of skills distribution), the probability of becoming friends with white children is four times lower than with a same-race counterpart. At the same time, it is two times more likely for the same child to become friends with children having similar skills than for children in the top decile of skills distribution. For a white low-skilled child (within the first quintile of skills distribution), the racial gap is lower (around 2.5 times), while the effect of skills is similar.

Technology parameters

Table 3.4 shows estimates for the technology of skill formation. I find a high degree of *static* substitutability between parents and peers, with a complementarity parameter (α_3) of approximately 0.95 (and associated elasticity of substitution of $\frac{1}{1-0.95} = 20$). I find a degree of self-productivity (α_1), i.e. the ability of skills to beget skills, of 0.75. This means that a 1% increase in current skills would predict an average of 0.75%

. This result is qualitatively in line with the previous research on the estimation of technology of skill formation (Cunha and Heckman, 2007, 2008; Cunha *et al.*, 2010; Agostinelli and Wiswall, 2016a). The magnitudes of the share parameters are meaningless due to the different scale between peers' skills (normalized at age 13 to have unit variance) and investments (in yearly hours). However, I find that the estimated value of 0.009 implies that to completely offset a change of one standard deviation in peers' skills, parental investments need to change by approximately four hours per week (for the mean parent). The return to scale of the combined parents–peers inputs (α_4) is 0.77, suggesting non-linear peer effects in skills dynamics even when parents and peers are approximately perfect substitutes (linear). I find that the total factor productivity is an increasing function of the neighborhood quality ($\gamma_{1,tfp}$=0.008) and of the age of children ($\gamma_{2,tfp}$=0.030). The coefficients for school quality represent each neighborhood-type mean and standard deviation of the school fixed effects. I find that average school quality is increased by neighborhood type (from -0.03 for low-income neighborhoods to 0.04 for high-income neighborhoods), while the standard deviation is decreased (from 0.26 for low-income neighborhoods to 0.18 for high-income neighborhoods). Finally, I find production shocks to be important in explaining the total variation in skills dynamics, with a coefficient for the standard deviation of shocks (σ_ξ) of 0.70.

Preferences Parameters

Panel A in Table 3.5 shows estimates for preference parameters. I find both utility for consumption and a child's skills to be relatively concave, with a higher degree of curvature for consumption relative to a child's skills. I find a relatively high degree of parental altruism towards a child's skills: parents care about their child's skills through ages 13 to 15 almost as much as their own consumption, while they care twice as much about the final continuation value for their child relative to their own current consumption. The last result underlines the importance of allowing for a different parameterization for the final period preference, which, as explained above, can capture a different developmental process for children, such as finishing high school, starting a job or going to college.

Wage-Income Process and Initial Conditions

Panel B in Table 3.5 shows estimates for the wage and non-labor income process. A mother's skills are very predictive of both, suggesting an elasticity of 0.44 and 1.03 for wages and non-labor. Table 3.6 shows estimates for the initial conditions by neighborhood type (low/medium/high family income). The normalized mean and variance for a mother's skills and a child's initial skills are for white families living in a low-income neighborhood (neighborhood 1). The other subpopulations' means and variances are relative to the normalized ones. I find that children from minorities start with a lower mean of initial skills relative to their reference white children. White children living in a higher family income neighborhood (neighborhoods 2 or 3) have higher mean initial skills relative to white children in lower-income neighborhoods. I find similar patterns in terms of mother's skills.

3.7.2 Sample Fit

Table B.2 and Table B.3 report the sample fit for the auxiliary regressions of investments and the dynamics of a child's skills. The model is able to replicate the empirical findings on parental investments and skill dynamics. Table B.2 also reports the 95% confidence interval to show that the simulated coefficients are not statistically different from the fitted coefficients. More importantly, Table B.2 shows that the model is able to replicate the switch in sign in the peers' skills, from the cross-sectional (negative) to the permanent (positive) variation in peers, through the different degrees of static versus dynamic complementarity of parental investments and peers' skills in skill formation.

Table B.3 shows that the estimated model is able to fit the auxiliary coefficients for the dynamic aspect of skill formation, suggesting that the estimated technology of skill formation provides the proper marginal productivity for the developmental inputs. The reported 95% confidence intervals suggest that the data and simulated coefficients are not statistically different.

In terms of neighborhood effects on child development, the estimated model is able to fit the differential patterns of skill formation between different neighborhoods and for different races. Figures B.2-B.7 show the sample fit for the mean and standard deviation of skills by age, race and for each of the three types of neighborhood I consider (low/medium/high income).

Figures B.8-B.9 show the sample fit for the homophily bias index for skills and race in different neighborhoods (low/medium/high income). Figure B.8 shows that the model tracks the findings on the skills homophily bias in friendship formation. The model replicates the fact that high-skilled children display a higher bias toward children with similar skills.

Figure B.9 shows that the model is able to replicate the findings of homophily bias by race in different neighborhoods. However, within the high-income neighborhood, while the data indicate a fall in the homophily bias index for Hispanic children relative to black children, the model indicates a common tendency of homophily bias between the two races.

3.8 The Exposure Effects of Environments on Child Development

In this section, I analyze the extent to which environments explain differences in children's final achievements. In order to calculate the treatment effect of better environments, I simulate the counterfactual dynamics of skills as the outcome of being moved to a different environment. I compute the treatment effect by simulating each child in each different environment and school and at any age I consider in the model.

Within the estimated model, I can now calculate the treatment effect of moving from any considered environment to any of the others. Specifically, consider the case where a child i is permanently moved from a environment e to a environment e' at age m. The individual treatment effect in this case is:

$$TE_i(e, e\text{'}, m) = y_i^*(e, e\text{'}, m) - y_i^*(e) \tag{3.28}$$

where $y_i^*(e)$ represents the baseline child percentile in the skill distribution at age 16, while $y_i^*(e, e\text{'}, m)$ represents the child percentile in the skill distribution at age 16 if the child is moved from her original environment e to a new environment e' at age m. To simplify the analysis and to have a comparable setting with the previous literature, I first characterize environments by their mean percentiles of skills of permanently based children at age 16 ($\overline{y}_e$). Second, I consider a specific parametric relationship between a child's outcome and the associated mean percentiles of children's skills at age 16 in each environment ($\overline{y}_e$):

$$y_i^* = \psi_0 + \psi_{1,m}\, \overline{y}_e + \epsilon_i \tag{3.29}$$

where $\overline{y}_e$ represents the associated mean percentile of children's skills in environment e. In this case, $\psi_{1,m}$ represents the average treatment effect of being permanently moved at age m from a environment e to a new environment e' which has a mean of children's skills at age 16 that is one percentile higher.

Finally, I can compute the exposure effects of environments, which represent the effects on skill formation of an additional year in a environment with a mean of children's skills at age 16 that is one percentile higher. Following equation (3.29), the exposure effects are simply $\psi_{1,m}$ - $\psi_{1,m+1}$.

Figure 3.5 shows that environments have sizable effects on skill formation. Moving a child at age 12 into a environment with a mean of children's skills at age 16 that is one percentile higher causes an increase of her skills by approximately 0.65 percentiles. This effect declines by age. I find an exposure effect of 0.048, which means that the outcomes of moved children converge to the outcomes of receiving children at a rate of 4.8% per year of exposure. In other words, moving a child to a better environment at 14 rather than at 13 years old causes the child to lose almost 5% of the benefit of moving. The exposure effects imply approximately a 15% higher benefit from moving to the same environment at 12 rather than at 15 years old. [38]

3.8.1 Model Validation: Comparison with Exposure Effects in Chetty and Hendren (2016a)

The specification in (3.29) allows me to compare my results with those in Chetty and Hendren (2016a). The authors implement the same specification to analyze the childhood exposure effects of environments for the United States. The authors consider individual income percentiles at age 24 as their variables of interest. Still, my results are comparable with those in Chetty and Hendren (2016a) under the reasonable assumption that expected individual income is defined by any monotone rank-

[38] My model starts at age 13 when children draw their skills after the realization of the environment where they live. In this exercise, when I move children at age 12, it means that I allow children to be in the environment before they draw their skills. In contrast, when I move them at age 13, I let them first draw their skills in the original environment and then move them into the new environment.

preserving function of children's skills (higher skills imply average higher individual income). [39]

Figure 3.5 shows that my model tracks the findings in Chetty and Hendren (2016a), who find that an additional year of exposure to a environment with a mean income of one percentile higher for permanently moved children increases a child's income later in life by approximately 0.04 percentiles. The authors also find exposure effects to be stronger for families above the median income distribution. Table 3.7 shows the comparison of model predictions and their estimates with respect to family income. The model's predictions are in line with the heterogeneous exposure effects.

It is still an open question, however, as to which factors drive the exposure effects. In the next section, I decompose the overall exposure effects into three classes of environment-specific amenities: peers, school quality and environment quality.

3.8.2 Decomposition of Childhood Exposure Effects

One advantage of my structural model is that I can now decompose the previous findings of exposure effects with respect to each of the components which characterized an environment in my model: *peers*, *school quality* and environment quality.

With the estimated model, I can replicate the previous simulated experiment of moving children within different scenarios to isolate the effects of each of the previous inputs. I first move children and compute the treatment effect of moving to different environments at different ages due exclusively to the associated change of peers (keeping the previous level of both school and environment quality fixed). Secondly, I compute the effect of moving children when the associated treatment is

[39] I could also estimate a relationship between individual income at age 24 and children's skills at age 16 and then use this estimated function to predict individual income for the simulation exercise. Any monotone relationship between these two variables would give me the same results in terms of rank effects.

composed by both changes in peers and the new school quality (keeping the previous level of environment quality fixed). Finally, I calculate the overall treatment effect of moving at different ages associated with the new peers, the new school and the new environment quality. In this way, I can ascertain the contribution of each component to the overall effects of childhood exposure to environments.

Table 3.8 shows the decomposition of the overall effect. Peers alone account for more than half of the childhood exposure effects, while school and environment quality account for the rest. This means that more than half of the effect of an additional year in a environment with mean skills one percentile higher for permanently moved children is caused by the child's social interactions. Whether a child leaves a disadvantaged environment affects the quality of the child's social interactions (in terms of peers' skills). The dynamic complementarity of skill formation causes this effect to have a higher return earlier than later throughout adolescence. Each additional year of adolescence spent in an environment with mean skills one percentile lower for permanently moved children worsens the child's skills at age 16 by 0.027 percentiles, exclusively through social interactions. The exposure to the same social interactions from age 12 to age 15 would cause a reduction of skills by approximately 0.08 percentiles.

The overall exposure effects are bigger for disadvantaged children, and in this case, peer effects account for almost two-thirds. Children who have a low endowment of skills and are from lower-income families are more likely to live in low-income environments, where they have higher chances of interacting with low-skilled peers due to the homophily bias effects in peer-group formation. This channel explains why this group of children experiences the greatest benefit from leaving these disadvantaged environments. The exposure effect in this case is approximately 0.05, and peers alone account for up to 60% of this finding. For this specific subgroup of children from low-income families, each additional year of adolescence spent in an environment with mean skills of one percentile lower for permanently moved children worsens a child's skills by around 0.034 percentiles at age 16, exclusively through social interactions. Figure 3.6 shows graphically the same result for lower-skilled children.

These results suggest the importance of social interactions alone in explaining the differences in developmental trajectories in children from different environments. In the next section, I analyze the effects of policies that target disadvantaged children and change the peers' composition between different environments, for example, policy that promote socioeconomic integration between environments or schools.

3.9 Policy Analysis

The decomposition of the exposure effects of environments suggests that more than half of the effects come from peers and social interaction. This result calls for policies which focus on changes in peers' composition between different environments to overcome the negative effects of growing up in disadvantaged environments. I analyze how this kind of policy, if implemented on a large scale (i.e. when a sizable fraction of children are moved into a new environment), can generate important equilibrium effects in skill dynamics through the changes in social interactions.

3.9.1 Large Scale Changes in Peers' Composition

The estimated model reveals that skills dynamics depend on a peers' composition and associated children's social interactions. In this section, I analyze the quantitative equilibrium effects, created by policies that change cohort compositions into different social contexts, on the dynamics of skills. Specifically, I want to understand the implications of changes in cohort compositions within both *receiving* and *sending* environments. In order to answer this question, I perform a simulated counterfactual analysis where I move disadvantaged children (and their parents), i.e. children with low skill endowment at age 13 and living in a poor environment, into a high-income environment. This policy has diverse effects on three subgroups of people: (i) parents and children who are moved; (ii) parents and children who live in the receiving (high-income) environment; (iii) parents and children who remain in the low-income environment.

The simulated policy targets a specific group of disadvantaged children: children who are within the first quartile of skills distribution at age 13 in the low-income environment. More than 70% of these children are from a racial minority and their initial (age 13) skills are, on average, about a standard deviation below the population mean of log-skills at age 13. On the other hand, the racial composition of the receiving environment is different: the racial minority makes up only 18% of the population. On average, receiving children are 0.12 standard deviations above the population mean of children's log-skills at age 13. I consider the policy effects when the moved children group is about 5% and 30% of the receiving cohort. These sizable changes in cohort composition allows me to analyze the equilibrium effects of the policy. In the last part of this section, I will discuss more in detail how the effects change as a function of the fraction of moved children.

Table 3.9 shows the effects of the policy that moves the smaller fraction of children (5% of the receiving population). Panel A reports the effects of the policy on children's skills for both the moved and receiving children. For each group, I report the baseline as well as the counterfactual skill dynamics. To assess the importance of equilibrium effects, I also report the counterfactual results without any dynamic equilibrium effects (column called "No Equilibrium"). In this case, I just solve the parent's behavioral problem without considering any equilibrium feedback effects from the endogenous response of other parents' behavior after the policy change. The first finding is that this policy has small effects for the receiving children, with a minimal decrease in skills at age 16 of 3%. On the other hand, I find that moved children increase their skills at age 16, on average, by 55% (0.4 of a standard deviation of children's skills at age 16). [40]

Panel B in Table 3.10 shows the effects of the policy change on parental investment decisions. I find that parents of moved children increase their investments overall, fostering the effects of the policy on their own child's skills dynamics. On the other hand, parents of receiving children do not respond to changes in peer composition.

Table 3.10 displays the effects for the larger policy. Panel A in Table 3.10 reports the effects of the policy on children's skills also for the remained children. In this case, the policy creates winners and losers: an average increase in skills at age 16 of 31% and 23%, respectively, for moved and remaining children is associated with an average decline of 15% in receiving children. An alternative interpretation of the results is in terms of the standard deviation of skills distribution: I find that moved and remaining children increase their log-skills on average by 0.22 and 0.17, respectively, of a standard deviation of children's log-skills at age 16, while, on average, log-skills for receiving children decrease approximately by 0.10 of a standard deviation. The results suggest that children who remained in the sending environment benefit from the outflow of the most disadvantaged companions.

Panel B in Table 3.10 shows the effects of the policy change on parental investment decisions. I find that parents of receiving children reduce their engagement with children due to a lower expected level of peers' skills. On the contrary, parents of moved and remaining children increase their investments due to the policy (positive) change in the expected future peers. Indeed, as suggested by the instrumental variable results, a positive (negative) *permanent* change in peer composition induces a positive

[40]The national population's standard deviation of children's log-skills at age 16 is 1.37

(negative) change in investments due to the dynamic complementarity. Figure 3.7 illustrates how the policy change has affected the equilibrium endogenous distribution of peers' skills and the relative parents' expectations, causing parents with better (worse) expectations to increase (decrease) their investments.

Finally, Table 3.9 and Table 3.10 underline the importance of accounting for equilibrium effects in this type of policy analysis. I find that equilibrium feedback effects tend to amplify the policy effects, and ignoring those would lead to biased policy predictions for children's final skills of approximately seven times smaller. Part of this gap is due to the erroneous predictions for investment decisions: in the absence of dynamic-equilibrium feedback effects, the *static* complementarity between parents and peers dominates the dynamic effects of the policy. In this case, a positive (negative) change in peer composition induces a negative (positive) change in investments.

Heterogenous Effects

In this section, I analyze whether the larger counterfactual policy (when the moved children group is 30% of the receiving cohort) generates heterogeneous effects between children due to the new counterfactual social network. I focus my attention to the two sources of potential homophily bias: skills and race. Figure 3.8 shows the return of policy (treatment effect) in % of skills at age 16 for different children in terms of their initial endowment of skills at age 13. The x-axis displays the percentiles of initial skill endowment for each subgroup. I find that children with lower skills at age 13 benefit the most when moved to a better environment. The treatment effect for children in the first decile of skills within the group of moved children is between 32%-35% in skills at age 16 (approximately 0.25 of a standard deviation for log-skills at age 16). This result is clear evidence of the role of segregated social interactions in child development: in the absence of context-specific network formation, the CES technology in (3.4) would predict that return of policy would increase in a child's endowment. [41] However, a discontinuous policy-induced change in peers creates higher benefits for children who are moved away from adverse social interactions. The policy induces a lower return for children with better initial skills, with the bottom of the effects of approximately 24% for children above the median of the moved group.

The heterogeneous effects in Figure 3.8 for receiving children reveal the importance of the counterfactual endogenous network formation. I find that children in the lower part of the skills distribution have the most sizable adverse effects within the group of receiving children. The treatment effect in this case is -0.14 of a standard deviation of log-skills at age 16. In fact, this group of children is the most exposed to social interactions with the new potential peers. For the same reason as the counterfactual change in social interactions, Figure 3.8 suggests that the policy return for children in the sending (low-income) environment are higher for low-skilled children, with a policy effect of approximately 0.20 of a standard deviation of log-skills at age 16 for children in the first skill decile, in contrast with 0.13 of a standard deviation for children in the highest decile.

Evidence from previous empirical studies indicates a potential racial difference in peer effects in children, pointing out that peer effects seem to be stronger *intra*-race and for minorities (see for example Hoxby, 2000; Angrist and Lang, 2004; Imberman *et al.*, 2012). [42] My quantitative exercise confirms the previous literature's results. Table 3.11 shows the decomposed results for receiving children by race. Children from minority groups are most adversely affected by the policy, with a reduction of

[41] This is because of the assumed complementarity between a child's endowment and other inputs: $\frac{\partial^2 \theta_{t+1}}{\partial \theta_t \partial \overline{H}_{i,t}} \geq 0$.

[42] Imberman *et al.* (2012) exploit the Katrina natural experiment to evaluate peer effects in receiving schools in Louisiana and Houston. The authors find negative peer effects on school attendance, disciplinary infractions and math scores for black children, although the last effect is statistically imprecise

approximately 42% and 35% (-0.30 and -0.25 of a standard deviation) of log-skills at age 16 for black and Hispanic children, respectively. On the other hand, I find that the effects on white children are smaller, with a reduction of log-skills at age 16, on average, of approximately 11% (0.08 of a standard deviation). [43] Panel B in Table 3.11 shows the counterfactual parental investment behavior. Again, groups that are more exposed to the *permanent* change in cohort composition (through homophily bias in social interactions) sizably reduce their parental involvement due to the dynamic complementarity between their current choice and the expected future peers.

The Scale Effects of Policy

During the previous policy analysis, I analyzed the effects of moving two different fractions of disadvantaged children into a high-income environment: 5% and 30% of the receiving population. Equilibrium effects on both receiving and sending environments were quantitatively different. In this section, I analyze the different implications for the same counterfactual policy as a function of different fractions of moved children. Specifically defining the original moved group as the *eligible* children, I now compute the equilibrium effects of the policy relative to the fraction of *eligible* children who are actually moved.

Figure 3.9 reports the average effects on children's log-skills at age 16 in the counterfactual economy as a function of the fraction of *eligible* moved children (x-axis) for the three subgroups of interest (moved, receiving and remaining children). The first result is that both the moved and receiving children are better off if the policy provides a relatively small group of moved children. For moved children, the

[43] The heterogeneous treatment effects by race for the smaller policy are qualitatively similar, see Table B.4

return of the policy rapidly drops as the fraction of eligible children rises in the new environment. Moving 3% or 30% of children creates a difference of approximately 20% in the final skills for moved children. This suggests that the probability of children from the same original environment continuing to interact with each other in the new social context is still high. For the receiving children, the decline is more gradual. The total change between moving nobody versus moving 30% of eligible children is approximately 15%, and it monotonically declines as more eligible children are moved into the environment. The second result is that the remaining children gain increasingly more out of the policy if the fraction of disadvantaged children who are moved out from that environment increases. A gain of 40% is guaranteed for remaining children if 30% of children are moved out from that environment. An increased outflow of the most disadvantaged children from the sending environment benefits children who remain, which is a result of the positive effects of the new peers' skills.

3.9.2 The Persistent Effects of Social Environments on Skills Dynamics

The last quantitative analysis focuses on underlining the persistent effects on skill dynamics of growing up in disadvantaged environments. To answer this question, I perform a simulated counterfactual policy which targets the same disadvantaged group of children as before, but now I boost their initial endowment at age 13 while keeping them living in the low-income environment. I compare their dynamics of skills throughout childhood with children with similar initial skills but living in better environments.

Table 3.12 reports the counterfactual results. I find that living in different social contexts permanently shapes the developmental trajectories of children. In particular, after starting from the same initial endowment, the dynamics of skills in the disadvantaged environment fail to keep up with the skills dynamics of children from the high-income environment. The growth rate of skills is approximately 60% higher for children in the higher-income environment during the age of 13–14 years, while at the end, the two groups' growth rates in skill converge. This leads to a total difference, at age 16, of approximately 57% in final skills.

Social interactions and children's skills composition play an important role in explaining this result. First, children living in the low-income environment have, on average, lower-skilled peers than children living in the high-income environment. Secondly, the two different social contexts determine parents' different expectations about their child's social interactions with peers, thus affecting their investment behavior.

Panel B in Table 3.12 shows that parents increase engagements with their offspring as the result of a higher initial skill endowment of their child. However, they are far from the parental investment levels of parents in high-income environments. The difference is approximately between 4 and 6 percentage points in terms of time allocated to child development. Figure 3.10 shows that part of this investment gap is due to differences in the expected peer effects throughout childhood. Hence, social influences determine patterns of skill inequality and, again, one key mechanism is the dynamic complementarity between parents and expected peers.

3.10 Conclusion

This paper studies the role of children's social interactions in the dynamics of children's skills. I estimate a tractable dynamic equilibrium model of parental investment and endogenous formation of peer groups. The model is estimated using information about friendships, children's test scores and parental investments in the National Longitudinal Study of Adolescent Health (Add Health). I exploit within school / across cohort variations in peers' composition to identify the degree of complementarity between parents and peers in producing a child's skills. I find that parents and peers are *static* substitutes and *dynamic* complementary inputs in child development. After validating my estimated model using findings in Chetty and Hendren (2016a) on environment exposure effects in children, I assess the importance of social interactions in skills dynamics with various policies.

This article underlines three main points: (i) social interactions and social context permanently shape the developmental trajectories of children; (ii) changing cohort composition and the relative social interactions generates *winners* and *losers* and the heterogeneous effects are due to the endogenous formation of new peers; (iii) neglecting the dynamic equilibrium effects of skill formation and social interactions would lead to biased predicted effects of policies.

I want to conclude this paper by considering an extension of this work. Specifically, one potential type of parental investment can be the choice of neighborhood where the family lives. In this case, parents have alternative margins in response to changes in peer composition, and to a certain extent, they can also decide to change where they live as a response to the previously considered policy. Modeling this second channel is challenging, because now the environment composition is also endogenous, and it becomes part of the equilibrium solution of the model. However, understanding the extent to which neighborhood decisions are influenced by children's social interactions is an important question in considering the effects of socioeconomic segregation on intergenerational mobility. Therefore, future work is needed.

Table 3.1: Sample Statistics

	Mean	Standard Deviation
	(1)	(2)
Child's Age	15.65	1.74
Fraction black	0.16	0.37
Fraction hispanic	0.17	0.38
Fraction white	0.67	0.47
N of reported friends (In-School)	4.48	3.58
Measures for skills:		
PPVT	64.26	11.14
English	2.83	0.98
Math	2.72	1.03
History	2.86	1.01
Science	2.82	1.01
Family's characteristics:		
Income ($ 1994)	42,844	27,724
Mother's education	13.13	2.35
Measures for parental investments : (activities in the last 4 weeks with mother)		
Gone shopping	0.72	0.44
Played a sport	0.08	0.28
Gone to a religious service	0.38	0.49
Talked about someone you are dating (or a party you went to)	0.47	0.50
Gone to a movie, play, museum, concert, or sports event	0.25	0.44
Had a talk about a personal problem you were having	0.39	0.49
Had a serious argument about your behavior	0.33	0.47
Talked about your school work or grades	0.63	0.48
Worked on a project for school	0.13	0.34
Talked about other things you are doing in school	0.54	0.50

Notes: This table reports the descriptive statistics for the sample I use in the estimation of the model. The number of reported friends is the number of nominated friends during the survey. The measures for parental investments are binary variables, which take value one if the activity was done, zero otherwise.

Data source: National Longitudinal Survey of Adolescent Health (Add Health).

Table 3.2: Parental Investments, Child's and Peers' Skills

	Dependent Variable Fraction (%) of Invested Parental Time	
	(1) Measurement Error Adjusted	(2) Measurement Error Adjusted and Instrumental Variables (IV)
Child Skills (Log)	2.660	2.120
	(0.316)	(0.668)
Peers' Skills (Log)	-1.441	0.720
	(0.650)	(0.354)
N of Children	14,267	14,267
Age Fixed Effects	✓	✓
School's Fixed Effects	✓	✓
Panel B: First Stage		
$Z_{1,i,t}$ (Minorities Children)		-0.104
		(0.052)
$Z_{2,i,t}$ (White Children)		0.082
		(0.037)
F-Stat Excl. Instruments		11.78
P-value		0.000

Notes: This table shows estimates for both model 3.2 (column 1) and 3.2 (column 2). The dependent variable is the fraction of invested parental time at age t and the covariates (log skills and log peers' skills) are at also at time t. All models also include controls for children's race, mother's skills and lagged family income. Standard errors in parenthesis are computing using a cluster bootstrap. The first stage statistics in column 2 shows the coefficients (and standard errors in parenthesis) of both excluded instruments for the first stage as well as the F-statistic of the joint null hypothesis that both coefficients are zero. Stock and Yogo provide critical values to test weak IV condition based on the F-stat of excluded instruments. Those critical values can be interpreted as a test with a 5 % significance level, of the hypothesis that the maximum relative bias (with respect to the OLS estimates) is 10% or at least 15%. In this case, the Stock and Yogo critical values for the F-stat of the excluded instruments are 19.93 (10%) and 11.59 (15%).
Data source: National Longitudinal Survey of Adolescent Health (Add Health).

Table 3.3: Estimates for the Network Formation Model

Parameter	Estimate	S.E.
Constant (δ_1)	-0.246	0.0172
Child's Log-Skills (δ_2)	0.088	0.0048
Black ($\delta_{3,1}$)	0.075	0.0023
Hispanic ($\delta_{3,2}$)	-0.005	0.0001
Both Black ($\delta_{4,1}$)	0.763	0.0317
Both Hispanic ($\delta_{4,2}$)	0.701	0.0298
Both White ($\delta_{4,3}$)	0.559	0.0475
Distance in Children's Skills (δ_5)	-0.038	0.0014
N of Children (Hundreds, $\delta_{6,1}$)	-0.890	0.0003
N of Children Squared (Hundreds, $\delta_{6,2}$)	0.001	0.0000
Distance in Children's Skills · %White ($\delta_{6,3}$)	- 0.063	0.0032
Distance in Children's Skills · %Black ($\delta_{6,4}$)	0.042	0.0025
Age (δ_7)	-0.050	0.0010
Additional Unobserved Heterogeneity ($\zeta_{i,j,t}$)		
Correlation with Skill Shocks	-0.404	0.0212
Standard Deviation	0.110	0.0095

Notes: This table shows the structural estimates for the child's utility for friendships in Equation (3.5). The child's utility shock is defined as $\nu_{i,j,t} = \widetilde{\nu}_{i,j,t} + \zeta_{i,j,t}$, where $\zeta_{i,j,t}$ and is correlated with the production function shock ($\xi_{i,t}$). The last part of the table shows the estimated correlation and standard deviation σ_ζ. The standard errors are computed using a cluster bootstrap.

Table 3.4: Estimates for the Technology of Skill Formation

Parameter	Estimate	S.E.
Child's Skills (α_1)	0.744	0.0682
Investments (Yearly Hours, α_2)	0.009	0.0014
Elasticity Investment vs Peers (α_3)	0.944	0.0270
Return to Scale (α_4)	0.767	0.0283
Std of Shocks (σ_ξ)	0.700	0.0461
Panel B: TFP		
Constant ($\gamma_{0,tfp}$)	-1.329	0.1256
Neighborhood Quality ($\gamma_{1,tfp}$)	0.008	0.0003
Age Trend ($\gamma_{2,tfp}$)	0.030	0.0008
Panel C: School-Quality Effects		
Low Income Neighborhood		
Mean ($\eta_{s,1}$)	-0.033	0.0350
Standard Deviation ($\sigma_{s,1}$)	0.262	0.0264
Medium Income Neighborhood		
Mean ($\eta_{s,2}$)	0.006	0.0277
Standard Deviation ($\sigma_{s,2}$)	0.244	0.0278
High Income Neighborhood		
Mean ($\eta_{s,3}$)	0.041	0.0318
Standard Deviation ($\sigma_{s,3}$)	0.188	0.0249

Notes: This table shows the estimates for the technology of children skill formation in Equation (3.17). Panel B reports the parameter estimates for the neighborhood-specific TFP defined in Section 3.5.2. Panel C reports the mean and standard deviation of school-quality for each neighborhood type. The standard errors are computed using a cluster bootstrap.

Table 3.5: Estimates for the Preferences and Income Variables

Parameter	Estimate	S.E.
Panel A: Auxiliary Coefficients for Investments		
Curvature on consumption (γ_1)	0.786	0.0046
Weight on Child‘s Skills (γ_2)	0.901	0.0030
Weight on Final Child‘s Skills (γ_4)	2.475	0.2455
Curvature on Child‘s Skills (γ_3)	0.562	0.0256
Curvature on Final Child‘s Skills (γ_5)	0.465	0.0011
Panel B: Parameters of Labor and Non-Labor Income		
Constant (Wage, $\kappa_{1,0}$)	2.750	0.0067
Mother‘s Skills (Wage, $\kappa_{1,1}$)	0.438	0.0048
Constant (Non-Labor Income, $\kappa_{2,0}$)	9.992	0.0174
Mother‘s Skills (Non-Labor Income, $\kappa_{2,1}$)	1.033	0.0113

Notes: Panel A shows the estimates for the utility parameters in Equation (3.9). Panel B reports the estimates for the wage and income process described in Section 3.5.2. The standard errors are computed using a cluster bootstrap.

Table 3.6: Estimates for Initial Conditions

Panel A: Mean Initial Child's and Mother's Skills

	Neighborhood 1		Neighborhood 2		Neighborhood 3	
	Child	Mother	Child	Mother	Child	Mother
Black	-0.47	-0.07	-0.40	0.36	-0.30	0.44
	(0.08)	(0.15)	(0.27)	(0.25)	(0.29)	(0.20)
Hispanic	-0.49	-0.93	-0.48	-0.77	-0.34	-0.36
	(0.11)	(0.19)	(0.26)	(0.19)	(0.25)	(0.19)
White	0.00	0.00	0.02	0.26	0.22	0.58
	(-)	(-)	(0.24)	(0.18)	(0.24)	(0.19)

Panel B: Variance-Covariance Initial Child's and Mother's Skills

	Neighborhood 1		Neighborhood 2		Neighborhood 3	
	Child	Mother	Child	Mother	Child	Mother
	0.65		0.87		0.89	
	(0.05)		(0.08)		(0.15)	
Black	0.20	0.61	0.31	0.67	0.30	0.64
	(0.08)	(0.14)	(0.09)	(0.17)	(0.16)	(0.14)
	0.84		1.10		0.78	
	(0.09)		(0.10)		(0.12)	
Hispanic	0.22	1.59	0.26	1.58	0.28	1.33
	(0.08)	(0.32)	(0.08)	(0.35)	(0.10)	(0.34)
	1.00		1.09		0.99	
	(-)		(0.09)		(0.13)	
White	0.48	1.00	0.37	0.74	0.36	0.78
	(0.07)	(-)	(0.04)	(0.19)	(0.06)	(0.17)

Notes: This table shows the estimates of initial conditions parameters by neighborhood-quality type (low-medium-high income) and race as described in Equation (3.16). The standard errors are computed using a cluster bootstrap.

Table 3.7: Model Validation from Exposure Effects in Children from Chetty and Hendren (2016a)

Panel A: Exposure Effects

	Baseline	
	Chetty-Hendren	Model
Neighborhood Exposure Effect	0.044	0.048
	(0.008)	

Panel B: Exposure Effects by Parental Income

	Below Median Income		Above Median Income	
	Chetty-Hendren	Model	Chetty-Hendren	Model
Neighborhood Exposure Effect	0.031	0.023	0.047	0.047
	(0.003)		(0.003)	

Notes: This table shows the comparison between model's predictions and findings in Chetty and Hendren (2016a) about the childhood exposure effects (baseline and heterogeneous effects by family income).

Table 3.8: Decomposition of Exposure Effects: Peers, School Quality and Neighborhood Quality

Panel A: Baseline Decomposition

	Baseline	Decomposition		
	Overall	Peers	Peers + School Quality	Peers + School Quality + Neighborhood Quality
Exposure Effect	0.048	0.027	0.038	0.048
(Percent)		(+55.58%)	(+23.81%)	(+20.60%)

Panel B: Decomposition for Disadvantaged Children

	Baseline	Decomposition		
	Overall	Peers	Peers + School Quality	Peers + School Quality + Neighborhood Quality
Low Family Income	0.056	0.034	0.046	0.056
(Percent)		(+61.49%)	(+20.13%)	(+18.38%)
Low Skills	0.054	0.030	0.042	0.054
(Percent)		(+56.40%)	(+22.54%)	(+21.05%)

Notes: This table shows the decomposition of childhood exposure effects in Chetty and Hendren (2016a) by: peers, school and neighborhood quality. Panel B shows the decomposition for disadvantaged children (both in terms of family income and child's skills).

Table 3.9: Counterfactual Effects on Skills and Investments (moved children are 5% of the receiving cohort)

Panel A: Effects on Children‘s Log-Skills (Mean)

	Moved Children				Receiving Children		
	Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)		Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)
Age 13	-0.94	-0.94	-0.94	Age 13	0.12	0.12	0.12
Age 14	-0.43	-0.23	-0.38	Age 14	0.87	0.85	0.87
Age 15	0.12	0.48	0.23	Age 15	1.55	1.52	1.55
Age 16	0.45	1.00	0.76	Age 16	2.12	2.09	2.11

Panel B: Effects on Parent‘s Investment Decision (Mean)

	Moved Children				Receiving Children		
	Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)		Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)
Age 13	17.73	22.16	17.22	Age 13	26.75	26.34	26.86
Age 14	17.50	21.32	16.33	Age 14	26.63	26.37	26.97
Age 15	10.83	10.48	9.18	Age 15	22.99	23.07	23.26

Notes: This table shows the counterfactual policy effects for moved and receiving when a fraction of moved children (5% of the receiving population) are moved into a high-income environment. For each subgroup, I compare the baseline results with the policy predictions (equilibrium effects). I also compute the predicted policy effects without equilibrium effects. The latter one is the predicted policy effects if I ignore equilibrium adjustments after the policy implementations. Panel B focus on the investments decisions for moved and receiving children. For each subgroup, I compare the mean predicted parental investments in the baseline case with the equilibrium and no-equilibrium effects.

Table 3.10: Counterfactual Effects on Skills and Investments (moved children are 30% of the receiving cohort)

Panel A: Effects on Children‘s Log-Skills (Mean)

	Moved Children				Receiving Children				Remained Children		
	Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)		Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)		Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)
Age 13	-1.00	-1.00	-1.00	Age 13	0.12	0.12	0.12	Age 13	-0.05	-0.05	-0.05
Age 14	-0.37	-0.28	-0.33	Age 14	0.87	0.80	0.87	Age 14	0.42	0.52	0.43
Age 15	0.24	0.40	0.34	Age 15	1.55	1.43	1.55	Age 15	0.87	1.05	0.88
Age 16	0.53	0.84	0.80	Age 16	2.12	1.97	2.11	Age 16	1.26	1.49	1.29

Panel B: Effects on Parent‘s Investment Decision (Mean)

	Moved Children				Receiving Children				Remained Children		
	Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)		Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)		Baseline	Counterfactual (Equilibrium)	Counterfactual (No Equilibrium)
Age 13	17.89	19.52	17.89	Age 13	26.75	24.74	27.22	Age 13	18.57	21.00	18.04
Age 14	18.00	18.62	17.15	Age 14	26.63	24.84	27.72	Age 14	18.84	21.16	18.16
Age 15	11.77	11.35	10.98	Age 15	22.99	22.82	23.92	Age 15	18.11	18.69	17.53

Notes: This table shows the counterfactual policy effects for moved and receiving when a fraction of moved children (30% of the receiving population) are moved into a high-income environment. For each subgroup, I compare the baseline results with the policy predictions (equilibrium effects). I also compute the predicted policy effects without equilibrium effects. The latter one is the predicted policy effects if I ignore equilibrium adjustments after the policy implementations. Panel B focus on the investments decisions for moved, receiving and remained children. For each subgroup, I compare the mean predicted parental investments in the baseline case with the equilibrium and no-equilibrium effects.

Table 3.11: Counterfactual Effects on Receiving Children by Race (moved children are 30% of the receiving cohort)

Panel A: Effects on Children‘s Log-Skills (Mean)

	Black			Hispanic			White	
	Baseline	Counterfactual (Equilibrium)		Baseline	Counterfactual (Equilibrium)		Baseline	Counterfactual (Equilibrium)
Age 13	-0.30	-0.30	Age 13	-0.34	-0.34	Age 13	0.22	0.22
Age 14	0.28	0.08	Age 14	0.45	0.30	Age 14	0.97	0.93
Age 15	0.96	0.61	Age 15	1.08	0.81	Age 15	1.66	1.58
Age 16	1.49	1.07	Age 16	1.59	1.24	Age 16	2.24	2.13

Panel B: Effects on Parent‘s Investment Decision (Mean)

	Black			Hispanic			White	
	Baseline	Counterfactual (Equilibrium)		Baseline	Counterfactual (Equilibrium)		Baseline	Counterfactual (Equilibrium)
Age 13	20.09	15.30	Age 13	21.97	17.72	Age 13	27.92	26.41
Age 14	18.85	14.46	Age 14	20.67	16.62	Age 14	28.02	26.72
Age 15	16.32	14.97	Age 15	16.69	14.99	Age 15	24.29	24.40

Notes: This table shows the counterfactual policy effects for receiving children (by race) when a fraction of moved children (30% of the receiving population) are moved into a high-income environment. For each subgroup, I compare the baseline results in skills and parental investments (Panel B) with the equilibrium counterfactual predictions.

Table 3.12: Counterfactual Effects on Skills and Investments of Earlier Interventions

Panel A: Effects on Children‘s Log-Skills (Mean)

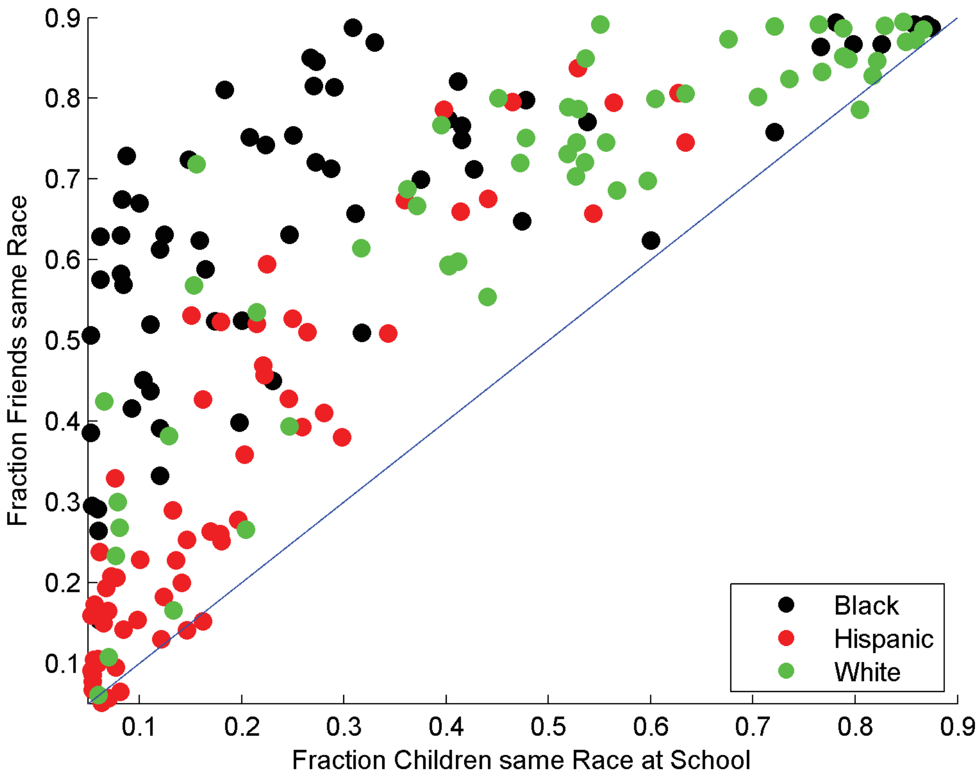
	Baseline	Counterfactual	Similar Children in Richer Environment
Age 13	-1.00	0.46	0.46
Age 14	-0.37	0.85	1.09
Age 15	0.24	1.31	1.74
Age 16	0.53	1.73	2.31

Panel B: Effects on Parent‘s Investment Decision (Mean)

	Baseline	Counterfactual	Similar Children in Richer Environment
Age 13	17.89	20.82	26.47
Age 14	18.01	20.76	26.20
Age 15	11.77	20.24	24.85

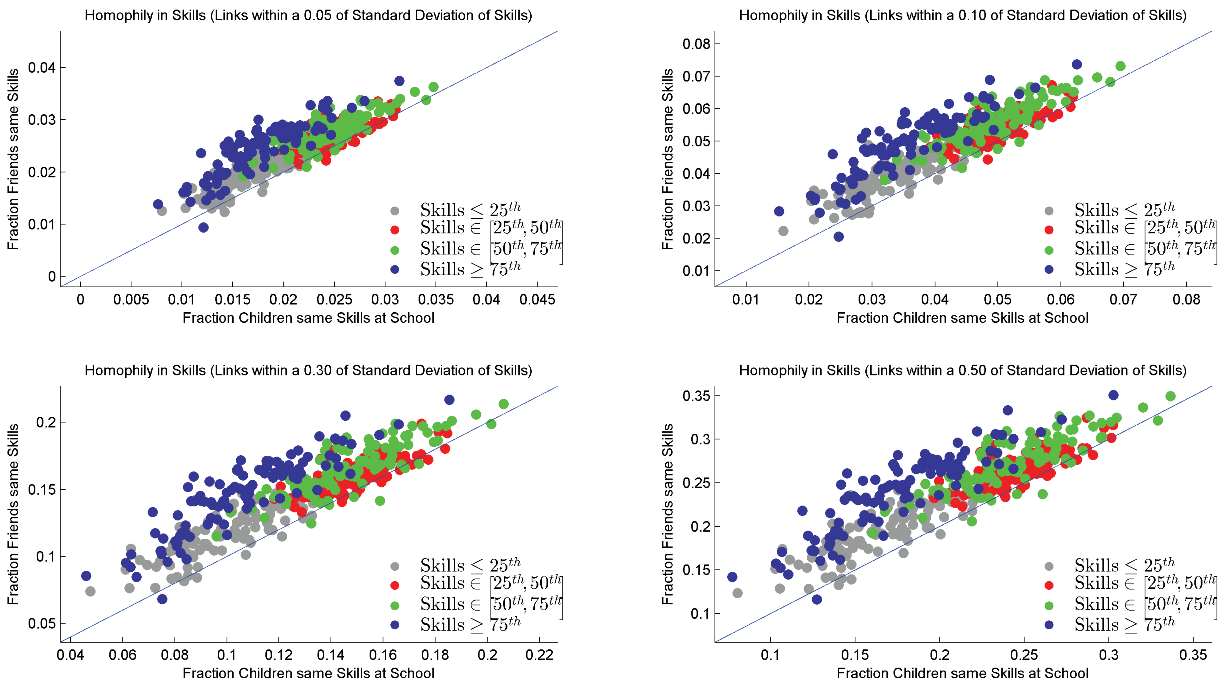
Notes: This table shows the counterfactual policy effects of fostering initial skills (age 13) in skill formation and parental investments (Panel B) for disadvantaged children. Column 1 and 2 report the baseline and counterfactual predictions. Column 3 reports the dynamics of skills and parental investments for children with same mean initial skills at age 13 but who are living in the high-income environment.

Figure 3.1: Homophily in Network Formation by Race



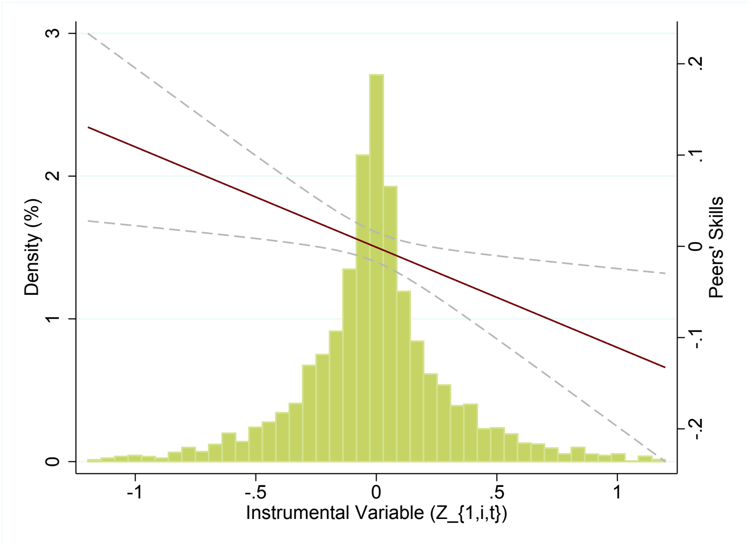
Notes: This figure displays analysis on network formation using the homophily index (see Coleman (1958)). In detail, letting $f_{x,s}$ be the average fraction of friends who are of the same race x at school s and $q_{x,s}$ to be the total fraction of children of race x in a given school s, the homophily-bias index looks as $HBI_{x,s} = \frac{f_{x,s}}{q_{x,s}}$.

Figure 3.2: Homophily in Network Formation by Skills

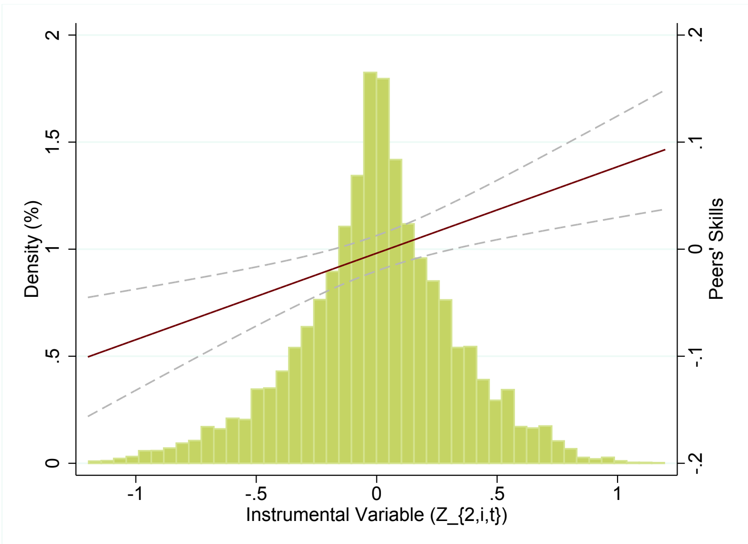


Notes: This figure displays analysis on network formation using the homophily index (see Coleman (1958)). In detail, letting $f_{x,s}$ be the average fraction of friends who are of similar skills level x at school s and $q_{x,s}$ to be the total fraction of similar children with skills level x in a given school s, the homophily-bias index looks as $HBI_{x,s} = \frac{f_{x,s}}{q_{x,s}}$.

Figure 3.3: First Stage Graphs



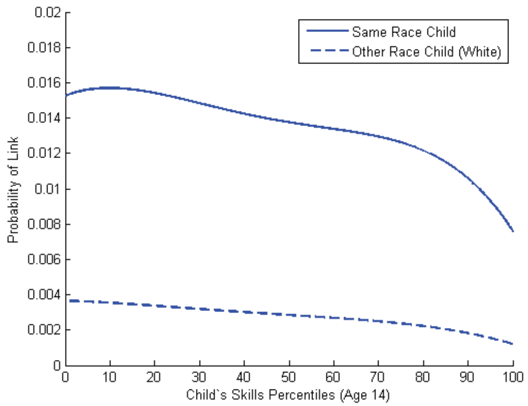
(a) First instrument: Fraction of Minority Children



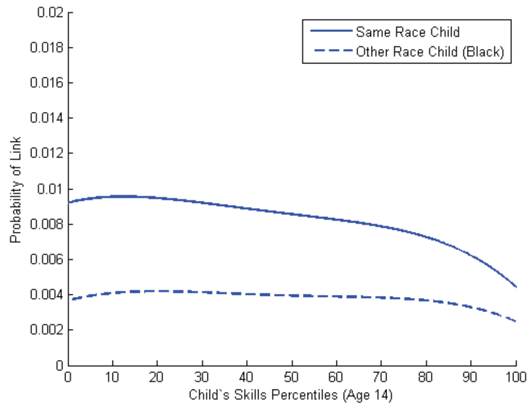
(b) Second instrument: Fraction of White Children

Notes: This figure shows graphically the first stage effects of the instrumental variables. Both graphs show the within-school variation of IV (x-axis) and the peers' skills (y-axes). Solid lines represents a linear regression of peer's skills on each of the two instrumental variables, after controlling for school fixed effects and all the covariates in Table 3.2. Dashed lines show 90% confidence intervals. The plotted values on the background show the overall variation of both instrumental variables (top and bottom 1% excluded).

Figure 3.4: The Probability Two Children Become Friends by Skills and Race



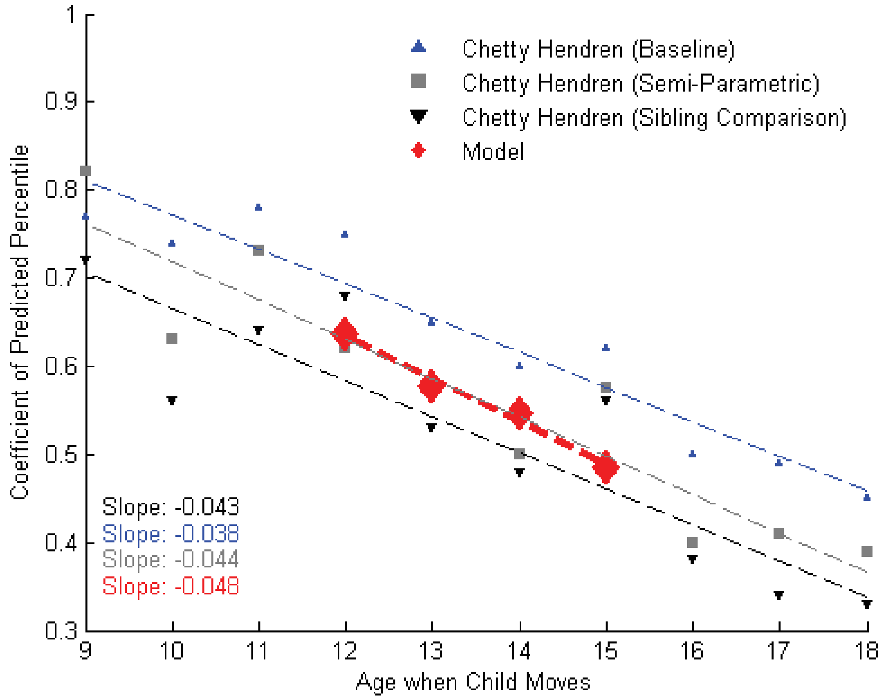
(a) Low Skill (First Quartile) Black Children



(b) Low Skill (First Quartile) White Children

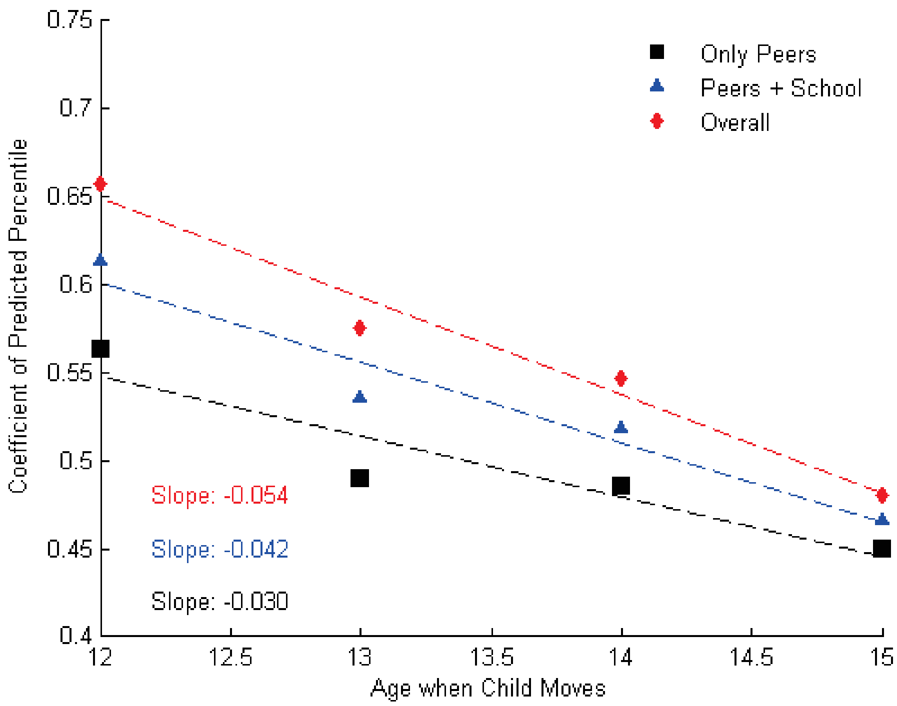
This figure displays the probability that two children become friends by children skills and race. Figure (a) shows the marginal probability for a black child with low skills (in the first quartile of skill distribution at age 14) to become friends with different children over the spectrum of skills and for different races. Figure (b) shows the same graph but for a white child with low skills (in the first quartile of skill distribution at age 14).

Figure 3.5: Model Validation from Exposure Effects in Children from Chetty and Hendren (2016a)



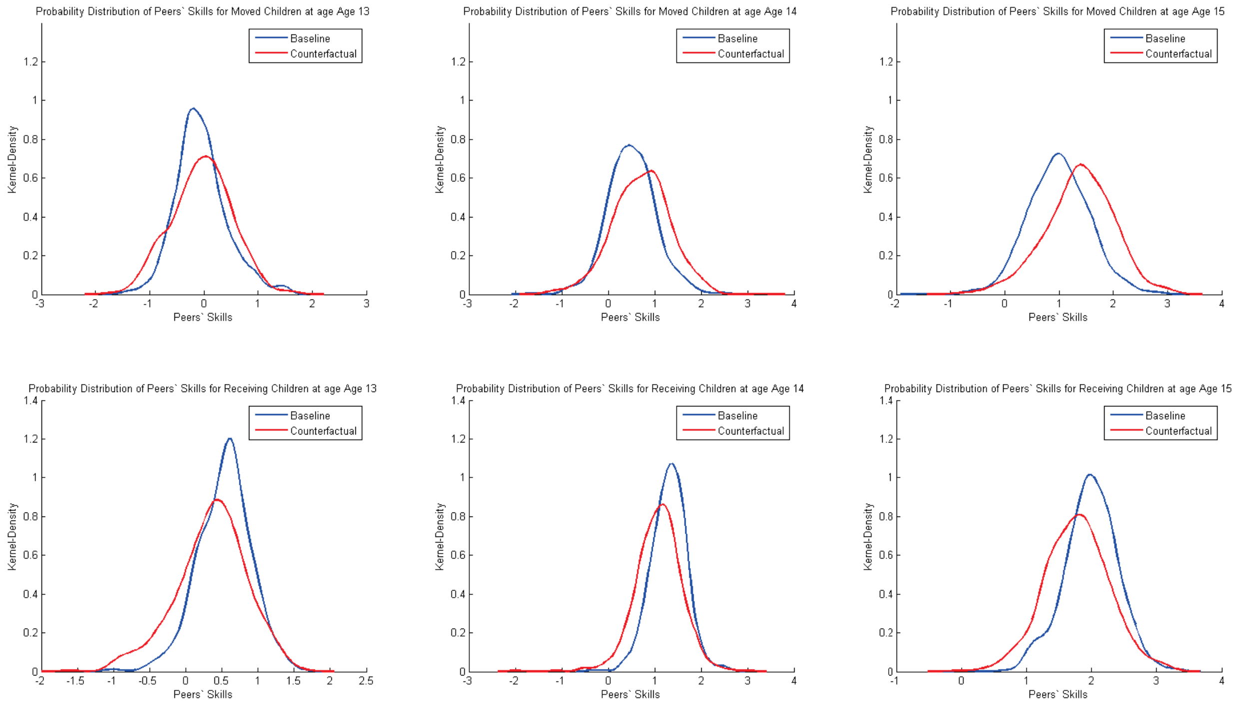
Notes: This figure displays the comparison between model's predictions and findings in Chetty and Hendren (2016a) about childhood exposure effects.

Figure 3.6: Decomposition Exposure Effects for Disadvantaged Children



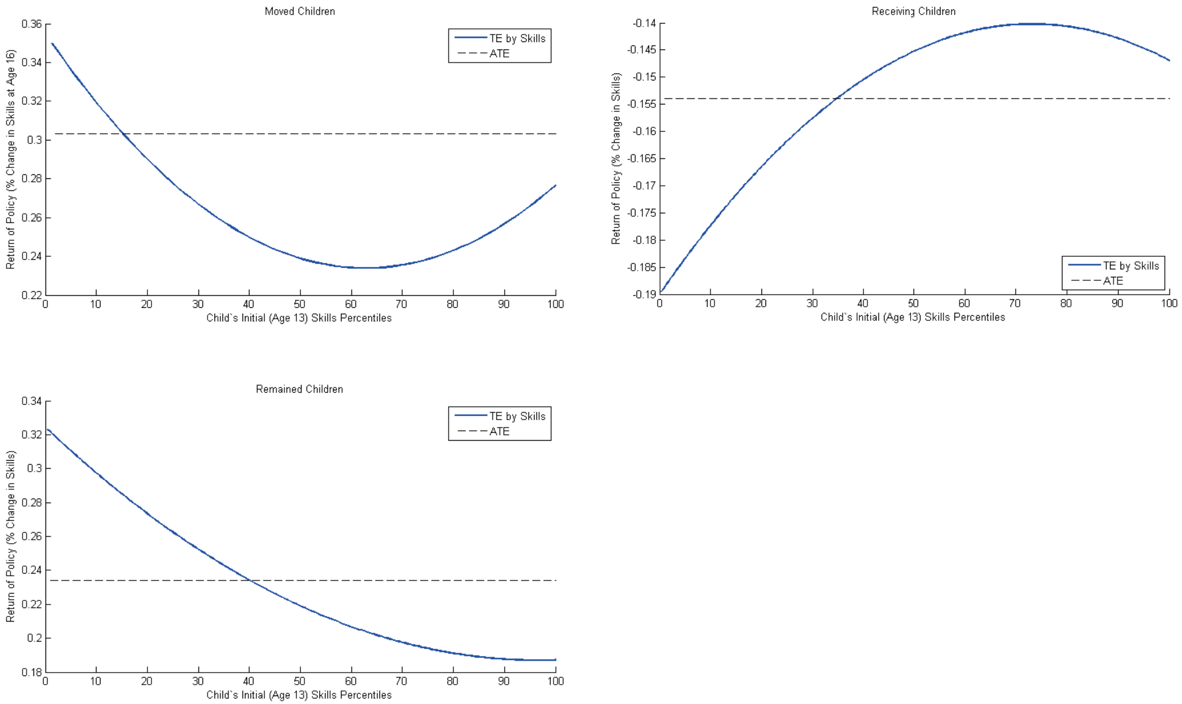
Notes: This figure displays the graphical decomposition for disadvantaged children of childhood exposure effects by: peers, school and neighborhood quality.

Figure 3.7: Peers' Skills Distribution (Baseline vs Counterfactual Economy)



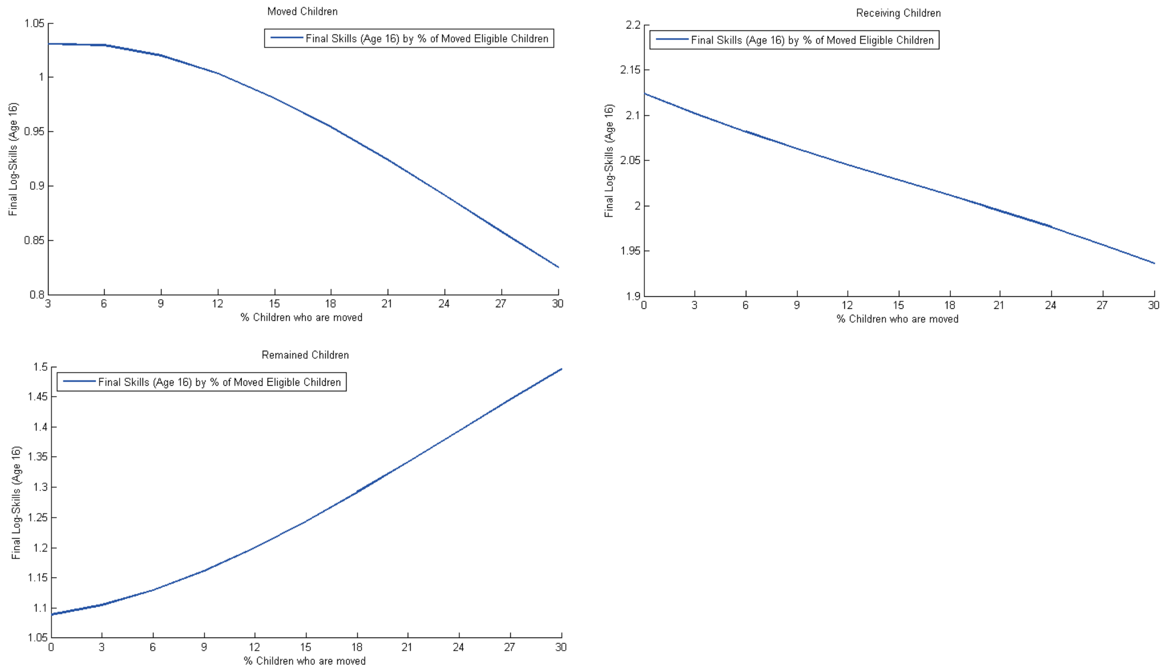
Notes: This figure shows the equilibrium distribution of peers' skills for moved and receiving children by age, before and after the policy is implemented. Parent's expectations about future peers' skills are based on the above distributions. The policy moved a fraction of disadvantaged children into a high-income environment (approximately 30% of the population of the receiving cohort).

Figure 3.8: Treatment Effect by Skills (Moved-Receiving-Remained Children)



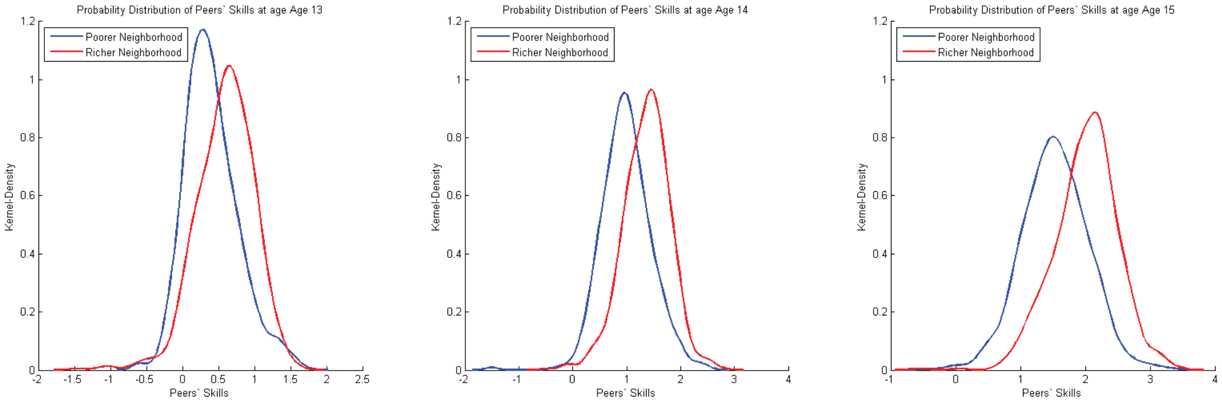
Notes: This figure shows the heterogeneous treatment effects on child's age 16 skills by initial (age 13) child's skills percentile for moved, receiving and remained children. The policy moved a fraction of disadvantaged children into a high-income environment (approximately 30% of the population of the receiving cohort).

Figure 3.9: Treatment Effect by Fraction of Moved Eligible Children (Moved-Receiving-Remained Children)



Notes: This figure shows the effect of the size of moved children for the policy returns for moved, receiving and remained children. Eligible children live in low-income environments and are below first quartile of age 13 skill distribution.

Figure 3.10: Peers' Skills Distribution (Low-Income vs High-Income)



Notes: This figure shows the equilibrium distribution of peers' skills for the children who received the initial boost of skills but still live in the low-income environment and their similar counterpart living in high-income environment. Parent's expectations about future peers' skills are based on the above distributions.

Chapter 4

CHILD DEVELOPMENT, FAMILY INCOME AND MATERNAL LABOR SUPPLY (WITH GIUSEPPE SORRENTI)

4.1 Introduction

Poverty represents one of the major threats to child development. In 2015, about 15 million children in the United States (21 percent of all children) were living in families with incomes below the federal poverty threshold (National Center for Children in Poverty, 2015). What effect does growing up in a disadvantaged family have on a child's achievements, and how can living conditions be improved to promote child development?

Support programs such as the Earned Income Tax Credit (EITC), the Food Stamp Program, and the Child Tax Credit attempt to reduce family poverty and especially that experienced by children. Many of these programs (e.g. the EITC) provide cash transfers on the condition that the recipient works (conditional cash transfers). Such conditions might shape child development by introducing a trade-off between the *income* effect, due to a surge in family income, and the *substitution* effect, due possibly to parental labor supply responses and a decrease in time parents spend with their child.

The arising trade-off poses an important question: is the change in family income more important than time spent with parents in shaping child development? In this study we answer this question by appraising the contemporaneous effect of changes in family income and maternal labor supply on cognitive and behavioral development of children. We implement an instrumental variable (IV) approach exploiting changes in the EITC benefits over time and shocks in the local labor demand as instruments for family income and maternal labor supply. In this sense, we bridge the gap between the literature dealing with the estimate of the effect of family income on child development and the literature on the effect of maternal labor supply and child-with-parents time. Moreover, we provide important insights on what policies can foster maternal employment and child development contemporaneously.

Family income is an important predictor of a child's success and future opportunities. Figure 4.1 shows the wide dispersion in children's achievements by family income. Both cognitive (Panel A) and behavioral (Panel B) development measures exhibit a steep income gradient, with high-achieving children placed in the top deciles of the after-tax family income distribution. The impact of family income on child development has been widely debated by economists. Previous studies such as Duncan *et al.* (2011), Levy and Duncan (1999), and Blau (1999) have found a positive relation between family economic conditions during childhood and child achievements. More recently, works such as Løken *et al.* (2012) and Dahl and Lochner (2012b) employ instrumental variable techniques to confirm this positive effect in Norway and in the U.S., respectively.

In addition to studies regarding the income effect, a vast body of economic literature associates maternal labor supply during childhood with possible negative effects on child development and future opportunities (Baum, 2003; Ruhm, 2004; Bernal, 2008; Carneiro and Rodriguez, 2009; Bernal and Keane, 2011; Hsin and Felfe, 2014; Carneiro *et al.*, 2015; Del Bono *et al.*, 2016; Fort *et al.*, 2017). As examples, according to Bernal and Keane (2011) each year of child care (versus maternal time input) before age 6 decreases test scores by 2.1 percent (0.11 standard deviations). Similarly, Carneiro *et al.* (2015) estimate that the probability of dropping out of high school decreases by 2 percent and wages increase by 5 percent at age 30 with the more time mothers spend with their children in the first months of life.

This paper reconciles these strands of the literature. For most families, an increase in income is due to an increase in maternal labor supply. In this case, a surge in monetary resources is associated with a potential decline in the time the mother spends with her offspring. To understand the possible trade-off between family income and maternal labor supply, we build upon the empirical model in Dahl and Lochner (2012b) by considering not only the role of family income but also the role of maternal hours worked in shaping child development. [1] The work by Dahl and Lochner (2012b) exploits quasi-experimental variation in the EITC to analyze the causal effect of family income on child achievement. However, the EITC is designed to incentivize individuals (including mothers) to work. [2] Mothers, and especially single mothers, are usually the main target group of these welfare programs and are most responsive to incentives (Meyer, 2002; Blundell and Hoynes, 2004; Blundell *et al.*, 2016). This affects the maternal allocation of time between working and parenting, with potential effects on children's test scores. More precisely, endogenous labor supply responses affect child development through two channels. An increase in maternal hours worked generates an income effect, with additional resources coming from a boost in labor income. At the same time, changes in maternal hours worked can also generate a substitution effect, with changes in the time that mothers allocate to child care (Heckman and Mosso, 2014a; Del Boca *et al.*, 2014a). Moreover, this paper is related to previous works that consider the effect of time and monetary resources on children by estimating a structural model of household choices and child development (see Del

[1] Dahl and Lochner (2017), after the analysis by Lundstrom (2017), adjust for a coding error in their previous work in the creation of the after-tax total family income. The results of the original and reviewed studies are similar.

[2] Hotz and Scholz (2003) and Nichols and Rothstein (2016) summarize theoretical and empirical findings about the effect of the EITC on maternal labor supply. Blundell *et al.* (2016) analyze the case of the U.K. and find substantial elasticities for women's labor supply (in particular for the group of single mothers).

Boca *et al.*, 2014a; Mullins, 2016).

An additional contribution of our study relates to the broad definition used for child development. While many works (see Dahl and Lochner, 2012b; Del Boca *et al.*, 2014a) exclusively focus on test scores for cognitive achievements, we extend the analysis to proxies for child noncognitive development. [3] As stated by Heckman and Rubinstein (2001), standard test scores only capture some of the multiple skills determining individual success and well-being. Moreover, early childhood interventions that boost personal traits such as self-discipline or motivation are usually deemed as extremely effective (Heckman, 2000). Socio-emotional skills are often more predictive of later-life success than cognitive skills. [4]

Our empirical analysis is based on the National Longitudinal Study of Youth 1979 (NLSY79) data set matched with its Children (NLSY79-C) section. This combined data set provides longitudinal information about measures of child development, family income, and hours worked by the mother. At the same time, the longitudinal structure allows us to account for individual unobserved heterogeneity through child fixed effects. Cognitive development is measured through children's achievements on the Peabody Individual Achievement Test (PIAT), a set of tests assessing proficiency in mathematics and reading. To study noncognitive development, we take advantage of the Behavior Problems Index (BPI). This comprehensive index is comprised of several different indicators for behavior such as aggressiveness or hyperactivity that are likely to shape children's future life opportunities.

Given the strong interdependence between maternal labor supply and family in-

[3] We also explore features related to early childhood development (1–7 years old).

[4] For example, data from the Perry Preschool Program, a high-quality U.S. preschool education program, suggest that increased academic motivation generates 30 percent of the effects on achievement and 40 percent on employment for females. Reduced externalizing behavior decreases lifetime violent crime by 65 percent, lifetime arrests by 40 percent, and unemployment by 20 percent. Visit heckmanequation.org/resource/early-childhood-education-quality-and-access-pay-off/ for a discussion of these results.

come, there is no suitable identifying source of variation that is likely to exclusively affect one variable of interest. Hence, in order to identify the single causal effect of either family income or maternal labor supply on child development, it is necessary to allow for the endogeneity of both inputs. To deal with this challenge, we exploit two instrumental variables. The first instrument is based on the longitudinal changes in monetary benefits of the EITC, one of the largest U.S. federal income support programs. This variation provides us with exogenous changes in family resources to allocate in child development. At the same time, only working people are eligible for EITC benefits, creating incentives for mothers to work. The second instrument is constructed by using longitudinal shocks in the local labor market demand. Shifts in local demand for labor affect equilibrium prices (wages) and, subsequently, the family income resources and the equilibrium labor quantity. [5]

Our instrumental variable analysis suggests different results for cognitive and behavioral development. An additional $1,000 in family income improves cognitive development by 4.4 percent of a standard deviation. [6] The same income change has no effect on child behavioral development. An additional $1,000 improves behavioral development by 1.3 percent of a standard deviation, and the result is not statistically significant.

We find that the income effect is counterbalanced by a negative effect of hours worked by the mother on child development. An increase in maternal labor supply of 100 hours per year causes a statistically significant decrease in both child cognitive

[5]We provide evidence throughout the paper that both identifying sources of variation do not confound other contemporaneous state-specific factors, like state-specific trends in children's achievements or changes in the per-pupil financial resources of schools in different states. Moreover, in the spirit of Goldsmith-Pinkham *et al.* (2017), we assess the validity of our labor demand shock instrument by formally testing for any parallel pre-trends between the instrument and child development. We reject the hypothesis of the existence of any pre-trends.

[6]This result is in line with the findings of Dahl and Lochner (2012b) and Dahl and Lochner (2017).

and behavioral development by approximately -6 percent and -5 percent of a standard deviation. The strong negative impact of the number of hours worked by the mother, both in terms of cognitive test scores and behavioral problems, encourages the debate in a new dimension: how to address concerns about the effect of maternal employment on child development.

We attempt to answer this question in the last part of our study. By using the time diary component of the American Time Use Survey (ATUS), we illustrate the mechanism underlying the negative impact of hours worked by the mother on child development. Similar to Sayer *et al.* (2004), Guryan *et al.* (2008), and Fox *et al.* (2013), we find that working mothers, conditional on income, invest less time in their children. As a consequence, labor market conditions play a role in shaping the effect of labor supply on child development.

We focus on the role of wages and show that, according to our results, an after-tax hourly wage up to $13.50 makes the substitution effect (less maternal time with the child) dominant over the income effect (higher earnings). With higher earnings, families face the option of substituting their decreased time investment with better and more productive alternatives (e.g. nonparental care, additional schooling, youth clubs, music activities, etc.).

We look for possible heterogeneous effects in different subgroups in order to highlight the potential importance of alternative inputs in the child development process. Behavioral development does not display evidence of heterogeneous impacts of income or hours worked by the mother. On the contrary, the negative effect of hours worked by the mother on cognitive development only appears in less educated, low-skilled, or single mothers. More educated and high-skilled mothers are likely to access to better nonparental child care options. Moreover, the differences in the labor supply effect can be reconciled with heterogeneous preferences for child care activities, generating different patterns of time allocation between working, child care, and leisure time (Guryan *et al.*, 2008).

We further investigate these channels by comparing the investment in the child by maternal employment status and family income. The Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID) collects detailed information about a wide set of children's activities and parental investment for a representative sample of U.S. families. Results obtained with this data set highlight some evidence of differential investments as a response to the maternal employment status when low-income families are compared to high-income families.

Policymakers might obtain several suggestions from our results. First, by showing the trade-off between the income and substitution effect in terms of child development, this work speaks to the growing body of literature about the effect of conditional versus unconditional cash transfers. Many income subsidies worldwide base monetary transfers on work requirements. In this context, only looking at the effect of income on child development might lead to biased policy predictions. Our results support the idea that policies aimed at fostering maternal labor supply can be beneficial to child development if integrated with specific consideration about a minimum wage or the taxation of family income. Alternatively, policies that encourage maternal employment in low-income families should also consider how to guarantee alternative sources of child care to support child development.

The remainder of the paper is structured as follows. Section 4.2 introduces the empirical model and the identification strategy. The data used for the analysis are presented in Section 4.3, while the results are described in Section 4.4. Section 4.5 sheds light on the mechanism underlying the main findings of the work. Section 4.6 concludes.

4.2 Methodology

4.2.1 Empirical Model

Child inputs strongly affect individual development and future opportunities. Our empirical model aims to capture the impact of family income and maternal hours worked on child development. We build upon the empirical model considered in Dahl and Lochner (2012b) by including the hours worked by the mother as an additional explanatory variable for child achievement. Specifically, our child outcome equation takes the following form:

$$y_{i,t} = \beta_0 + \alpha_0\, t + \alpha_1\, I_{i,t} + \alpha_2\, L_{i,t} + x_i'\, \beta_{1,t} + x_{i,t}'\, \beta_2 + \eta_i + \epsilon_{i,t}\ , \tag{4.1}$$

where $y_{i,t}$ represents the child's outcome in period t. [7] In our empirical analysis, we focus specifically on both child cognitive and behavioral development. $I_{i,t}$ and $L_{i,t}$ reflect the after-tax total family income and the maternal labor supply (hours worked) at time t. x_i and $x_{i,t}$ represent exogenous observed family i fix and time-varying characteristics. η_i reflects unobserved family specific heterogeneity (which can capture any permanent unobserved family factor as well as child unobserved ability). We allow for an age-trend effect in children's outcomes (α_0). Finally, we define $\epsilon_{i,t}$ as the additional time-varying unobserved heterogeneity in the child's outcome, which may include unobserved child developmental shocks. Taking first differences to eliminate family fixed effects leads to the following empirical specification:

$$\Delta y_{i,t} = \alpha_0 + \alpha_1 \Delta I_{i,t} + \alpha_2 \Delta L_{i,t} + x_i'\beta_1 + \Delta x_{i,t}'\beta_2 + \Delta\epsilon_{i,t}\ , \tag{4.2}$$

where $\beta_1 = \beta_{1,t+1} - \beta_{1,t}$ allows us to control for differential growth in children's outcomes by observable characteristics (e.g. gender, age, race, etc.). [8] Equation (4.2)

[7] We consider periods to be the child's age, and we use these two concepts interchangeably.

[8] The alternative, more general approach is to allow for a semiparametric model of differential age effects of observable characteristics on outcome growth by age.

constitutes the baseline empirical model of this study, while α_1 and α_2 are the parameters identifying the income and maternal labor supply effect on child achievement. The coefficient α_1 expresses the effect of changes in family income on changes in child achievement, while α_2 captures the mother's labor supply effect on changes in child achievement.

To recover the parameters in equation (4.2) and to deal with the endogeneity of both family income and maternal hours worked, we implement an IV estimation strategy. Children from disadvantaged backgrounds are likely to experience contextual conditions affecting their development in the presence of substantial positive income shocks. Similarly, changes in maternal labor supply are likely to be linked to other unobservable characteristics affecting child development. Our IV approach tackles the endogeneity issues by exploiting two sources of exogenous longitudinal variation: (i) changes in the EITC benefits, one of the largest federal income support programs; and (ii) shocks in the local labor market demand.

4.2.2 Instrumental Variables

The identification of equation (4.2) is particularly challenging due to the endogeneity of both family income and maternal labor supply. Changes in family resources and *intra*-family labor market decisions can be correlated with family-specific unobserved permanent shocks, which threatens the validity of a standard OLS approach. We deal with this issue by implementing an IV approach based on two instruments: longitudinal changes in the EITC schedule, and longitudinal variation in labor demand shocks measured as geographical changes in sectoral compositions of local economies. The identification of the parameters in our linear specification in equation (4.2) requires two necessary conditions for the instruments: relevance and exogeneity. Here, we describe in detail the two instrumental variables. The discussion about the relevance of the instruments is postponed to Section 4.4.1, in which the results are presented.

Longitudinal Changes in EITC Benefits

When the EITC was introduced in 1975, it was a modest program that aimed to improve economic and social conditions of low-income families with dependent children. After its introduction, the EITC was progressively expanded (e.g. in 1986, 1990, 1993, etc.) to become the largest cash transfer program for low-income families with dependent children (Eissa and Liebman, 1996). In 2013, the total federal EITC reached $63 billion shared by 27 million individuals. In 2015, the program was responsible for lifting about 6.5 million people out of poverty, including 3.3 million children (Center on Budget and Policy Priorities, 2016).

The credit is conditioned on three eligibility criteria: (i) the taxpayer needs to report a positive earned income; (ii) the adjusted gross income and earned income must be below a certain year-specific threshold; and (iii) the taxpayer needs to have at least one qualifying child. [9] Therefore, the EITC's primary incentive is to increase the labor supply (Nichols and Rothstein, 2016). The provision of work incentives is typical of many welfare programs, and as shown in Blundell *et al.* (2016) in the U.K., mothers, and especially single mothers, are usually the most responsive target to these incentives.

As shown in Figure 4.2, the EITC income thresholds and benefits have changed over time. We plot the different amounts of received transfers conditional on after-tax family income, keeping all the family characteristics (e.g. marital status, number of dependent children, etc.) fixed. Focusing on a single year, it is possible to observe the structure of the EITC program and, specifically, the three phases that characterize the program. In the phase-in, the credit is a pure earnings subsidy. This is followed by

[9] A few exceptions to the last criterion were introduced in 1994.

a flat phase after which the credit starts to gradually phase-out. Individual incentives and behaviors regarding labor supply may differ according to the family structure and the position on the schedule. In particular, mothers who fall into the phase-out part of the schedule may have incentive to reduce their hours worked. However, Meyer (2002) provides evidence, at least for single mothers, that the past expansions in the EITC schedules did not show this type of response.

Figure 4.2 shows the EITC federal schedule expansions over time. Families with an after-tax income of around \$15,000 received a transfer of around \$1,000 in 1987 or 1989. The same families received an amount that was 400 percent higher (around \$4,000) in 1999. We exploit this variation of the EITC schedules over time to predict changes in family income and changes in maternal labor supply.

We start by showing the premise underlying the EITC's effects on our variables of interest. EITC benefits affect family income in two ways: (i) *directly* through the tax credit transfer; and (ii) *indirectly* through labor supply responses. Consider the following after-tax total family income ($I_{i,t}$) decomposition:

$$I_{i,t} = \underbrace{w_{i,t} \cdot L_{i,t}(EITC_{i,t}) + \widetilde{I}_{i,t}}_{I_{i,t}^{pre-tax}} + EITC_{i,t}(I_{i,t}^{pre-tax}) - \tau_{i,t}(I_{i,t}^{pre-tax}) \; , \qquad (4.3)$$

where $I_{i,t}^{pre-tax}$ represents the pre-tax family income, composed of the mother's pre-tax earnings ($w_{i,t} \cdot L_{i,t}(\cdot)$) and other sources of income ($\widetilde{I}_{i,t}$). $EITC_{i,t}(\cdot)$ and $\tau_{i,t}(\cdot)$ represent respectively the EITC schedule and income tax schedule as a function of pre-tax family income.

The IV strategy is based on changes in the EITC schedules over time. However, directly using changes in received EITC benefits would make the instrument invalid as a change in the transfer that families receive is a function of both policy changes in the EITC schedules and the endogenous response in family income. Indeed, family income endogenously changes in response to several factors such as individual labor supply choices, changes in marital status or household structure, etc.

To exploit only policy changes in the EITC schedules, we construct the instrumental variable as in Dahl and Lochner (2012b). We calculate the change in EITC benefits due to changes in the EITC schedules over time based on the predicted family income change that would have happened in any case, keeping fixed the family structure and characteristics to avoid possible endogenous changes in family composition and characteristics. In this way, our instrumental variable captures only the longitudinal variation in monetary benefits due to the changes in EITC schedules.

Specifically, our instrument takes the form:

$$\Delta EITC_{i,t}^{IV}\,(I_{i,t-1}^{pre-tax}) = EITC_{i,t}(\widehat{E}\left[I_{i,t}^{pre-tax}|I_{i,t-1}^{pre-tax}\right]) - EITC_{i,t-1}(I_{i,t-1}^{pre-tax})\ , \quad (4.4)$$

where $\widehat{E}\left[I_{i,t}^{pre-tax}|I_{i,t-1}^{pre-tax}\right]$ represents the predicted family income as a function of lagged pre-tax income. We follow Dahl and Lochner (2012b), and we use a fifth order polynomial of past income together with an indicator for positive lagged pre-tax income to predict current pre-tax income. For each family, the predicted changes over time in the benefits in equation (4.4) are now only a function of changes in schedules.

However, there is a possible concern underlying the definition of the instrumental variable in equation (4.4). In a cross-sectional perspective, differences in imputed changes in EITC benefits are explained by the previous period's pre-tax family income ($I_{i,t}^{pre-tax}$), as well as the predicted family income change ($\widehat{E}\left[\Delta I_{i,t}^{pre-tax}|I_{i,t-1}^{pre-tax}\right]$). We take into account this concern by introducing a control function for family income ($\Phi(I_{i,t-1}^{pre-tax})$) and augmenting our model specification as follows:

$$\Delta y_{i,t} = \alpha_0 + \alpha_1 \Delta I_{i,t} + \alpha_2 \Delta L_{i,t} + x_i' \beta_1 + \Delta x_{i,t}' \beta_2 + \Phi(I_{i,t-1}^{pre-tax}) + \Delta\epsilon_{i,t}\ . \quad (4.5)$$

With the inclusion of the income control function in the model, the validity of our first instrument relies on the assumption that no unobserved heterogeneity potentially correlated with lagged pre-tax family income is left. This condition translates into the following mean independence condition:

$$E(\Delta\epsilon_{i,t}|\Delta EITC_{i,t}\,(I_{i,t-1}^{pre-tax}))=0\;, \tag{4.6}$$

where $\Delta\epsilon_{i,t}$ represents the error term in equation (4.5). In other words, condition (4.6) assumes that our control function captures the true relationship between the expected unobserved heterogeneity and lagged pre-tax income. To fulfill this requirement, we introduce a generalization of the control function in Dahl and Lochner (2012b) and we exploit a flexible Taylor expansion of $\Phi(\cdot)$ about the point of predicted income for a fixed EITC schedule change:

$$\begin{aligned}\Phi(I_{i,t-1}^{pre-tax}) \approx\; & \Phi\left(\widehat{E}\left[I_{i,t}^{pre-tax}|I_{i,t-1}^{pre-tax}\right]\right)\\ & +\sum_{n=1}^{k}\frac{\Phi^{(n)}\left(\widehat{E}\left[I_{i,t}^{pre-tax}|I_{i,t-1}^{pre-tax}\right]\right)}{n!}\cdot\left(\widehat{E}\left[I_{i,t}^{pre-tax}|I_{i,t-1}^{pre-tax}\right]-I_{i,t-1}^{pre-tax}\right)^{n}\;.\end{aligned} \tag{4.7}$$

The control function in equation (4.7) reconciles with the one implemented in Dahl and Lochner (2012b) in the limited cases in which they assume the control function to have the same functional form used to estimate the predicted family income ($n=0$ order of approximation and $\Phi(I_{i,t-1}^{pre-tax})=\widehat{E}\left[I_{i,t}^{pre-tax}|I_{i,t-1}^{pre-tax}\right]$).

Finally, we discuss a further possible threat related to the use of the EITC instrument. EITC changes may induce different responses in maternal labor supply for particular subpopulations. This can potentially compromise the monotonicity assumption of the instrument, which allows us to interpret the IV results as the local average treatment effect (see Imbens and Angrist, 1994). Monotonicity is an untestable assumption, but we focus on specific subgroups of our sample that have potentially different labor supply responses to EITC changes. Specifically, we separately focus on heterogeneous responses to the EITC with respect to lagged maternal employment status. [10] No evidence of possible nonmonotone responses to EITC changes arise in our framework, and all the results remain unchanged. At least in the above-considered dimension, our empirical strategy seems robust to potential heterogeneous responses in EITC changes.

Labor Demand Shocks

We use as a second instrument the spatial differential effects of long-term aggregate trends on local labor markets. Different local labor markets are characterized by different economic sectoral compositions, inducing different expositions to aggregate structural changes in the economy. Ideally, we would identify differences in exogenous labor demand changes, unrelated to the supply side, that shift the equilibrium of local labor market outcomes. We then could use this variation to predict changes in family income and mother's labor supply. Following the approach first developed by Bartik (1991) and used in many previous empirical works (see for example Blanchard and Katz, 1992; Autor and Duggan, 2003; Luttmer, 2005; Aizer, 2010; Notowidigdo, 2011; Bertrand *et al.*, 2015; Diamond, 2016; Charles *et al.*, 2015, 2017), we construct an empirical analogue of the above-mentioned thought experiment by considering the cross-state differences in industrial composition and aggregate growth in the employment level.

We exploit heterogeneous labor demand shocks for women by state and educational attainment. We define a group (or cell) "*se*" as the aggregation index for people living in a state s with a level of education e. For each variation unit se, we create labor demand shocks as national changes in industry-specific employment rates weighted by the industry female employment share at the baseline year. For our

[10] Dahl and Lochner (2017) use a similar approach and allow for different effects for EITC changes relative to the mother's employment status.

empirical analysis, we fix the baseline year at 1980, as our empirical analysis focuses on the period from 1988 to 2000 (see Section 4.3 for more details). [11]

Any observation i that belongs to the specific cell se is matched with the following instrumental variable value:

$$LabDemShocks^{IV}_{i,t} = \sum_{ind} (\ln E_{ind,-s,t} - \ln E_{ind,-s,1980}) \frac{E_{ind,se,1980}}{E_{se,1980}} , \tag{4.8}$$

where $(\ln E_{ind,-s,t} - \ln E_{ind,-s,1980})$ is (approximately) the percentage change in the aggregate employment rate in industry ind relative to 1980. To calculate this statistic for each state s, we consider all states except state s to avoid possible concerns of endogeneity (Goldsmith-Pinkham *et al.*, 2017). $\frac{E_{ind,se,1980}}{E_{se,1980}}$ represents the 1980 female employment share of industry ind for a specific education group e in state s. The instrumental variable constructed in equation (4.8) can be interpreted as the average long-term growth in employment rates by state and educational achievement.

Figure 4.3 graphically shows the variation of labor demand we exploit. For the sake of clarity, we report only the first (1988) and the last year (2000) covered by our sample and two levels of educational attainment (high school dropout and college graduate). However, in the empirical analysis, we construct the instrumental variable for all years of our analysis and for four types of educational levels: high school dropout, completed high school, some college, and completed college.

Figure 4.3 displays extensive changes in the employment rate over time and between different states. First, low- and high-educated mothers display opposite dynamics in employment rates. High school dropouts experience an overall decline in employment rate, with an average change of -0.34 percent from 1988 to 2000. On the contrary, the employment rate for college graduates increased by 0.40 percent from 1988 to 2000. Second, changes in employment rates from 1980 to 2000 are hetero-

[11]Moreover, we choose the 1980 as the baseline year instead of an earlier decade as the earlier versions of census data sets are only 1 percent samples instead of 5 percent samples.

geneous among states, with a standard deviation of 0.55 percent for low educated and 0.15 percent for highly educated women. The greatest declines in high school dropouts between 1988 to 2000 are shown in North Carolina, South Carolina, and Rhode Island, with a decline of -1.96, -1.80, and -1.68 percent, respectively. The greatest increases in employment rates for college graduate women are displayed in the District of Columbia, New York, and Massachusetts, with an increase of 1.41, 0.95, and 0.93 percent, respectively.

Conditional Independence. A recent paper by Goldsmith-Pinkham *et al.* (2017) shows that exploiting the labor demand shocks in equation (4.8) "is equivalent to using local industry shares as instruments, with variation in the common industry component of growth only contributing to instrument relevance." Hence, we can define our identifying assumption as the mean independence of the change in developmental, unobserved shocks ($\Delta\epsilon_{i,t}$) from 1988–2000 and the employment shares during 1980 for each state and education level:

$$E(\Delta\epsilon_{i,t}|LabDemShocks_{i,t}^{IV}) = 0\ . \tag{4.9}$$

The condition in equation (4.9) does *not* state that cross-sectional differences in children's unobserved skills from 1988–2000 are uncorrelated with the state-specific employment shares in 1980. This last statement would be difficult to defend because of unobserved specific differences between states, which would directly affect the *level* of skills (e.g. school-quality differences) and would be potentially correlated with the industrial composition of that state. Instead, our conditional independence condition points toward the dynamic aspect of child development, assuming that the unobserved *changes* in children's skills during 1988–2000 are uncorrelated with the state-specific industrial compositions in the U.S. in 1980.

To deal with some potential concerns underlying the condition in equation (4.9), we introduce an augmented specification of the model in equation (4.1) with potential state-specific trends in children's skills formation. In this way, we control for potential unobserved changes in state-specific factors that can affect the change in children's skills and, at the same time, can be confounding with the variation in local labor demand shocks (i.e. state-specific trends in school quality). All the results remain unaffected by the inclusion of state trends. [12]

Finally, following the suggestion in Goldsmith-Pinkham *et al.* (2017), we assess whether any parallel pre-trends between our instrumental variable and child development could jeopardize the validity of our identification strategy. Specifically, Goldsmith-Pinkham *et al.* (2017) recommend testing whether future values of the instrumental variable are predictive for the current second stage residuals. We do not find evidence of pre-trends.

Exclusion Restriction. The conditional independence is sufficient to interpret as causal the reduced form effect of labor demand shocks on child achievement. However, we need the *exclusion restriction* to hold in order to interpret our IV estimates as the causal effect of family income and labor supply. The exclusion restriction requires labor demand shocks to affect children's outcomes through either changes in after-tax family income or changes in maternal labor supply, and not directly in any other way.

One concern potentially undermining the exclusion restriction relates to the fact that local labor demand shocks might affect employment and the allocated resources in the education industry. We address this concern in Section 4.4.1 by showing that baseline results do not change if we augment the model with the change in per-pupil total revenues and per-pupil total current expenditures by state and over time. This evidence suggests that our instrument does not affect children's achievement through

[12] See Section 4.4.1 for the analysis.

changes in the education system.

4.2.3 The Two-Stage Least Squares Estimator

We aim to estimate the causal impact of family income and maternal labor supply on measures for child development (y). We analyze child development by focusing on proxies for both cognitive and noncognitive development. Specifically, we exploit (i) individual scores in a combined math-reading standardized test as a proxy for children's cognitive development; and (ii) a standardized index for children's behavioral problems. [13] As discussed, we use longitudinal changes in the EITC schedule and longitudinal variation in labor demand shocks, measured as geographical changes in sectoral compositions of local economies, as instruments for family income and hours worked by the mother.

In this framework, for each of the endogenous variables $\Delta W \in \{\Delta I, \Delta L\}$ (changes in income or changes in hours worked by the mother), we estimate the following first stage:

$$\Delta W_{i,t} = \gamma_0 + \gamma_1 \Delta EITC_{i,t}^{IV} + \gamma_2 LabDemShocks_{i,t}^{IV} + x_i' \gamma_3 + \Delta x_{i,t}' \gamma_4 + \Phi(I_{i,t-1}^{pre-tax}) + \Delta u_{i,t} \ , \tag{4.10}$$

where i represents the child and t the time period. $\Delta EITC_{i,t}^{IV}$ is the change, with respect to the previous period, in the EITC schedule experienced by children i. $LabDemShocks_{i,t}^{IV}$ stays for labor demand shocks at time t (with respect to the baseline year 1980) experienced by children i in state s and with maternal education background e. To allow for differential growth rates in test scores in children with different (observable) characteristics, the vector X_{it} contains variables for children's gender, age, race, and number of siblings. The same vector also contains the third order polynomial control function for income previously discussed in Section 4.2.2.

[13] We carefully introduce all details about the two outcomes of interest in the next section.

$\Delta u_{i,t}$ defines the error term (in difference). The second stage is:

$$\Delta y_{i,t} = \alpha_0 + \alpha_1 \widehat{\Delta I_{i,t}} + \alpha_2 \widehat{\Delta L_{i,t}} + x_i' \beta_1 + \Delta x_{i,t}' \beta_2 + \Phi(I_{i,t-1}^{pre-tax}) + \Delta \epsilon_{i,t} \ , \quad (4.11)$$

where $\widehat{\Delta I_{i,t}}$ and $\widehat{\Delta L_{i,t}}$ are the predicted changes in family after-tax income and hours worked by the mother obtained through the first stage estimates.

4.3 Data

The baseline empirical analysis exploits three different data sets: the National Longitudinal Study of Youth 1979 (NLSY79), the Current Population Survey (CPS), and the 1980 Census Integrated Public Use Microdata Series (IPUMS). While we could estimate the model using information only from the NLSY79, two potential concerns arise. First, the detailed level of heterogeneity in the construction of the labor demand shocks could suffer from small cell problems with the NLSY79 data. Second, this sample may not necessarily be informative of labor market conditions in later years at national or regional levels, as the NLSY79 is representative of U.S. Americans between 14 and 21 years of age in 1979. Therefore, we use the U.S. 1980 Census Data to calculate the employment share for each industry and group *se* at the baseline year (1980) and the longitudinal dimension of the CPS to compute the industry-specific changes in employment rates.

The National Longitudinal Study of Youth 1979 (and Children). Information about children and their families is obtained by matching the information of the mothers in the original National Longitudinal Study of Youth 1979 (NLSY79) to the additional children's survey (NLSY79-C). This matched data set (C-NLSY) results from a survey conducted every 2 years from 1986 to 2014. The sample selection rule adopted is simple; observational units are the children with information about the two main outcomes of interest, namely cognitive and behavioral development. Because the children are surveyed every two years, our empirical analysis of the model in equation (4.2) is based on 2-year changes (differences). In view of the above, our results should be interpreted as the effects of biennial changes in family income and maternal labor supply on biennial changes in children's cognitive and behavioral development.

Cognitive development is measured through achievements in math and reading activities. Specifically, we exploit the Peabody Individual Achievement Test (PIAT), a set of tests assessing proficiency in mathematics (math), oral reading and word recognition (reading recognition), and the ability to derive meaning from printed words (reading comprehension). We standardized each of the three test scores to obtain a measure with a mean of zero and a standard deviation of one. [14] We repeat the same procedure to compute an aggregate measure of math-reading achievement as the average of the three standardized single test scores.

The second outcome of interest is the Behavior Problems Index (BPI) score used as a proxy for children's noncognitive development. The BPI was created by Nicholas Zill and James Peterson to measure the frequency, range, and type of childhood behavior problems for children age four and older (Peterson and Zill, 1986). In the C-NLSY data set, five indicators for behavioral problems are collected: antisocial behavior, anxious behavior, headstrong behavior, hyperactive behavior, and peer conflicts behavior. Each index is transformed to obtain a positive scale so that higher values correspond to fewer behavioral problems. Hence, a higher index score corresponds to a higher-achieving (in terms of behavior) child. We standardize each single index to obtain a measure with a mean of zero and a standard deviation equal to 1. [15] We compute a comprehensive index, which is the mean of the five single indexes.

[14] This standardization is made on the random sample of test takers. Obviously, for several reasons based on the sample selection rule adopted in our framework, not all these observations are part of the estimation sample.

[15] This standardization is made on the random sample of individuals reporting BPI indexes.

Information about child achievement and demographics is matched with family and mothers' information such as family income, marital status, education level, etc. We exclude from the analysis children whose mothers changed marital status in two consecutive periods. We want to avoid exploiting changes in family income that are due to changes in the presence of a husband in the family. We also restrict the analysis to the period between 1988 and 2000 for two main reasons: (i) to avoid mixing EITC changes with large changes in the U.S. tax system such as the Tax Reform Act of 1986 and the two tax cuts of 2001 and 2003; and (ii) to avoid confounding the aggregate effects of the great recession after 2007.

Finally, we use information about family income and the procedure introduced in Section 4.2.2 to compute both the after-tax family income and the federal EITC for each family and period by using the TAXSIM program by Daniel Feenberg and the National Bureau of Economic Research. [16]

The Current Population Survey (CPS). The CPS data set is representative of the U.S. civilian non-institutional population. We use an integrated version of the CPS from Integrated Public Use Microdata Series (IPUMS). This data set allows us to collect data about the yearly female employment rate for each cell *se* previously described in Section 4.2.2.

1980 Census Integrated Public Use Microdata Series (IPUMS). We use the 1980 U.S. Census data from IPUMS to construct in the most precise way the employment shares for the baseline year (1980) by industry, state, and education level. Census data contain enough observations to calculate the mean employment rate for each cell defined as the combination of industry, state of residence, and education level, and to deal with possible small cell problems.

[16] TAXSIM is an ongoing project of Dan Feenberg of the NBER and his collaborators. It allows one to calculate "federal and state income tax liabilities from survey data." See Feenberg and Coutts (1993) for further details.

Table 4.1 reports the descriptive statistics for the two main samples used in the baseline analysis, one for cognitive development as measured by the combined math-reading standardized test score, and one for the analysis of the BPI used as a proxy for noncognitive development. The two samples have similar characteristics.

The average performance on the math test is slightly more than 40 (out of 100) points, between 44 and 47 (out of 100 points) for the reading recognition test, and between 40 and 43 (out of 100 points) for reading comprehension. The average BPI is 3.2 for both samples. [17] The average family in the sample reports an after-tax income of around $38,000 (median=$31,000), while mothers spend on average around 1,200 hours per year working. Children are assessed biennially with PIAT tests and BPI tests starting at ages 5 and 4, respectively, until they reach the age of 18. Children in our estimating sample are, on average, approximately 10 years old. The sample is perfectly balanced in terms of gender, while it overrepresents ethnic minorities such as blacks (32–34 percent) and Hispanics (20 percent). Only 9 percent of the sample consists of an only child, 37–38 percent have one sibling, and 53–54 percent have two or more siblings. About 65 percent of mothers are married in both estimating samples. Finally, few mothers (8 percent) are college graduates; 71 percent have at most a high school diploma.

[17] Table 4.1 also shows the values for the single five components of the BPI score.

4.4 Baseline Results

4.4.1 The Effect of Family Income and Maternal Labor Supply on Child Development

First Stage Estimates

Table 4.2 illustrates the first stage results for both the math-reading test score (columns 1–2) and the BPI score (columns 3–4). [18] All the models, at both the first and second stages, are estimated by clustering standard errors at the family level to allow for serial correlation of the error term over time and between siblings.

The diagnostic tests for the first stage (bottom part of the table) suggest that the instruments work well in our specification for both the math-reading and the behavioral analysis. Neither under- nor weak identification seem to constitute a threat to our estimates.

We start by analyzing the first stage for family income. In terms of coefficients estimates, changes in the EITC schedule generate a positive effect on family income (columns 1 and 3). A \$1,000 change in the schedule induces a \$1,026 increase in after-tax family income when math-reading test score is analyzed and \$1,101 when behavioral problems are considered. Our point estimates for the effect of changes in the EITC on family income are comparable with respect to those estimated by Dahl and Lochner (2017) and Lundstrom (2017).

Additionally, shocks in the labor demand positively affect family income. Indeed, a shift in the labor demand directly affects worker compensation and family resources. We find that an increase by 1 percent in the employment rate relative to 1980 predicts an increase of \$1,659 (math-reading first stage) or \$2,067 (BPI first stage) in after-tax

[18] For the sake of brevity, we report here only a subset of the first stage coefficients. Table C.1 reports the entire set of first stage coefficients for individual characteristics.

family income.

In columns (2) and (4) of Table 4.2 we present the first stage of hours worked by the mother. In our sample, the EITC schedule changes induce, on average, positive shifts in the maternal labor supply. The overall positive effect is generated from several different effects such as the differential impact on the extensive versus intensive margin or the differential effect for different subgroups of the population (Eissa and Liebman, 1996; Hoynes and Essa, 1996). A $1,000 change in the EITC schedule explains an average increase of around 150 hours worked per year by mothers. The effect is similar for the math-reading sample (column 2) and the BPI sample (column 4). This effect is aligned with the findings in the EITC literature summarized in Nichols and Rothstein (2016): while earlier estimates indicated that the main effect of the EITC on labor supply was in terms of extensive margins, more recent studies have found evidence of nonzero, although small, intensive margin effects.

The second instrument labor demand shocks induce changes in hours worked. We find that a 1 percent change in the employment rate relative to 1980 induces a change of around 32 (24) hours worked per year by the mother. This means that, for the average mother who works 1,258 hours per year (see Table 4.1), a 1 percent change in the employment rate in her local labor market causes an increase of approximately 1.83 percent of her labor supply. The evidence from the first stage suggests that, in our sample, labor demand shocks affect both family income and maternal labor supply, although in the last case, the coefficient is only weakly significant. [19]

Two potential concerns need to be discussed in this framework. First, we neglect possible labor supply responses by the spouse, in the case of married couples, induced by EITC changes and shocks in the labor demand. The EITC literature previously

[19] The coefficient is significant at the 10 percent level in the math-reading sample, while it is statistically insignificant in the BPI sample.

estimated small changes for the male labor supply caused by EITC changes (Hotz and Scholz, 2003; Nichols and Rothstein, 2016). However, equation (4.2) includes this endogenous reaction as part of the error term, potentially jeopardizing our identification strategy. We analyze whether the instruments are predictive of changes in the spouse labor supply to test this hypothesis. We estimate our baseline first stage specification with changes in the spouse labor supply as dependent variable. Table C.2 reports the results. Neither changes in the EITC nor labor demand shocks in the women's labor market significantly predict changes in the spouse labor supply.

A second hypothetical concern relates to the possible existence of state-specific trends in children's skills formation that might constitute a threat to the exclusion restrictions. The conditional independence of the instrument based on labor demand shocks requires that unobserved *changes* in children's skills from 1988–2000 are not correlated with the state-specific industrial compositions in the U.S. in 1980. We estimate a model that augments the baseline with the inclusion of a full set of state fixed effects to capture state trends over time. [20] First stage diagnostic tests (see Table C.5) are improved when state fixed effects are also included in the baseline model. First stage coefficients remain almost unaltered in this new setting. This suggest that, even controlling for state trends in children's skill formation, our results do not change. [21]

[20] The state-specific trends in model (4.1) become state fixed effects in our main specification (4.5). To see this point, consider our initial specification

$$y_{i,t} = \beta_0 + \alpha_{0,s}\, t + \alpha_1\, I_{i,t} + \alpha_2\, L_{i,t} + x_i'\, \beta_{1,t} + x_{i,t}'\, \beta_2 + \eta_i + \epsilon_{i,t}\ ,$$

where $\alpha_{0,s}$ is the coefficient for the state-specific trend. Taking the differences, we have

$$\Delta y_{i,t} = \alpha_{0,s} + \alpha_1 \Delta I_{i,t} + \alpha_2 \Delta L_{i,t} + x_i'\beta_1 + \Delta x_{i,t}'\beta_2 + \Delta\epsilon_{i,t}\ ,$$

where $\alpha_{0,s}$ is the state fixed effect in the difference model.

[21] We show below that second stage estimates are also unaffected by the inclusion of state trends over time.

Second Stage Estimates

Cognitive Development. We start by analyzing children's cognitive development as measured by the math-reading test score. Table 4.3 reports second stage estimates for the effect of family income and maternal hours worked. [22] Ordinary least squares (OLS) estimates in column (1) suggest a weak and positive effect (0.1 percent of a standard deviation) of income on children's achievement, while the effect of hours worked is zero. These estimates suffer from various forms of bias. Unobserved dynamics in the quality of child care and family circumstances can correlate with the effect of family income and maternal hours worked on children's development. Furthermore, measurement error is likely to affect both the measures for income and for hours worked, generating potential attenuation bias for both estimates. Finally, Løken *et al.* (2012) show that, even in the absence of endogeneity, the OLS and IV estimands can be substantially different due to differential weighting of the marginal effects.

Instrumental variable estimates in column (2) address these concerns by correcting the endogeneity of family income and maternal hours worked. Family income positively affects child cognitive achievement. A $1,000 increase in family after-tax income, *ceteris paribus*, generates an increase of 4.4 percent of a standard deviation in the math-reading test score. This result, although a different estimation framework, is aligned with Dahl and Lochner (2017). [23]

Maternal hours worked induce a significant negative effect on children's performance. A 100-hour per year increase in maternal work, *ceteris paribus*, leads to a 6 percent of a standard deviation decrease in children's math-reading test score. The

[22] The full set of coefficients, including those for individual characteristics, is reported in Table C.3.

[23] This consideration also applies in the case of OLS estimates.

size of the effect is comparable with previous findings. Bernal (2008) finds that the mother's working full-time and using child care for 1 year is associated with a 1.8 percent reduction in the child's test score (0.13 standard deviations). Bernal and Keane (2011) estimate a 2.1 percent decrease in test score as response to one year of child care instead of (single) mother care.

This finding is important when it comes to analyzing the overall effect of changes in labor earnings on child development. Indeed, policies that foster maternal labor supply, like income transfers based on employment-status criteria, generate two opposing effects: a positive income effect and a possible substitution effect induced by parental hours worked. In the next sections, we carefully analyze the drivers of the negative effect of hours worked on child development. The aim is to provide insights on how to design policies and interventions that contemporaneously foster maternal employment and child development. To anticipate the intuition, the effect of hours worked is driven by changes in parental inputs and in the quality of alternative sources of child care. Moreover, the wage rate plays a role in determining whether the income effect dominates the substitution effect of hours worked. Indeed, the wage paid shapes the marginal contribution of maternal hours worked in fostering family income.

Behavioral Development. Table 4.4 shows the analysis of behavioral development as measured by the BPI score. [24] OLS estimates display a close-to-zero effect of family income and a negative (-0.1 percent of a standard deviation), statistically insignificant effect of hours worked. IV estimates in column (2) suggest that the coefficient for family income is positive (1.3 percent of a standard deviation), although smaller than the one for cognitive development, and statistically insignificant. This result seems to suggest a differential impact of family income on the accumulation process of cognitive and noncognitive skills. While changes in family income considerably affect cognitive

[24]Table C.4 shows the full set of coefficients, including the ones for individual characteristics.

development, noncognitive development appears less sensitive (at least in the short term) to shocks in family income.

On the other hand, the effect of labor supply on noncognitive development fairly mimics the one for cognitive development. Maternal hours worked negatively affect child behavioral development. A 100-hour per year increase of maternal work causes a 5.2 percent of a standard deviation decrease in behavioral development.

The importance of accounting for the contemporaneous effects of family income and maternal labor supply on child development emerges with the analysis of the two factors in isolation. The analysis of family income without consideration of possible endogenous changes in labor supply creates a risk of underestimating the pure income effect on child development. At the same time, the analysis of labor supply without accounting for the induced income effect underestimates the (negative) effect of labor supply on child development.

Table 4.5 shows the results of the analysis. In column (1), we use our identification strategy to estimate the effect of family income in isolation on children's cognitive development. The point estimate suggests an income effect of 1.7 percent of a standard deviation. [25] In terms of comparison with the baseline model of our study (column 3), the lower point estimate for the effect of family income in column (1) is hardly surprising. The coefficient for family income captures both the positive income effect on child development and the negative effect induced by one of the main determinants

[25] Dahl and Lochner (2017) find that the effect of an additional $1,000 of family income induces children's cognitive development to increase by 4.1 percent of a standard deviation. We replicate their empirical model with our estimating sample, and we find a comparable income effect of 2.5 percent of a standard deviation. We interpret the differences in estimates as the result of differences in the compliers' groups, as a result of different sample selection criteria. In fact, Dahl and Lochner (2017) trim the data according to whether families have a relatively large change in after-tax family income between two years (see the Online Appendix for specific details). These sample selection criteria are reasonable and well-motivated in the paper, given the authors' interest in analyzing the effect of marginal changes in family resources on child development. However, in our case, sizable changes in family income can be due to changes in the extensive margin of maternal labor supply. The latter represents a valuable identifying source of variation of the causal effect of maternal hours worked on child development if the extensive margin shifts are induced by our instrumental variables.

of positive income shocks, namely increases in individual labor supply. Behavioral development (columns 4 and 6) displays the same pattern. The coefficient for family income becomes considerably smaller in size, -0.3 versus 1.3 percent of a standard deviation, in the model using only family income as the endogenous regressor. The previous explanation for cognitive development also applies to this case.

Columns (2) and (5) focus on maternal hours worked in isolation. Coefficients display a smaller effect of maternal labor supply both for cognitive and behavioral development when compared to the reference baseline models in columns (3) and (6), respectively. For cognitive development, the effect switches from -2.1 to -6 percent of a standard deviation. For behavioral development, the change moves from -4 to -5.2 percent of a standard deviation. These changes confirm that the coefficient for maternal labor supply, when analyzed in isolation, captures both the labor supply effect and the positive income effect induced by increases in individual labor supply.

Given the strong interdependence between maternal labor supply and family income, there is no suitable identifying source of variation that is likely to exclusively affect one variable of interest. Hence, as shown above, in order to identify the single causal effect of either family income or maternal labor supply on child development, it is necessary to allow for the endogeneity of both inputs.

We will now discuss some potential threats to our IV framework validity. As introduced, the possible existence of state-specific trends in children's skill formation might undermine our exclusion restrictions. We take into account this potential concern by augmenting the model with state-specific trends in children's skills formation. Such inclusion does not affect the results. [26] Table C.5 shows that point estimates for the effect of changes in family income and hours worked are almost unchanged

[26]See Section 4.4.1 for the first stage analysis of this model with state-specific trends in children's skills formation.

with respect to the models without state fixed effects. The replication of all the other analyses of the study including state fixed effects does not remarkably alter any of the results. [27] For this reason, we have decided to report in Table C.5 the baseline estimates obtained by including state fixed effects in the model, while in the rest of the work we report results without controlling for state fixed effects.

The exclusion from the set of regressors of variables capturing school financial and economic resources might bias our results by violating the exclusion restriction for the labor demand shocks instrument (see discussion in Section 4.2.2). In Table C.6, we deal with this potential concern by including changes over time of school finances and economic resources at the state level, therefore testing whether these variables were part of the error term of the model.

We use data about school resources from the CDD National Public Education Financial Survey, and we focus attention on two different measures. [28] First, we collect data on total revenues per pupil, measured as the total revenues from all sources divided by the fall membership. Second, we collect the total current expenditure per pupil, defined as the total current expenditure for public elementary and secondary education divided by the fall membership. We augment the baseline model by adding both variables expressed in difference with respect to the previous period.

Results highlight two main patterns. On the one hand, neither changes over time in revenues nor expenditures are statistically significant predictors of child cognitive and behavioral development. On the other hand, point estimates for both family income and hours worked by the mother are unchanged with respect to the specifi-

[27] Results are available upon request.

[28] The CDD National Public Education Financial Survey has a primary purpose of making available to the public an annual state-level collection of revenues and expenditures for public education for students in prekindergarten through grade 12.

cations without controls for school financial and economic resources. Also first stage diagnostic tests, as shown by the tests in the bottom part of the table, are unaffected in this new model specification.

The work by Goldsmith-Pinkham *et al.* (2017) points out that, in general, labor demand shocks can include pre-trends that can indirectly affect the dependent variable, which may jeopardize the validity of our identification strategy. To test this hypothesis, Goldsmith-Pinkham *et al.* (2017) recommend testing whether future values of the instrumental variable are predictive for current second stage residuals. Table C.7 shows the hypothesis testing for the presence of pre-trends. We test for pre-trends with different specifications with different lagged variables, up to a maximum of 6 lagged years (3 model-periods as observations are collected every 2 years). We do not find evidence of pre-trends. In all cases, future labor demand shocks are not predictive of past child test scores. The only exception appears for the most adjacent case of the 1-period lag for cognitive measures. However, by extending the analysis to 2 periods or 3 periods of lagged variables, any relationship between future labor demand shocks and cognitive test scores arises.

Furthermore, as the instrument for labor demand shocks is state-specific, we address the potential concern due to possible endogenous household changes in state of residence from one period to another. In our sample, a very small fraction of families change their state of residence in two following periods. [29] To be conservative, we replicate our baseline analysis and restrict the sample to those households maintaining the same state of residence across two consecutive periods. The analysis in Table C.8 does not pinpoint any significant effect on results.

[29] In our estimation samples, there are 581 (math-reading sample) and 690 (BPI sample) cases of mothers who changed their state of residence during the two-year intervals when test scores and behavioral indexes are measured. In both cases, it represents approximately 5 percent of the entire sample.

Finally, tax reforms may have heterogeneous effects within groups (Hoynes and Essa, 1996). Because of the structure of the EITC benefits, mothers who are working and fall into the phase-out section of the schedule may have incentive to reduce their hours worked, compromising the monotonicity assumption of the EITC instrument. This assumption is needed to interpret our IV results as the local average treatment effect (see Imbens and Angrist, 1994). Even if monotonicity is untestable, we consider the potential heterogeneous effects induced by the change in the EITC for employed versus non-employed mothers. [30]

Table 4.6 shows the results. First stage estimates do not provide any evidence of failure of the monotonicity assumption. The effect of changes in the EITC schedule on family income is positive both for employed and non-employed mothers. Shocks in the labor demand display a similar coefficient with respect to the one of the baseline analysis. Changes over time in the EITC benefits also positively affect maternal labor supply. Second stage results are similar to the ones in the baseline analysis. The effect of family income on the math-reading test score is positive and strongly significant, while maternal hours worked negatively affect the child's cognitive development. Additionally, the analysis of behavioral development (column 2) conveys the same message as the one in the baseline analysis of BPI.

In sum, this analysis shows that, at least in the above-considered dimension, we cannot reject the monotonicity assumption. Moreover, our results are not affected by using possible heterogeneous responses to changes in the EITC schedule as possible instruments for family income and maternal hours worked.

[30]Information about employment status refers to the previous period to mitigate possible endogeneity concerns.

Decomposition of the Overall Effects

We focus here on the analysis of each single component of our aggregate measures for cognitive and behavioral development. Such decomposition is important as it allows us to understand whether the overall effect shown in the baseline analysis is general or is driven by some specific measures for children's achievements. Table 4.7 reports the decomposition of the combined math-reading test score in its three single components: math, reading recognition, and reading comprehension. The three tests in isolation confirm the existence of a positive and significant effect of family income on test performance counterbalanced by a negative impact of hours worked by the mother. The income effect appears slightly smaller in size (2.9 and 3 percent of a standard deviation) for math and reading comprehension (columns 1 and 3) when compared to reading recognition (column 2). In terms of hours worked, the effect is particularly sizable for reading recognition (-7 percent of a standard deviation) and reading comprehension (-4.9 percent of a standard deviation), while it is smaller for math (-3.6 percent of a standard deviation).

This evidence is suggestive of possible channels underlying the effect of maternal hours worked. At least two mechanisms potentially explain the results: (i) an endogenous reallocation of maternal time that values more schooling activities rather than reading; and (ii) a productivity gap of maternal time between math and reading.

We replicate the same decomposition analysis for indexes for behavioral development (Table 4.8). We analyze the following five components: antisocial behavior, anxious behavior, headstrong behavior, hyperactive behavior, and peer conflicts behavior. With the exception of hyperactive behavior (column 4), behavioral problems are not affected by family income. On the contrary, hours worked display a negative and significant (with the exception of anxious behavior in column 2) effect on behavioral problems, with point estimates bounded between -3.6 and -4.8 percent of a standard deviation.

The analysis of single behavioral indexes suggests similar insights with respect to the aggregate BPI index. Family income seems to play a very marginal role in shaping, at least in the short term, children's behavioral problems. Concurrently, the time spent with the mother is a relevant input in terms of children's behavioral development.

4.4.2 Early Childhood Development

Until this point we have considered measures for cognitive performance and behavioral problems for children older than 5 and 4 years old, respectively. We now extend the analysis to early childhood development. The C-NLSY data set contains information about temperament measures collected between ages 1–7. We focus our attention on three specific measures collected for children in this age range: compliance, insecure attachment, and sociability. [31] As for BPI, these measures are also expressed on a positive scale, meaning that higher values correspond to fewer temperament problems. We standardize each of the three measures to make an index with a zero mean and a unitary standard deviation. Because compliance and insecure attachment are collected for children in the same age range, we also construct an aggregate average index of the two.

Table 4.9 illustrates the analysis of the effect of family income and maternal hours worked on early childhood development. We estimate each model as in the baseline analysis. Despite the lower level of precision due to the reduced sample size, point estimates show a similar pattern to the one identified in the main analysis on older

[31] The NLSY79 contains other measures for child development in this age range. However, these are the only measures repeated over time, which therefore allow a dynamic analysis in first differences.

children.

The coefficient for family income is always positive and similar in size to that of the baseline model for the math-reading test score. For example, a $1,000 change in family income explains a (statistically insignificant) increase of 4.6 percent of a standard deviation in the compliance score (column 1). At the same time, the coefficients for maternal hours worked are negative, with a range between -1.0 (sociability, column 4) and -5.3 (compliance and insecure attachment, column 3) percent of a standard deviation. These magnitudes are similar to those found with respect to cognitive and behavioral development.

The analysis of early childhood provides supportive evidence that at this developmental stage there might also be a contemporaneous effect of family income and maternal hours worked on child development. The pattern seems to mimic the one identified for cognitive and behavioral development. Due to the limited sample size, the effects on early childhood development are not precise and require further analysis to infer more conclusive insights.

4.5 Hours Worked and Child Development: To the Roots of the Result

In this section we study the mechanisms behind the negative impact of maternal hours worked on child development. This understanding is crucial for designing policies that contemporaneously foster maternal employment and child development.

4.5.1 Time Investment in the Child

Parental inputs determine child development (Cunha and Heckman, 2008; Cunha *et al.*, 2010; Del Boca *et al.*, 2014a; Heckman and Mosso, 2014a; Agostinelli and Wiswall, 2016a). The choice to increase maternal labor supply may generate a displacement effect in terms of maternal investment in the formation of children's skills.

It is then important to establish whether maternal hours worked affect parental investment in the child. [32]

Time diary data allow us to observe maternal response in terms of time investment in the child as a result of her labor supply. We combine data from the American Time Use Survey (ATUS) and the American Heritage Time Use Survey (AHTUS), which provide information about the amount of time people spend doing various activities, such as paid work, child care, volunteering, and socializing. [33] For similarities with our estimating sample in the C-NLSY, we focus our attention on households with at least one child in the same age range of the baseline analysis in the period 1985–2003. [34]

We collect all the information about family income, hours worked by the mother, education of the mother, household composition (e.g. single-head household, number of children, etc.), child's age, and four measures of parental investment in child development. The available measures for parental investment are physical child care, helping with homework, reading and playing with the child, and a residual category containing other forms of child care. We also construct an aggregate measure that is the sum of the four mentioned child care activities. All the measures for time investment are expressed in hours per week.

Figure 4.4 shows the estimates of the five regressions of each time investment measure on family income and maternal hours worked plus a set of controls for mother's age, household composition, number of siblings, child's age, and year fixed effects. Each panel of the figure represents the regression coefficient, together with its 95

[32] The mother, as a response to an increase of hours worked, may decide to decrease parental inputs and child investment or to decrease leisure activities to try to keep the amount of time devoted to the child fixed.

[33] See www.ipums.org/timeuse.shtml for further details.

[34] Our sample selection is based on the availability of the surveys. We start with the 1985 AHTUS. We use the 2003 ATUS to increase the sample size of the analysis.

percent confidence interval, for maternal hours worked (Panel A) or family income (Panel B) on each measure for time investment in child care activities. [35]

As shown in Panel A, maternal hours worked are negatively correlated with parental time investment in all five considered activities. As an example, an increase of 1 hour worked per week predicts a 4-minute decline per week in child care time (total child care). In other words, the result is equivalent to an average decrease in child care of approximately 2 hours per week if the mother starts working full time (from 0 to 35 hours per week).

Panel B reports the results for family income obtained from the same models. All the coefficients are close to zero and statistically insignificant. In our sample, higher family income does not correlate with changes in parental time investment in the child. These results only suggest general insights; they do not deal with factors such as the quality of time parents spend with their children. Section 4.5.4 discusses that.

4.5.2 Income versus the Substitution Effect: The Role of Wages

We exploit the results of the main analysis to explain the drivers behind the average negative impact of maternal hours worked on child development. An increase in maternal hours worked generates an income effect (higher earnings) and a substitution effect (displacement of maternal time) (Heckman and Mosso, 2014a; Del Boca *et al.*, 2014a).

Given the specification in equation (4.1), the causal effect of maternal hours worked on child achievement can be deconstructed in these two mechanisms as follows:

$$\frac{\partial E\left[y_{i,t}|L_{i,t}\right]}{\partial L_{i,t}} \equiv \underbrace{\alpha_1 \cdot \frac{\partial E\left[I_{i,t}|L_{i,t}\right]}{\partial L_{i,t}}}_{\text{Income Effect}} + \underbrace{\alpha_2}_{\text{Substitution Effect}} . \tag{4.12}$$

[35]It is important to recall that although the effect of maternal hours worked and family income are displayed in different panels, their coefficients are contemporaneously estimated with the same regression. Appendix Table C.9 shows the regression results.

By decomposing the total family income in the mother's after-tax earnings ($w_{i,t} \cdot L_{i,t}$) where $w_{i,t}$ represents the wage, and any other source of income ($\widetilde{I}_{i,t}$), we can rewrite equation (4.12) as: [36]

$$\frac{\partial E\left[y_{i,t}|L_{i,t}\right]}{\partial L_{i,t}} \equiv \alpha_1 \cdot \left(w_{i,t} + \frac{\partial E\left[\widetilde{I}_{i,t}|L_{i,t}\right]}{\partial L_{i,t}} \right) + \alpha_2 \ . \tag{4.13}$$

Equation (4.13) conveys a clear message: the effect of hours worked on children's achievement is ambiguous in sign and heterogeneous within the population. Given a wage rate $w_{i,t}$, the total effect in equation (4.13) depends on the relative magnitude of the income effect (α_1) in contrast to the substitution effect (α_2). Additionally, the income effect depends on the specific wage rate $w_{i,t}$, suggesting heterogeneous effects of maternal hours worked on children's outcomes. We investigate heterogeneity according to mother and child characteristics in the next section, while here we focus the attention on the role played by wages.

The effect of hours worked by the mother strictly depends on factors such as labor market conditions. The recognition of sufficiently high wages potentially overcomes the substitution effect induced by decreased maternal time invested in child development. This is likely driven by the fact that the mother might be able to substitute her own input by purchasing higher quality alternative sources of child care (e.g. nonparental care, additional school, youth clubs, sport and music activities, etc.).

Figure 4.5 graphically shows the importance of the paid wage by exploiting our baseline results for the effect of maternal hours worked on child development. [37] The analysis assumes that other sources of income do not respond to changes in maternal

[36] This is the case when the mother is already working ($L_{i,t} > 0$). For the extensive margin case, the causal effect is $\frac{E[y_{i,t}|L_{i,t}]}{\partial L_{i,t}} = \alpha_1 \cdot \left(w^*_{i,t} + \frac{\partial E[\widetilde{I}_{i,t}|L_{i,t}]}{\partial L_{i,t}} \right) + \alpha_2$, where $w^*_{i,t}$ is the counterfactual wage she would receive once she works.

[37] The figure is based on the estimates in Table 4.3, column (2). As we do not find a significant income effect for behavioral development, we base this analysis exclusively on cognitive development.

labor supply, and the income effect is determined only by changes in earnings. The figure shows the heterogeneous effect of maternal hours worked on children's cognitive development with respect to maternal hourly wage. The intersection of the solid line (effect of hours worked) with the dashed horizontal line representing a zero net income-substitution effect, highlights that up to a corresponded wage of around \$13.50 per hour, the effect induced by the extra labor income (income effect) is not enough to compensate for the loss in child development induced by decreased maternal input (substitution effect). For wages higher than \$13.50 per hour, the income effect dominates the substitution effect.

In the background of Figure 4.5, we plot the wage distribution for both single and married mothers in our NLSY79 estimation sample, which provides an intuition on the determinants of the negative effect of maternal hours worked on child cognitive development. The biggest fractions of the wage distributions are located below the wage threshold corresponding to a zero or positive effect of maternal labor supply on child achievements. These results call for a policy discussion regarding the importance of labor market conditions and opportunities especially when it comes to women and, specifically, mothers. Moreover, such findings should spark a discussion on fiscal reforms and the minimum wage.

4.5.3 Heterogeneous Effects of Maternal Hours Worked

In this section we replicate the baseline analysis by focusing on various subpopulations of interest. The aim is to understand whether the negative effect of hours worked on child development might be driven by differences in the quality of the alternative child care inputs used in substitution of maternal inputs or by other child characteristics such as race or age. Bernal and Keane (2011) show that informal care (grandparents, siblings, etc.) has adverse effects on child development as measured through test scores, and that more than 75 percent of single mothers use informal care. Mothers with a higher educational level or with higher skills are likely to use higher quality alternative inputs for their children, therefore possibly mitigating the negative impact induced by their increase in individual labor supply.

The analysis is based on five different sources of heterogeneity: maternal educational level, the Armed Forces Qualification Test (AFQT) as a proxy for maternal skills, maternal marital status, child's race, and child's age. [38] We compare maternal educational levels by dividing the sample in two groups: mothers with at most a completed high school degree (*Low education*) and mothers with some college education or more (*High education*). In terms of maternal skills, we separate mothers according to the median value of the AFQT test by labeling the ones with lower-than-the-median AFQT as *Low AFQT*, and those with higher-than-the-median AFQT as *High AFQT*. We analyze marital status by comparing married mothers with unmarried mothers. To take into account the possible differential effects of hours worked for minority populations, we also compare the white population with the black and Hispanic populations. Finally, the effect induced by maternal labor supply might be larger when the child is younger and needs more supervision and parental care. We look at potentially heterogeneous impacts of family income and maternal hours worked on child development according to child's age by dividing the sample into children under and over 12 years old. [39]

Table 4.10 reports estimates by subpopulations according to mother's education

[38] The Armed Forces Qualification Test (AFQT) was derived from the Army General Classification Test in 1950, and it is widely recognized as a reliable measure of mental ability. The AFQT score is not available for all the observations in the sample. Therefore, the sample size for this analysis is slightly reduced with respect to the one in the baseline models.

[39] To assess the importance of the heterogeneous treatment effects in our estimating sample, we decompose our predicted exogenous changes in our two endogenous variables in a two-stage least squares fashion, where we allow the second stage coefficients for income and hours worked to vary by mother's level of education, AFQT, marital status, child's race, and child's age. We implement a family-level clustered bootstrap procedure (100 repetitions) to adjust standard errors.

(Panel A), AFQT score (Panel B), and marital status (Panel C). Column (1) displays the analysis of the combined math-reading test score. The differential impact of family income appears negligible. Coefficients are similar across subgroups for all sources of heterogeneity. [40]

The impact of maternal hours worked is indeed characterized by high heterogeneity. Considering maternal education as a source of heterogeneity, the negative effect of hours worked shown in the baseline analysis seems to be driven by the subgroup of mothers with a low educational level. For this group of mothers, an increase of 100 hours worked per year explains a decrease in standardized test scores by 5.8 percent of a standard deviation. The effect for the more educated counterpart is zero. The analysis of maternal skills and marital status unveils similar heterogeneous patterns. Maternal hours worked do not affect child cognitive development when mothers have high AFQT, while the effect of hours worked is negative and significant (-6.4 percent of a standard deviation) for low-AFQT mothers. Concerning marital status, the coefficient for hours worked is significant and negative (-6.9 percent of a standard deviation) for unmarried mothers, while the effect of maternal labor supply is statistically insignificant for married mothers.

The presented heterogeneous analysis suggests that parents from more advantaged backgrounds and with more resources, proxied by education, skills level, and marital status, might employ high-quality alternative inputs for the child when there is an increase in individual labor supply. Alternatively, they are able to more productively substitute the quantity of time with the quality of time devoted to their children.

The heterogeneous impact of maternal labor supply on child development is not confirmed for behavioral development (Table 4.10, column 2).The effect of family in-

[40] A small but more pronounced difference appears for marital status, with unmarried mothers displaying a slightly larger effect than married mothers.

come and hours worked is similar across groups. More precisely, any differential impact of maternal hours worked across subpopulations is detected neither for mother's education (Panel A), nor for mother's AFQT (Panel B), nor marital status (Panel C). These results suggest potentially different mechanisms underlying the cognitive and behavioral skill production functions. In particular, it is easier to substitute for parental time with activities related to cognitive development but more difficult to substitute for parental time with activities related to a child's behavioral development. Further research on this point is needed.

Table 4.11 extends the analysis to child characteristics. In terms of cognitive development (column 1), the analysis by race (Panel A) displays similar effects (around 4.6 percent of standard deviation) of family income across subgroups. We find a negative effect of maternal hours worked for both the subgroups of white and black or Hispanic. Although the point estimates across race subgroups are not significantly different, it is interesting to notice that the point estimate is larger in magnitude (-6.9 percent of a standard deviation) for black or Hispanic children than for white children (-4.7 percent of a standard deviation). Also the analysis by age (Panel B) highlights an interesting pattern. While the effect of family income is similar across age groups, the impact of maternal hours worked is more relevant for younger children (<12 years old). Relatively younger children report a statistically significant negative effect of maternal labor supply (-7.6 percent of a standard deviation), while the same coefficient is statistically insignificant and smaller (-5.3 percent of a standard deviation) for relatively older children. The evidence of heterogeneous impact by age suggests that the importance of parental input and investment in shaping child development is confirmed in all stages of childhood, although it seems to be dominant when the child is relatively younger.

When behavioral development is considered, child characteristics do not display heterogeneous patterns (Table 4.11, column 2). In general, the income effect is always statistically insignificant and similar across subpopulations. The effect of maternal hours worked is indeed negative and strongly significant for all the subpopulations of interest.

4.5.4 Employment, Child's Activities, and Quality of Child Care

Section 4.5.1 shows that maternal hours worked are negatively correlated with parental investment in child care. We analyze whether maternal employment status and family income play a role in explaining differences in the type and quality of investments. In particular, we investigate to what extent our heterogeneous results of maternal hours worked in Section 4.5.3 might depend on the quality of alternatives sources of child care.

We draw on data from the Child Development Supplement (CDS), a research component of the Panel Study of Income Dynamics (PSID), to analyze investment in child development. [41] In 1997, the PSID complemented its main data collection with additional information on 0-12 years old children and their parents. The aim was to provide researchers with a comprehensive, nationally representative, and longitudinal data set of children and their families with which to study the process of early human capital formation. We focus on the 1997 wave of the CDS (CDS-I) as it contains a wide set of information about parental investment in the child, child's activities, and time diary data for 3,563 children from 2,394 families.

Table 4.12 shows the analysis of a set of proxies for parental investment in child development. We compare values across four different subgroups of households:

[41] The PSID is a longitudinal study of a representative sample of U.S. individuals and the families in which they reside. Since 1968, the PSID has collected data on family composition changes, housing and food expenditures, marriage and fertility histories, employment, income, time spent on housework, health, consumption, wealth, and more. See psidonline.isr.umich.edu for further information about the data set.

low-income and non-employed mother (LI,NE), low-income and employed mother (LI,E), high-income and non-employed mother (HI,NE), and high-income and employed mother (HI,E). This comparison allows us to disentangle: (i) differences in maternal investment and child's activities according to family income level, and (ii) the difference in investment and child's activities between employed and non-employed mothers conditional on family income. Low- and high- income families are defined according to the median value for family income in the CDS-I sample ($35,000). The employment status refers to the year 1997. The table reports average values for the four subgroups (columns 1–4), together with the difference between employed and non-employed mothers conditional on income group (columns 5 and 7), and its statistical significance (columns 6 and 8). [42]

Panel A of the table depicts proxies for parenting styles. Behavior such as encouraging child's hobbies, showing physical affection, attending parenting classes, having the child cared for by others, or the use of rules to discipline the child display a similar pattern. Both low- and high- income families report insignificant changes across employment status (column 1(3) versus column 2(4)) or the change is similar across income groups (column 5 versus column 7).

On the other hand, diverging patterns arise in terms of monitoring activities perpetrated by parents. Low-income families put into practice more monitoring when the mother is employed. For example, employed mothers report higher levels of control over child's companions (+3 percent), activities after school (+6 percent), and homework time (+8 percent) when compared to the non-employed counterpart. Mothers with high incomes behave in the opposite way. In this case, we observe a decrease in

[42] Unless differently specified (e.g. in the case of a time diary), all variables in the table are constructed as dummy variables. The questionnaire contemplates "Yes/No" answers (e.g. encourage hobbies) for some of the investments or activities, while in other cases, a more detailed list of options is available (e.g. "Very likely", "Somewhat likely", "Not sure how likely", "Somewhat unlikely", "Not at all likely"). Appendix C.2 explains variable definitions and construction in detail.

investment for employed mothers (-11, -13, and -10 percent, respectively).

Panel B focuses on investment in their child's scholastic performance by parents. We observe a diverging pattern across income groups when we analyze activities such as contacting the faculty, keeping a closer eye on the child's activities, lecturing the child, encouraging the child to work harder, and helping the child with schoolwork. Results in column (6) highlight that any significant change is detected for low-income mothers. These mothers do not react differently to possible poor scholastic performance when they are employed as opposed to when they are not employed. Mothers from high-income families behave differently. They increase contact and discussion with faculty by around 7 percent (p-val=0.01) relative to non-employed mothers. They lecture their child more (+6 percent, p-val=0.04), and they prompt the child to work harder more often (+7 percent, p-val=0.04).

In Panel C, we analyze family environment scales to describe the environment to which each child is exposed. Scales are obtained as the combination of information collected in the data set (e.g. parental reaction to child's behavior, ways of showing physical affection to the child, etc.). [43] Four different scales are available: the general home scale, the cognitive stimulation scale, the emotional support scale, and the parental warmth scale. High-income families outperform low-income families. Concerning the maternal employment status, we find that the presence of employed mothers is almost always correlated with an increase in home scales. The increase is similar across income groups, although slightly larger in size for low-income families.

Finally, in Panel D of the table we use time diary data to study differences in the daily activities of the child. School attendance is similar across income groups. In general, children from families with non-employed mothers attend less school (around

[43] Refer to psidonline.isr.umich.edu and to the CDS-I User Guide Supplement for additional information about the construction of family environment scales.

12,000 seconds per day) than children with employed mothers (around 16,000 seconds per day). If the average school quality differs across income groups (e.g. high-income mothers living in better neighborhoods with high quality schools, etc.) this might produce a differential effect related to maternal employment.

We then focus on activities usually considered as potentially detrimental for child development. [44] Time spent watching television highlights an interesting pattern: in both income groups, children with employed mothers tend, or at least declare, to watch less television. This is probably due to a lower amount of time spent at home. However, while the average decrease in television watching in low-income families is 221 seconds per day, the same decrease is double for children from high-income families (522 seconds per day). [45] Similarly, maternal employment is correlated with an increase in the time spent playing electronic games exclusively for the subgroup of children from low-income families. Indeed, the employed versus non-employed differential is sizable (+172 seconds per day, p-val=0.10) for children from low-income families, while it is close to zero and statistically insignificant (+27 seconds per day, p-val=0.73) for children from high-income families.

Educational activities, such as art and sculpture, highlight an opposite income-related pattern. Children from high-income families do not display any significant change due to maternal employment status (-30 seconds per day, p-val=0.56), while a significant decrease arises for low-income families when employed and non-employed mothers are compared (-119 seconds per day, p-val=0.01). The change in time devoted to reading and looking at books is similar across employment statuses for both income groups. Children from low-income families tend to increase the time devoted to visits to other persons as a response to maternal employment relatively more than children

[44] A consistent fraction of individuals in the sample report zero seconds for such activities; this explains the apparently low average values displayed in the table.

[45] These values are statistically insignificant for both income groups.

from high-income families.

The evidence in this section suggests potential differential patterns in the quality of investment in child development for employed and non-employed mothers across income groups. Employed, high-income mothers tend to substantially decrease their control over their child's activities (when compared to their high-income, non-employed counterpart), unless they become aware of their child's poor academic performance. Low-income mothers behave in the opposite way by increasing control as a response to employment (lower trust in the alternative inputs used). Moreover, children from low-income families seem to engage more with respect to their wealthier counterparts in activities potentially detrimental for their development as a response to maternal employment. These results might help explain the overall effect of maternal hours worked on child development and the stronger impact of labor supply (on cognitive development) for mothers from low socio-economic backgrounds.

4.6 Conclusion

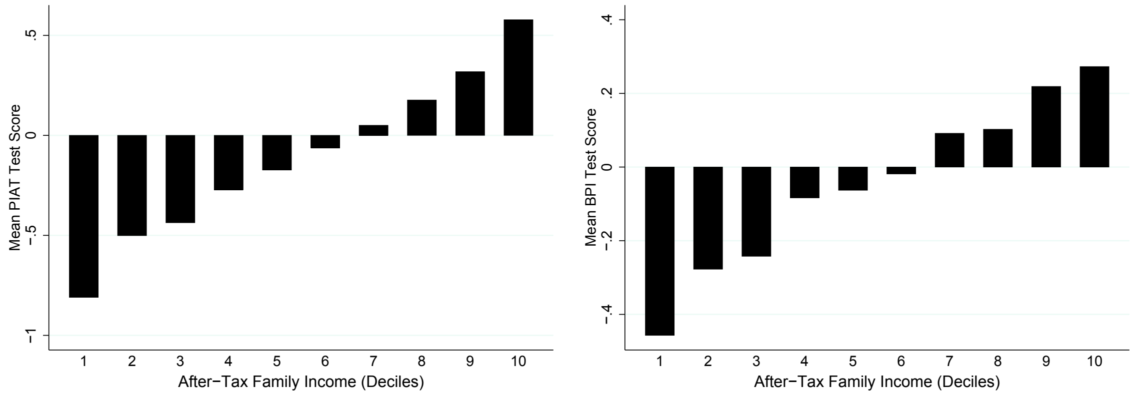
This paper unveils the contemporaneous effect family income and maternal hours worked have in shaping child development. We combine the analysis of cognitive and noncognitive development. We exploit children's performance on standardized tests to measure cognitive development. We use indicators of behavioral problems to gauge noncognitive development.

We find that family income has a sizable and positive effect on cognitive development, while the income effect is negligible (although positive) on behavioral development. The effect of maternal hours worked is the same across outcomes. On average, hours worked by the mother negatively affect both cognitive and behavioral development.

We shed light on the mechanism behind the negative effect of maternal hours worked on child development. Working mothers invest less time in child care. As a consequence, the choice of alternative sources of child care becomes crucial; this choice is likely to be affected by economic factors. We decompose the overall effect of maternal hours worked on child development into an income effect (higher earnings) and a substitution effect (less maternal time). We find that the substitution effect tends to dominate the income effect when the after-tax hourly wage is less than $13.50 per hour. With higher earnings, families are able to substitute their decreased time investment with better and more productive alternatives. In line with this explanation, we show that the average effect (on cognitive development) is mainly driven by low-income, less-educated families and that the employment effect on investment in the child differs according to family income.

Several policy suggestions derive from our results. The trade-off between the income and substitution effect in terms of child development encourages a debate about the effect of conditional versus unconditional cash transfers. Income subsidies that provide monetary transfers based on work requirements might produce heterogeneous impacts in terms of child development. Our analysis confirms that policies aimed at fostering maternal labor supply benefit child development when considered in conjunction with well-researched policies concerning the optimal level of family income taxation or the optimal minimum wage. Alternatively, policies that encourage maternal employment in low-income families should also guarantee alternative sources of child care to support child development.

Figure 4.1: Children's Outcomes by After-Tax Family Income Deciles

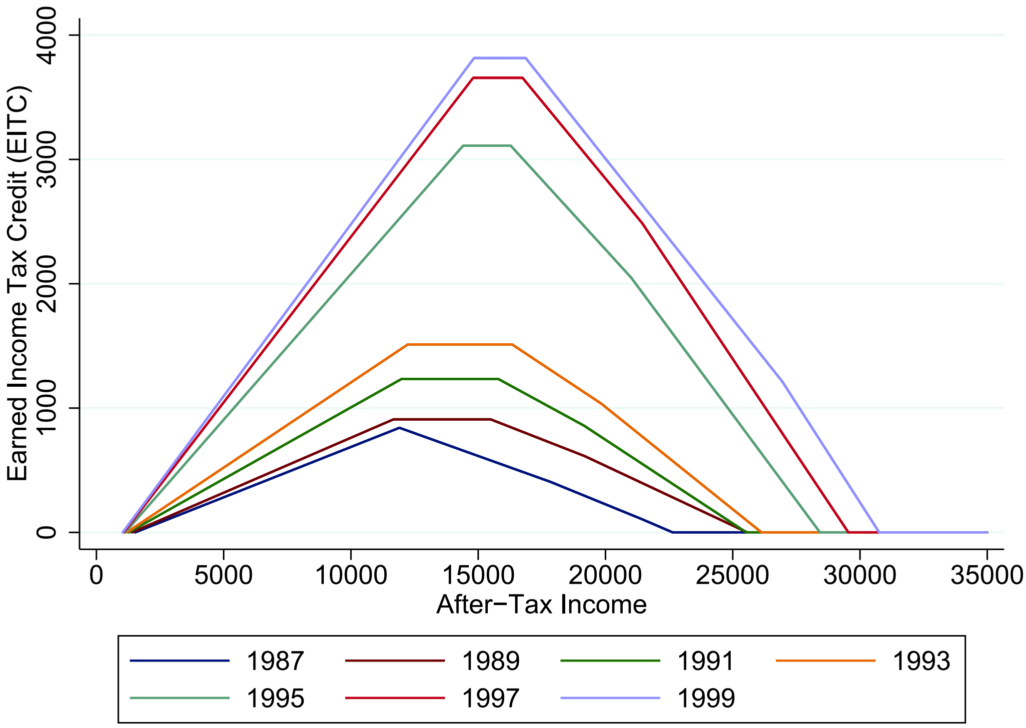


Panel A: Average PIAT Score

Panel B: Average BPI Score

Notes: This figure shows the average cognitive and behavioral outcomes of children by after-tax family income deciles. For cognitive measures we consider the average standardized PIAT score between the math, reading recognition, and reading comprehension indexes. The behavioral measure is the average of the standardized Behavior Problems Index (BPI). The after-tax family income is calculated using the TAXSIM program. Source: C-NLSY.

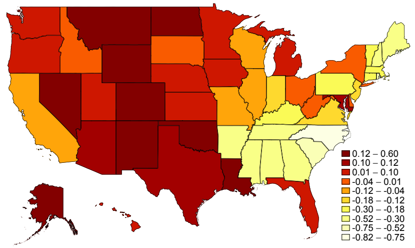
Figure 4.2: The EITC Expansion



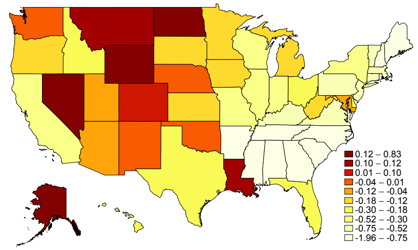
Notes: This figure shows the changes in the federal EITC schedule for families with two children. The after-tax family income is in real (2000) dollars. We calculate the EITC benefits over time using the TAXSIM program.

Figure 4.3: Labor Demand Shocks

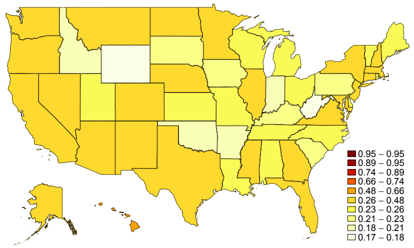
Panel A: High School Dropouts, 1988



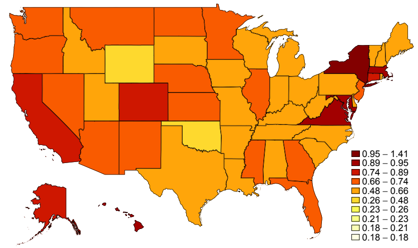
Panel B: High School Dropouts, 2000



Panel C: College Graduates, 1988

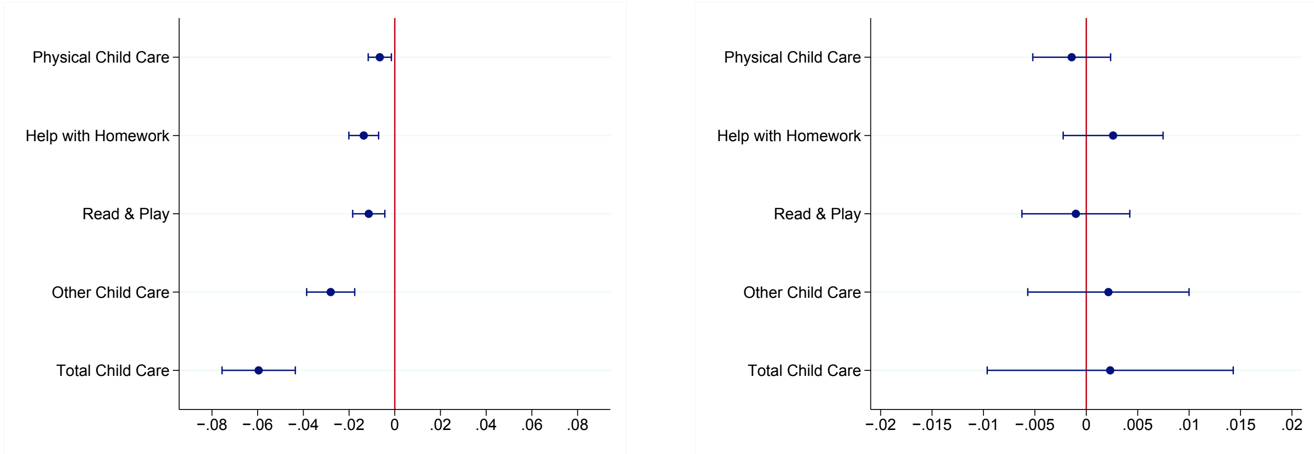


Panel D: College Graduates, 2000



Notes: This figure shows the variation in labor demand shocks between states and over time for less educated (school dropouts) and highly educated (college graduates) women. Panels A–B show the variation of labor demand shocks for the less educated group. Panels C–D show the variation of labor demand shocks for the highly educated group. Sources: CPS and Census 1980.

Figure 4.4: Time Allocated to Child Care, Mother's Hours Worked, and Family Income

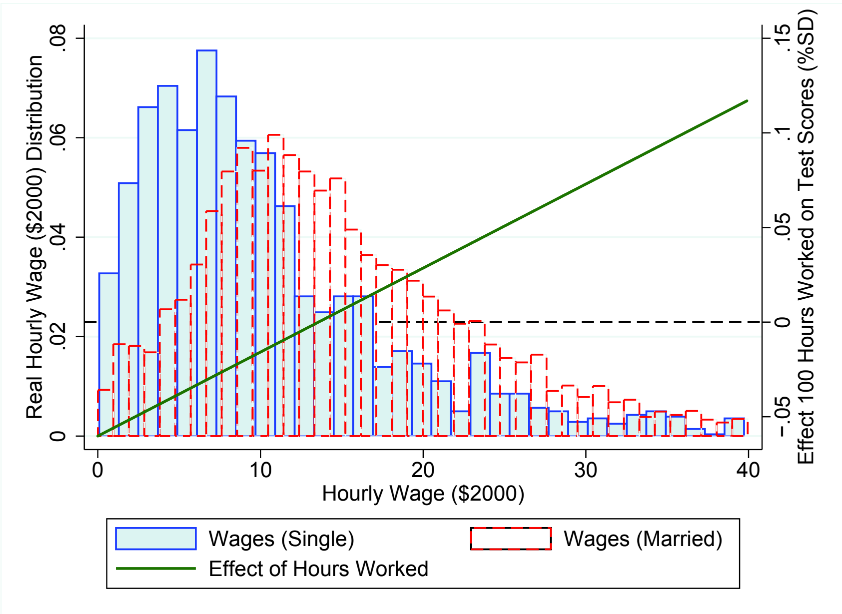


Panel A: Child Care and Mother's Hours Worked

Panel B: Child Care and Family Income

Notes: This figure shows the effect of hours worked and family income on time (hours per week) allocated to child care activities. Panel A displays the regression coefficients (with a 95% confidence interval) for the effect of hours worked on each measure for time investment in child care activities. Panel B displays the regression coefficients (with a 95% confidence interval) for the effect of family income on each measure for time investment in child care activities. See text for further details. Sources: ATUS and AHTUS.

Figure 4.5: The Effect of Maternal Labor Supply on Child Achievement



Notes: This figure shows the causal effect of maternal hours worked on child achievement as a function of mothers' hourly wage rate (green line). The plotted values in the background show the empirical distributions of real hourly wages ($2000) for single and married mothers (top 5% excluded). The solid line represents the overall effect of maternal labor supply (income and substitution effects) based on our baseline results in Table 4.3, column (2).

Table 4.1: Summary Statistics

	Combined Math-Reading		Behavior Problems Index	
	Mean	St.Dev.	Mean	St.Dev.
	(1)	(2)	(3)	(4)
Math	43.62	13.55	40.54	15.28
Reading recognition	47.29	16.05	43.98	17.57
Reading comprehension	42.60	13.70	40.02	14.97
Behavior Problems Index	3.22	1.13	3.23	1.13
Antisocial	4.49	1.59	4.50	1.59
Anxious	3.29	1.47	3.32	1.47
Headstrong	2.64	1.67	2.64	1.67
Hyperactive	3.23	1.60	3.20	1.60
Peer conflicts	2.49	0.84	2.49	0.84
Family income	37,775	30,132	38,463	30,701
Hours worked (Y)	1,258	986	1,234	982
Age	10.69	2.31	10.11	2.57
Male	0.50	0.50	0.50	0.50
White	0.46	0.50	0.48	0.50
Black	0.34	0.47	0.32	0.47
Hispanic	0.20	0.40	0.20	0.40
No siblings	0.09	0.28	0.09	0.29
One sibling	0.37	0.48	0.38	0.49
Two or more siblings	0.54	0.50	0.53	0.50
Mother's marital status:				
Married	0.63	0.48	0.65	0.48
Mother's education:				
High school dropout	0.22	0.41	0.21	0.40
High school graduate	0.49	0.50	0.50	0.50
Some college	0.21	0.41	0.21	0.41
Graduated college	0.08	0.27	0.08	0.28
Observations	12,288		13,777	

Notes: This table shows the summary statistics of our estimating samples. Columns (1) and (2) refer to the estimating sample for the analysis of child cognitive development (combined Math-Reading test score). Columns (3) and (4) consider the estimating sample for the analysis of child behavioral development (Behavior Problems Index, BPI). Income is after-tax income and it is measured in year 2000 dollars. Hours worked are yearly hours. Source: C-NLSY

Table 4.2: First Stage Estimates

	Combined Math-Reading		Behavior Problems Index	
	ΔIncome	ΔHours Worked	ΔIncome	ΔHours Worked
	(1)	(2)	(3)	(4)
ΔEITC	1.026**	1.481***	1.101**	1.488***
	(0.488)	(0.282)	(0.482)	(0.280)
LabDemShocks	1.659***	0.322*	2.067***	0.245
	(0.395)	(0.186)	(0.405)	(0.178)
SW Chi-sq. (Under id)	13.21	14.40	21.89	20.57
P-value	0.00	0.00	0.00	0.00
SW F (Weak id)	13.19	14.38	21.86	20.54
P-value	0.00	0.00	0.00	0.00
KP (Weak id)	6.42	6.42	10.43	10.43
Observations	12,288	12,288	13,777	13,777

Notes: This table shows the estimates for both our first stage models. Dependent variable: ΔIncome (columns 1 and 3), and ΔHours worked (columns 2 and 4). Columns (1) and (2) refer to the estimating sample for the analysis of child cognitive development (combined Math-Reading test score). Columns (3) and (4) consider the estimating sample for the analysis of child behavioral development (Behavior Problems Index, BPI). For each analysis, the two endogenous variables are: changes in income (ΔIncome) and changes in maternal hours worked (ΔHours). The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks ($LabDemShocks$). Income and the EITC are measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.3: Income, Hours Worked, and Child Test Scores

	Combined Math-Reading	
	OLS	IV
	(1)	(2)
ΔIncome	0.001*	0.044***
	(0.000)	(0.015)
ΔHours worked	0.000	-0.060**
	(0.001)	(0.024)
Observations	12,288	12,288

Notes: This table shows the estimates for our analysis of child cognitive development. Dependent variable: Combined Math-Reading test score. Column (1) reports the OLS estimates. Column (2) shows the IV estimates. The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks ($LabDemShocks$). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.4: Income, Hours Worked, and Child Behavior

	Behavior Problems Index	
	OLS	IV
	(1)	(2)
ΔIncome	0.000	0.013
	(0.000)	(0.009)
ΔHours worked	-0.001	-0.052**
	(0.001)	(0.022)
Observations	13,777	13,777

Notes: This table shows the estimates for our analysis of child behavioral development. Dependent variable: Behavior Problems Index (BPI). Column (1) reports the OLS estimates. Column (2) shows the IV estimates. The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks (*LabDemShocks*). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.5: The Effect of Family Income and Hours Worked in Isolation

	Combined Math-Reading			Behavior Problems Index		
	IV	IV	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)
ΔIncome	0.017**		0.044***	-0.003		0.013
	(0.007)		(0.015)	(0.007)		(0.009)
ΔHours worked		-0.021*	-0.060**		-0.040**	-0.052**
		(0.011)	(0.024)		(0.018)	(0.022)
First Stage Tests (Income/Hours):						
SW Chi-sq. (Under id)	19.37	29.19	13.21/14.40	27.68	29.09	21.89/20.57
P-value	0.00	0.00	0.00/0.00	0.00	0.00	0.00/0.00
SW F (Weak id)	9.67	14.57	13.19/14.38	13.82	14.53	21.86/20.54
P-value	0.00	0.00	0.00/0.00	0.00	0.00	0.00/0.00
KP (Weak id)	9.67	14.57	6.42	13.82	14.53	10.43
Observations	12,288	12,288	12,288	13,777	13,777	13,777

Notes: This table shows the estimates for our analysis of child cognitive development (columns 1–3) and child behavioral development (columns 4–6). Dependent variable: Combined Math-Reading test score (columns 1–3), and Behavior Problems Index (BPI) (columns 4–6). Columns (1) and (4) show the impact of family income in isolation. Columns (2) and (5) show the impact of maternal hours worked in isolation. Columns (3) and (6) show the contemporaneous impact of family income and maternal hours worked. All estimates are IV estimates. For comparison purposes, the coefficient for the effect of family income estimated in Dahl and Lochner (2017) is equal to 0.041. See their work for further details. In columns (1) to (6), the two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks (*LabDemShocks*). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models in columns (1) to (6) include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). The same models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.6: Heterogeneous Effect of EITC Changes: Mother's Employment

	Combined Math-Reading IV (1)	Behavior Problems Index IV (2)
ΔIncome	0.032***	0.002
	(0.009)	(0.008)
ΔHours worked	-0.039***	-0.028**
	(0.011)	(0.013)
First Stage Coefficients:		
ΔIncome:		
ΔEITC*Employed$_{(t-1)}$	1.557**	1.744***
	(0.640)	(0.638)
ΔEITC*Non-Employed$_{(t-1)}$	0.637	0.614
	(0.504)	(0.512)
LabDemShocks	1.613***	2.009***
	(0.395)	(0.404)
ΔHours worked:		
ΔEITC*Employed$_{(t-1)}$	0.544	0.553
	(0.365)	(0.355)
ΔEITC*Non-Employed$_{(t-1)}$	2.166***	2.194***
	(0.294)	(0.298)
LabDemShocks	0.403**	0.330*
	(0.189)	(0.181)
First Stage Tests (Income/Hours):		
SW Chi-sq. (Under id)	21.47/62.24	30.61/63.37
P-value	0.00/0.00	0.00/0.00
SW F (Weak id)	10.71/31.07	15.28/31.64
P-value	0.00/0.00	0.00/0.00
KP (Weak id)	7.06	10.05
Observations	12,288	13,777

Notes: This table shows the IV estimates for our robustness analysis. Dependent variable: Combined Math-Reading test score (column 1), and Behavior Problems Index (BPI) (column 2). First stage estimates are obtained by interacting the EITC instrument with the mother's lagged employment status. Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.7: Single Test Scores

	Math	Reading Recognition	Reading Comprehension
	IV	IV	IV
	(1)	(2)	(3)
ΔIncome	0.029**	0.055***	0.030**
	(0.012)	(0.018)	(0.013)
ΔHours worked	-0.036*	-0.070**	-0.049**
	(0.021)	(0.029)	(0.022)
Observations	12,288	12,288	12,288

Notes: This table shows the IV estimates for each single PIAT test score. Dependent variable: Math test score (column 1), Reading Recognition test score (column 2), and Reading Comprehension test score (column 3). The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks ($LabDemShocks$). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.8: Single Behavior Problems Index

	Antisocial	Anxious	Headstrong	Hyperactive	Peer Conflicts
	IV	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)
ΔIncome	0.012	-0.007	0.015	0.020**	0.009
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
ΔHours worked	-0.048**	-0.027	-0.046**	-0.036*	-0.041*
	(0.022)	(0.019)	(0.021)	(0.021)	(0.025)
Observations	13,777	13,777	13,777	13,777	13,777

Notes: This table shows the IV estimates for each single BPI score. Dependent variable: Antisocial behavior (column 1), Anxious behavior (column 2), Headstrong behavior (column 3), Hyperactive behavior (column 4), and Peer Conflicts behavior (column 5). The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks ($LabDemShocks$). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.9: Income, Hours Worked, and Early Childhood Development

	Compliance IV (1)	Insecure Attachment IV (2)	Compliance and Ins.Attach. IV (3)	Sociability IV (4)
ΔIncome	0.046	0.020	0.046	0.011
	(0.031)	(0.022)	(0.029)	(0.020)
ΔHours worked	-0.039	-0.044	-0.053	-0.010
	(0.043)	(0.034)	(0.039)	(0.045)
Age range	1–7	1–7	1–7	2–7
Observations	4,807	4,884	4,656	2,969

Notes: This table shows the IV estimates for our analysis of early childhood temperament development. Dependent variable: Compliance score (column 1), Insecure Attachment score (column 2), Combined Compliance and Insecure Attachment score (column 3), and Sociability score (column 4). The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks (*LabDemShocks*). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.10: Heterogeneous Effects: Mother Characteristics

	Combined Math-Reading IV (1)	Behavior Problems Index IV (2)
	Panel A: Mother's Education	
ΔIncome*HS or less	0.031**	0.012
	(0.015)	(0.010)
ΔIncome*Some college or more	0.030**	0.013
	(0.016)	(0.011)
ΔHours worked*HS or less	**-0.058****	**-0.054****
	(0.024)	(0.021)
ΔHours worked*Some college or more	**0.001**	**-0.049****
	(0.028)	(0.024)
Observations	12,288	13,777
	Panel B: Mother's AFQT	
ΔIncome*Low AFQT	0.030**	0.016
	(0.015)	(0.010)
ΔIncome*High AFQT	0.033**	0.018*
	(0.016)	(0.010)
ΔHours worked*Low AFQT	**-0.064****	**-0.052****
	(0.025)	(0.022)
ΔHours worked*High AFQT	**0.001**	**-0.073*****
	(0.028)	(0.023)
Observations	11,939	13,348
	Panel C: Mother's Marital Status	
ΔIncome*Married	0.038**	0.016
	(0.016)	(0.010)
ΔIncome*Unmarried	0.044***	0.013
	(0.017)	(0.011)
ΔHours worked*Married	**-0.010**	**-0.065****
	(0.030)	(0.029)
ΔHours worked*Unmarried	**-0.069****	**-0.052****
	(0.028)	(0.022)
Observations	12,288	13,777

Notes: This table shows the IV heterogeneous effects of income and maternal hours worked on child development. Dependent variable: Combined Math-Reading test score (column 1), and Behavior Problems Index (BPI) (column 2). We divide mothers according to: (i) Panel A: educational attainments (high school (HS) diploma or less vs. some college or more); (ii) Panel B: AFTQ score (below or above the median); and (iii) Panel C: marital status (married vs. unmarried). The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks ($LabDemShocks$). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are obtained through a family-level clustered bootstrap procedure based on 100 repetitions and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.11: Heterogeneous Effects: Child Characteristics

	Combined Math-Reading IV (1)	Behavior Problems Index IV (2)
	Panel A: Child's Race	
ΔIncome*Black or Hispanic	0.046**	0.014
	(0.018)	(0.009)
ΔIncome*White	0.047**	0.015
	(0.019)	(0.010)
ΔHours worked*Black or Hispanic	**-0.069****	**-0.050****
	(0.031)	(0.023)
ΔHours worked*White	**-0.047**	**-0.068*****
	(0.032)	(0.022)
Observations	12,288	13,777
	Panel B: Child's Age	
ΔIncome*Below 12	0.048**	0.012
	(0.019)	(0.009)
ΔIncome*Above 12	0.049**	0.015
	(0.020)	(0.010)
ΔHours worked*Below 12	**-0.076****	**-0.055****
	(0.031)	(0.023)
ΔHours worked*Above 12	**-0.053**	**-0.055****
	(0.033)	(0.022)
Observations	12,288	13,777

Notes: This table shows the IV heterogeneous effects of income and maternal hours worked on child development. Dependent variable: Combined Math-Reading test score (column 1), and Behavior Problems Index (BPI) (column 2). We divide children according to: (i) Panel A: race (white vs. black or Hispanic); and (ii) Panel B: age (below 12 years old vs. above 12 years old). The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks ($LabDemShocks$). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are obtained through a family-level clustered bootstrap procedure based on 100 repetitions and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table 4.12: Maternal Employment Status, Investment in the Child, and Child's Activities

	(LI,NE) (1)	(LI,E) (2)	(HI,NE) (3)	(HI,E) (4)	(LI,E)-(LI,NE) (5)	p-val (6)	(HI,E)-(HI,NE) (7)	p-val (8)
	Panel A: Parenting							
Encourage hobbies	0.92	0.91	0.96	0.94	-0.01	0.64	-0.01	0.38
Phys. affection (times past week)	8.43	9.51	15.55	13.98	1.07	0.16	-1.57	0.36
Parenting class pre-birth	0.15	0.14	0.20	0.18	-0.02	0.36	-0.03	0.18
Parenting class	0.24	0.20	0.31	0.26	-0.04	0.06	-0.05	0.03
Never cared by others	0.57	0.24	0.45	0.15	-0.33	0.00	-0.30	0.00
Use of rules	0.58	0.51	0.54	0.50	-0.07	0.02	-0.04	0.19
Control who the child is with	0.55	0.58	0.59	0.47	0.03	0.32	-0.11	0.00
Control activities after school	0.60	0.66	0.70	0.57	0.06	0.08	-0.13	0.00
Set homework time	0.70	0.78	0.82	0.72	0.08	0.01	-0.10	0.00
	Panel B: Reaction to Poor Scholastic Performance							
Contact faculty ($\geq$ 6 y.o.)	0.84	0.82	0.81	0.88	-0.02	0.47	0.07	0.01
Closer eye on activities	0.84	0.84	0.89	0.88	0.00	0.84	-0.01	0.80
Lecture a child	0.80	0.81	0.74	0.80	0.01	0.80	0.06	0.04
Tell child to work harder	0.81	0.80	0.66	0.73	-0.01	0.84	0.07	0.04
Help with schoolwork	0.80	0.82	0.75	0.76	0.02	0.39	0.01	0.78
	Panel C: Family Environment Scales							
Full home	17.39	18.10	19.90	20.18	0.71	0.00	0.28	0.13
Cognitive stimulation	8.67	9.24	10.04	10.13	0.57	0.00	0.09	0.44
Emotional support	8.72	8.86	9.86	10.05	0.15	0.14	0.19	0.09
Parental warmth	4.46	4.47	4.59	4.48	0.01	0.67	-0.11	0.00
	Panel D: Time Diaries (in seconds per day)							
School	12,161	16,323	12,745	16,743	4,162	0.00	3,998	0.00
TV	6,492	6,271	5,769	5,247	-221	0.49	-522	0.12
Electronic games	365	538	335	361	172	0.10	27	0.73
Art, sculpture	242	123	244	214	-119	0.01	-30	0.56
Books	248	238	350	337	-10	0.83	-13	0.81
Books ($\geq$ 4 y.o.)	280	248	332	334	-32	0.59	2	0.97
Visiting others, socializing	409	526	261	288	117	0.40	28	0.76

Notes: This table shows several measures for investment in the child development process using the CDS supplement of the PSID data set. All measures refer to children aged 0–12 in 1997. LI means low family income (below $35,000), HI means high family income (above $35,000). NE means that the mother is non-employed in 1997, E means that the mother is employed in 1997. All the variables (if not differently specified) excepted time diaries are indicator variables. Time diaries variables (Panel D) are expressed in seconds per day and refer to weekdays only.

Chapter 5

CONCLUSIONS

Chapter 2 develops new identification concepts and associated estimators for the process of skill development in children. One of the key empirical challenges in this context is that the various measures of children's skills are in general imperfect and arbitrarily located and scaled. We introduce the concept of known location and scale production technologies, which are the type of technologies actually estimated in many previous papers, and show that for these technologies, standard measurement assumptions non-parametrically identify the production technology, up to the normalization of initial period skills. Importantly, we show non-parametric identification for these cases without re-normalizing latent skills each period which can bias the production technology. For production functions which do not have a known location or scale, additional assumptions are necessary, and we provide empirically grounded assumptions which are sufficient for identification of these more general technologies. Our paper provides the first analysis of these crucial identification tradeoffs, and hopefully will serve as a useful guide for future work.

Based on our identification results, we develop a robust method of moments estimator and show that it can be implemented using a sequential algorithm. Our estimator does not require strong assumptions about the marginal distribution of measurement errors or the latent factors. We estimate the skill production process using data for the United States and a flexible parametric model of skill development allowing for non-constant returns to scale, dynamics in TFP, and for parental investment to endogenously depend on unobserved children's skills.

Our empirical results show a pattern of rapid skill development from age 5 to 14.

We find that as children age, not only does their mean skill level increase, but the level of skill inequality also increases. Our parameter estimates reveal that investments are more productive at early ages and in particular for disadvantaged children. Our findings of a positive return to income transfers at early ages, especially for poorer households, is largely consistent with prior evidence of a positive effect of income on a number of child outcomes (see Dahl and Lochner, 2012a; Loken *et al.*, 2012) using different sources of identification. Our results suggest that family income is a better "target" than initial children's skills for children's skills. Lastly, our finding that that the estimated policy effects would be substantially smaller if one estimated a restricted technology or ignored measurement error demonstrates the critical importance of allowing for general technologies and correcting estimates for measurement error.

Chapter 3 studies the role of children's social interactions in the dynamics of children's skills. I estimate a tractable dynamic equilibrium model of parental investment and endogenous formation of peer groups. The model is estimated using information about friendships, children's test scores and parental investments in the National Longitudinal Study of Adolescent Health (Add Health). I exploit within school / across cohort variations in peers' composition to identify the degree of complementarity between parents and peers in producing a child's skills. I find that parents and peers are *static* substitutes and *dynamic* complementary inputs in child development. After validating my estimated model using findings in Chetty and Hendren (2016a) on environment exposure effects in children, I assess the importance of social interactions in skills dynamics with various policies.

This article underlines three main points: (i) social interactions and social context permanently shape the developmental trajectories of children; (ii) changing cohort composition and the relative social interactions generates *winners* and *losers* and the heterogeneous effects are due to the endogenous formation of new peers; (iii) neglecting the dynamic equilibrium effects of skill formation and social interactions would lead to biased predicted effects of policies.

I want to conclude this paper by considering an extension of this work. Specifically, one potential type of parental investment can be the choice of neighborhood where the family lives. In this case, parents have alternative margins in response to changes in peer composition, and to a certain extent, they can also decide to change where they live as a response to the previously considered policy. Modeling this second channel is challenging, because now the environment composition is also endogenous, and it becomes part of the equilibrium solution of the model. However, understanding the extent to which neighborhood decisions are influenced by children's social interactions is an important question in considering the effects of socioeconomic segregation on intergenerational mobility. Therefore, future work is needed.

Chapter 4 unveils the contemporaneous effect family income and maternal hours worked have in shaping child development. We combine the analysis of cognitive and noncognitive development. We exploit children's performance on standardized tests to measure cognitive development. We use indicators of behavioral problems to gauge noncognitive development.

We find that family income has a sizable and positive effect on cognitive development, while the income effect is negligible (although positive) on behavioral development. The effect of maternal hours worked is the same across outcomes. On average, hours worked by the mother negatively affect both cognitive and behavioral development.

We shed light on the mechanism behind the negative effect of maternal hours worked on child development. Working mothers invest less time in child care. As a consequence, the choice of alternative sources of child care becomes crucial; this choice is likely to be affected by economic factors. We decompose the overall effect of maternal hours worked on child development into an income effect (higher earnings) and a substitution effect (less maternal time). We find that the substitution effect tends to dominate the income effect when the after-tax hourly wage is less than $13.50 per hour. With higher earnings, families are able to substitute their decreased time investment with better and more productive alternatives. In line with this explanation, we show that the average effect (on cognitive development) is mainly driven by low-income, less-educated families and that the employment effect on investment in the child differs according to family income.

Several policy suggestions derive from our results. The trade-off between the income and substitution effect in terms of child development encourages a debate about the effect of conditional versus unconditional cash transfers. Income subsidies that provide monetary transfers based on work requirements might produce heterogeneous impacts in terms of child development. Our analysis confirms that policies aimed at fostering maternal labor supply benefit child development when considered in conjunction with well-researched policies concerning the optimal level of family income taxation or the optimal minimum wage. Alternatively, policies that encourage maternal employment in low-income families should also guarantee alternative sources of child care to support child development.

REFERENCES

Abdulkadiroglu, A., J. D. Angrist and P. A. Pathak, “The elite illusion: Achievement effects at Boston and New York exam schools”, Econometrica **82**, 137–196 (2014).

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

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

Aizer, A., “The Gender Wage Gap and Domestic Violence”, American Economic Review **100**, 4, 1847–1859 (2010).

Ammermueller, A. and J.-S. Pischke, “Peer effects in european primary schools: Evidence from the progress in international reading literacy study”, Journal of Labor Economics **27**, 3, 315–348 (2009).

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

Andrews, D. W. K. and G. Soares, “Inference for parameters defined by moment inequalities using generalized moment selection”, Econometrica **78**, 1, 119–157 (2010).

Angrist, J. D. and K. Lang, “Does school integration generate peer effects? Evidence from Boston’s METCO program”, American Economic Review **94**, 5, 1613–1634 (2004).

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

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir and M. Rubio-Codina, “Estimating the production function for human capital: Results from a randomized control trial in colombia”, NBER Working Paper , 21740 (2015a).

Attanasio, O., S. Cattan, E. F. C. Meghir and M. Rubio−Codina, “Estimating the production function for human capital: Results from a randomized controlled trial in Colombia”, NBER Working Paper n. 20965 (2017a).

Attanasio, O., C. Meghir and E. Nix, “Human capital development and parental investment in india”, NBER Working Paper , 20965 (2015b).

Attanasio, O., C. Meghir and E. Nix, “Human capital development and parental investment in India”, NBER Working Paper n. 21740 (2017b).

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

Auerbach, E., “Identification and estimation of models with endogenous network formation”, (2016).

Autor, D. H. and M. G. Duggan, “The Rise in the Disability Rolls and the Decline in Unemployment”, The Quarterly Journal of Economics **118**, 1, 157–206 (2003).

Badev, A., “Discrete games in endogenous networks: Theory and policy”, Working Paper (2016).

Bala, V. and S. Goyal, “A noncooperative model of network formation”, Econometrica **68**, 5, 1181–1229 (2000).

Bartik, T. J., *Who Benefits from State and Local Economic Development Policies?* (Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research, 1991).

Baum, C. L., “Does Early Maternal Employment Harm Child Development? An Analysis of the Potential Benefits of Leave Taking”, Journal of Labor Economics **21**, 2, 409–448 (2003).

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

Bernal, R. and M. P. Keane, “Child Care Choices and Children’s Cognitive Achievement: The Case of Single Mothers”, Journal of Labor Economics **29**, 3, 459–512 (2011).

Bertrand, M., E. Kamenica and J. Pan, “Gender Identity and Relative Income within Households”, The Quarterly Journal of Economics **130**, 2, 571–614 (2015).

Bifulco, R., J. M. Fletcher and S. L. Ross, “The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health”, American Economic Journal: Economic Policy **3**, 1, 25–53 (2011).

Blanchard, O. and L. F. Katz, “Regional Evolutions”, Brookings Papers on Economic Activity **23**, 1, 1–76 (1992).

Blau, D. M., “The Effect of Income on Child Development”, The Review of Economics and Statistics **81**, 2, 261–276 (1999).

Blume, L. E., W. A. Brock, S. N. Durlauf and Y. M. Ioannides, “Identification of social interactions”, Handbook of Social Economics, J. Benhabib, A. Bisin, and M. Jackson **Amsterdam: North Holland** (2011).

Blume, L. E., W. A. Brock, S. N. Durlauf and R. Jayaraman, “Linear social interactions models”, Journal of Political Economy **123**, 2, 444–496 (2015).

Blundell, R., M. Costa Dias, C. Meghir and J. Shaw, “Female Labor Supply, Human Capital, and Welfare Reform”, Econometrica **84**, 5, 1705–1753, URL `http://dx.doi.org/10.3982/ECTA11576` (2016).

Blundell, R. and H. Hoynes, “Has In-Work Benefit Reform Helped the Labour Market?”, in “Seeking a Premier League Economy”, edited by R. Blundell, D. Card and R. Freeman (University of Chicago Press, 2004).

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**, 1468–1479 (2013b).

Bramoullé, Y., H. Djebbari and B. Fortin, “Identification of peer effects through social networks”, Journal of Econometrics **150**, 1, 41 – 55 (2009).

Brock, W. A. and S. N. Durlauf, “Discrete choice with social interactions”, Review of Economic Studies **68**, 2, 235–260 (2001a).

Brock, W. A. and S. N. Durlauf, “Interactions-based models”, Handbook of Econometrics by J. Heckman and E. Leamer **5**, Amsterdam: North-Holland, 52–75 (2001b).

Brock, W. A. and S. N. Durlauf, “Linear social interactions models”, Journal of Econometrics **140**, 52–75 (2007).

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

Calvó-Armengol, A., E. Patacchini and Y. Zenou, “Peer effects and social networks in education”, Review of Economic Studies **76**, 4, 1239–1267 (2009).

Card, D. and L. Giuliano, “Can tracking raise the test scores of high-ability minority students?”, American Economic Review **106**, 10, 2783–2816 (2016).

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).

Carneiro, P., K. V. Løken and K. G. Salvanes, “A Flying Start? Maternity Leave Benefits and Long-Run Outcomes of Children”, Journal of Political Economy **123**, 2, 365–412 (2015).

Carneiro, P. and M. Rodriguez, “Evaluating the Effect of Maternal Time on Child Development Using the Generalized Propensity Score. Working Paper”, (2009).

Carrell, S. E., M. Hoekstra and E. Kuka, “The long-run effects of disruptive peers”, NBER Working Paper n. 22042 (2016).

Carrell, S. E., B. I. Sacerdote and J. E. West, “From natural variation to optimal policy? The importance of endogenous peer group formation”, Econometrica **81**, 3 (2013).

Caucutt, E. M. and L. Lochner, "Early and late human capital investments, borrowing constraints, and the family", (2017).

Caucutt, E. M., L. Lochner and Y. Park, "Correlation, consumption, confusion, or constraints: Why do poor children perform so poorly?", NBER Working Papers , 21023 (2015).

Center on Budget and Policy Priorities, *Policy Basics: The Earned Income Tax Credit* (Washington DC: Center on Budget and Policy Priorities, 2016).

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).

Chandrasekhar, A. G. and R. Lewis, "Econometrics of sampled networks", (2016).

Charles, K., E. Hurst and M. Notowidigdo, "Housing Booms and Busts, Labor Market Opportunities, and College Attendance. NBER Working Paper n. 21587", (2015).

Charles, K., E. Hurst and M. Notowidigdo, "Housing Booms, Manufacturing Decline, and Labor Market Outcomes", The Economic Journal , forthcoming (2017).

Chernozhukov, V., H. Hong and E. Tamer, "Estimation and confidence regions for parameter sets in econometric models", Econometrica **75**, 5, 1243–1284 (2007).

Chetty, R. and N. Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects", (2016a).

Chetty, R. and N. Hendren, "The impacts of neighborhoods on intergenerational mobility II: County-level estimates", (2016b).

Chetty, R., N. Hendren and L. Katz, "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity project", American Economic Review **106**, 4 (2016a).

Chetty, R., N. Hendren, F. Lin, J. Majerovitz and B. Scuderi, "Childhood environment and gender gaps in adulthood", American Economic Review Papers and Proceedings (2016b).

Christakis, N. A. and J. H. Fowler, "Connected: The surprising power of our social networks and how they shape our lives", Little, Brown and Company (2009).

Christakis, N. A., J. H. Fowler, G. W. Imbens and K. Kalyanaraman, “An empirical model for strategic network formation”, NBER Working Paper n. 16039 (2010).

Coleman, J. S., “Relational analysis: the study of social organizations with survey methods”, Hum Organ **17**, 28–36 (1958).

Cunha, F., “Investments in children when markets are incomplete”, Working Paper (2013a).

Cunha, F., “Investments in children when markets are incomplete”, (2013b).

Cunha, F., I. Elo and J. Culhane, “Eliciting maternal expectations about the technology of cognitive skill formation”, NBER working paper , 19144 (2013).

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 for the formation of skills”, 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).

Currarini, S., M. O. Jackson and P. Pin, “Identifying the roles of race-based choice and chance in high school friendship network formation”, Proceedings of the National Academy of Sciences **107**, 11 (2010).

Dahl, G. B. and L. Lochner, “The impact of family income on child achievement: Evidence from the earned income tax credit”, American Economic Review **102**, 5, 1927–1956 (2012a).

Dahl, G. B. and L. Lochner, “The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit”, American Economic Review **102**, 5, 1927–1956 (2012b).

Dahl, G. B. and L. Lochner, “The Impact of Family Income on Child Achievement: Evidence from Changes in the Earned Income Tax Credit: Reply”, American Economic Review **107**, 2, 629–631 (2017).

Datta, M., K. Reffett and L. Wozny, “Comparing recursive equilibria in economies with dynamic complementarities and indeterminacy”, Economic Theory **Forthcoming** (2017).

De Giorgi, G., M. Pellizzari and S. Redaelli, “Identification of social interactions through partially overlapping peer groups”, American Economics Journal: Applied Economics **2**, 2, 241–275 (2010).

Del Boca, D., C. Flinn and M. Wiswall, “Household Choices and Child Development”, The Review of Economic Studies **81**, 1, 137–185 (2014a).

Del Boca, D., C. Flinn and M. Wiswall, "Household choices and child development", Review of Economic Studies **81**, 1, 137–185 (2014b).

Del Boca, D., C. Flinn and M. Wiswall, "Transfers to households with children and child development", Economic Journal **Forthcoming** (2016).

Del Bono, E., M. Francesconi, Y. Kelly and A. Sacker, "Early Maternal Time Investment and Early Child Outcomes", The Economic Journal **126**, 596, F96–F135, URL `http://dx.doi.org/10.1111/ecoj.12342` (2016).

Diamond, R., "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000", American Economic Review **106**, 3, 479–524 (2016).

Duncan, G., P. Morris and C. Rodrigues, "Does money really matter? estimating impacts of family income on young children's achievement with data from random-assignment experiments", Dev Psychol **47**, 5, 1263–1279 (2011).

Dutta, B., S. Ghosal and D. Ray, "Farsighted network formation", Journal of Economic Theory **122**, 2, 143–164 (2005).

Eissa, N. and J. B. Liebman, "Labor Supply Response to the Earned Income Tax Credit", The Quarterly Journal of Economics **111**, 2, 605–637 (1996).

Feenberg, D. and E. Coutts, "An Introduction to the TAXSIM Model", Journal of Policy Analysis and Management **12**, 1, 189–194 (1993).

Fort, M., A. Ichino and G. Zanella, "The Cognitive Cost of Daycare 0-2 for Children in Advantaged Families. Working Paper", (2017).

Fox, L., W. J. Han, C. Ruhm and J. Waldfogel, "Time for Children: Trends in the Employment Patterns of Parents, 1967–2009", Demography **50**, 1, 25–49 (2013).

Fu, C. and J. Gregory, "Estimation of an equilibrium model with externalities: Combining the strengths of structural models and quasi experiments", (2017).

Fu, C. and N. Mehat, "Ability tracking, school and parental effort, and student achievement: A structural model and estimation", Working Paper (2016).

Gayle, G.-L., L. Golan and A. M. Soytas, "What accounts for the racial gap in time allocation and intergenerational transmission of human capital?", (2015).

Gayle, G.-L., L. Golan and A. M. Soytas, "What is the source of the intergenerational correlation in earnings?", (2016).

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

Goldsmith-Pinkham, P. and G. W. Imbens, "Social Networks and the Identification of Peer Effects", Journal of Business & Economic Statistics **31**, 3, 253–264 (2013).

Goldsmith-Pinkham, P., I. Sorkin and H. Swift, “Bartik Instruments: What, When, Why, and How. Working Paper”, (2017).

Graham, B. S., “Homophily and transitivity in dynamic network formation”, (2016).

Graham, B. S., “An econometric model of network formation with degree heterogeneity”, Econometrica **85**, 4, 1033–1063 (2017).

Griffith, A., “How many friends do you have? An empirical investigation into censoring−induced bias in social network data”, (2017).

Guryan, J., E. Hurst and M. Kearney, “Parental Education and Parental Time with Children”, Journal of Economic Perspectives **22**, 3, 23–46 (2008).

Halpern, C. T., “Smarter teens delay sexual contacts”, PPFY (Pregnancy Prevention for Youth) Network Newsletter **3**, 4–5 (2000).

Hanushek, E. A., J. F. Kain, J. M. Markman and S. G. Rivkin, “Does peer ability affect student achievement?”, Journal of Applied Econometrics **18**, 5, 527–544 (2003).

Heckman, J. and S. Mosso, “The Economics of Human Development and Social Mobility”, Annual Review of Economics **6**, 689–733 (2014a).

Heckman, J. J., “Policies to Foster Human Capital”, Research in Economics **54**, 1, 3–56, URL `http://www.sciencedirect.com/science/article/pii/S1090944399902259` (2000).

Heckman, J. J. and D. V. Masterov, “The productivity argument for investing in young children”, Review of Agricultural Economics **29**, 3, 446–493 (2007).

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

Heckman, J. J. and Y. Rubinstein, “The Importance of Noncognitive Skills: Lessons from the GED Testing Program”, American Economic Review **91**, 2, 145–149, URL `http://www.aeaweb.org/articles?id=10.1257/aer.91.2.145` (2001).

Hotz, V. J. and J. K. Scholz, “The Earned Income Tax Credit”, in “Means-Tested Transfer Programs in the United States, Volume I”, edited by R. A. Moffitt (University of Chicago Press, 2003).

Hoxby, C., “Peer effects in the classroom: Learning from gender and race variation”, NBER Working Paper n. 7867 (2000).

Hoynes, H. and N. Essa, “The Hours of Work Response of Married Couples: Taxes and the Earned Income Tax Credit”, in “Tax Policy and Labor Market Performance”, edited by J. Agell and P. B. Sorensen (MIT Press, 1996).

Hsin, A. and C. Felfe, “When Does Time Matter? Maternal Employment, Children’s Time With Parents, and Child Development”, Demography **51**, 5, 1867–1894 (2014).

Imbens, G. W. and J. D. Angrist, “Identification and Estimation of Local Average Treatment Effects”, Econometrica **62**, 2, 467–475 (1994).

Imberman, S. A., A. D. Kugler and B. I. Sacerdote, “Katrina’s children: Evidence on the structure of peer effects from hurricane evacuees”, American Economic Review **102**, 5, 2048–2082 (2012).

Jackson, M. O., *Social and Economic Networks* (Princeton University Press, 2008).

Jackson, M. O. and A. Wolinsky, “A strategic model of social and economic networks”, Journal of Economic Theory **71**, 1, 44–74 (1996).

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).

Lavy, V., M. D. Paserman and A. Schlosser, “Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom”, The Economic Journal **122**, 559, 208–237 (2012).

Lavy, V. and A. Schlosser, “Mechanisms and impacts of gender peer effects at school”, American Economic Journal: Applied Economics **3**, 2, 1–33 (2011).

Lazzati, N., “Treatment response with social interactions: Partial identification via monotone comparative statics”, Quantitative Economics , 6, 49–83 (2015).

Levy, D. and G. Duncan, “Using Sibling Samples to Assess the Effect of Childhood Family Income on Completed Schooling. Working Paper”, (1999).

Loken, K. V., M. Mogstad and M. Wiswall, “What linear estimators miss: Re-examining the effects of family income on child outcomes”, American Economic Journal: Applied Economics **4**, 2, 1–35 (2012).

Løken, K. V., M. Mogstad and M. Wiswall, “What Linear Estimators Miss: The Effects of Family Income on Child Outcomes”, American Economic Journal: Applied Economics **4**, 2, 1–35 (2012).

Lundstrom, S., “The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit: Comment”, American Economic Review **107**, 2, 623–628, URL http://www.aeaweb.org/articles?id=10.1257/aer.20150988 (2017).

Luttmer, E. F. P., “Neighbors as Negatives: Relative Earnings and Well-Being”, The Quarterly Journal of Economics **120**, 3, 963–1002 (2005).

Manski, C. F., “Identification of endogenous social effects: The reflection problem”, Review of Economic Studies **60**, 3, 531–542 (1993).

Mele, A., “A structural model of segregation in social networks”, Working Paper **Net Institute**, 10-16 (2010).

Meyer, B. D., "Labor Supply at the Extensive and Intensive Margins: The EITC, Welfare, and Hours Worked", American Economic Review **92**, 2, 373–379, URL http://www.aeaweb.org/articles?id=10.1257/000282802320191642 (2002).

Mincer, J., "Investment in human capital and personal income distribution", Journal of Political Economy **66**, 4, 281–302 (1958).

Mullins, J., "Improving child outcomes through welfare reform", (2016).

Nichols, A. and J. Rothstein, "The Earned Income Tax Credit", in "Economics of Means-Tested Transfer Programs in the United States, Volume I", edited by R. A. Moffitt (University of Chicago Press, 2016).

Notowidigdo, M., "The Incidence of Local Labor Demand Shocks. NBER Working Paper n. 17167", (2011).

Olivetti, C., E. Patacchini and Y. Zenou, "Mothers, peers and gender identity", Working Paper (2016).

Pakes, A., J. Porter, K. Ho and J. Ishii, "Moment inequalities and their application", Econometrica **83**, 1, 315–334 (2015).

Patacchini, E., E. Rainone and Y. Zenou, "Student networks and long-run educational outcomes: The strength of strong ties", CEPR Discussion Papers 9149, C.E.P.R. Discussion Papers (2012).

Patacchini, E. and Y. Zenou, "Juvenile delinquency and conformism", Journal of Law, Economics and Organization **28**, 1, 1–31 (2012).

Patacchini, E. and Y. Zenou, "Social networks and parental behavior in the intergenerational transmission of religion", Quantitative Economics **7**, 969–995 (2016).

Pavan, R., "On the production of skills and the birth order effect", Working Paper (2015).

Peterson, J. and N. Zill, "Marital Disruption, Parent-Child Relationships, and Behaviour Problems in Children", Journal of Marriage and the Family **48**, 295–307 (1986).

Ruhm, C. J., "Parental Employment and Child Cognitive Development", Journal of Human Resources **39**, 1, 155–192 (2004).

Sacerdote, B., "Peer effects with random assignment: Results for Dartmouth roommates", The Quarterly Journal of Economics **116**, 681–704 (2001).

Sacerdote, B., "Experimental and quasi-experimental analysis of peer effects: Two steps forward?", Annual Review of Economics **6**, 253–272 (2014).

Sayer, L. C., S. M. Bianchi and J. P. Robinson, "Are Parents Investing Less in Children? Trends in Mothers' and Fathers' Time with Children", American Journal of Sociology **110**, 1, 1–43 (2004).

Sheng, S., "A structural econometric analysis of network formation games", Working Paper (2014).

Todd, P. and K. Wolpin, "On the specification and estimation of the production function for cognitive achievement", Economic Journal **113**, 3, F3–F33 (2003).

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).

Topkis, D. M., "Equilibrium points in nonzero-sum n-person submodular games", SIAM Journal of Control and Optimization **17**, 6, 773–787 (1979).

Chapter A

SUPPLEMENTARY MATERIAL FOR CHAPTER 1

A.1 Proofs

A.1.1 Proof of Lemma 1

Proof. First, we note that with $Z_{t,m}$, $G_t(\theta_t, I_t)$, $\mu_{t,m}$ and $\lambda_{t,m}$ known, we then identify the distribution of the measurement error, given

$$\varphi_{\frac{\epsilon_{t,m}}{\lambda_{t,m}}}(x) = \frac{\varphi_{\widetilde{Z}_{t,m}}(x)}{\varphi_{\ln\theta_{t,m}}(x)}$$

Given the one-to-one mapping between characteristic functions and distributions, we identify the marginal density of $\frac{\epsilon_{t,m}}{\lambda_{t,m}}$. Since $\lambda_{t,m}$ is known, we also identify the marginal density of $\epsilon_{t,m}$, $F_{\epsilon_{t,m}}(\epsilon)$.

Next, consider the following conditional expectation:

$$E(Z_{t+1,m}|\ln\theta_t = a, \ln I_t = \ell) = \mu_{t+1} + \lambda_{t,m}E(\ln\theta_{t+1}|\ln\theta_t = a, \ln I_t = \ell)$$
$$+E(\epsilon_{t+1,m}|\ln\theta_t = a, \ln I_t = \ell)$$

where $E(\epsilon_{t+1,m}|\ln\theta_t = a, \ln I_t = \ell) = 0$ given Assumption 1 ($\epsilon_{t+1,m}$ independent of $\ln\theta_t$ and $\ln I_t$).

Iterating expectations and substituting for $\ln\theta_t = \frac{Z_{t,m}-\mu_{t,m}-\epsilon_{t,m}}{\lambda_{t,m}}$, we have the following:

$$E(Z_{t+1,m}|\ln\theta_t = a, \ln I_t = \ell) = \int E(Z_{t+1,m}|\frac{Z_{t,m}-\mu_{t,m}-\epsilon}{\lambda_{t,m}} = a, \ln I_t = \ell, \epsilon)dF_{\epsilon_{t,m}}(\epsilon)$$

Again applying Assumption 1 ($\epsilon_{t,m}$ independent of $Z_{t+1,m}$), we have

$$= \int E(Z_{t+1,m}|\frac{Z_{t,m}-\mu_{t,m}-\epsilon}{\lambda_{t,m}} = a, \ln I_t = \ell)dF_{\epsilon_{t,m}}(\epsilon)$$

Re-writing again, we have

$$= \int E(Z_{t+1,m}|Z_{t,m} = \lambda_{t,m}a + \mu_{t,m} + \epsilon, \ln I_t = \ell)dF_{\epsilon_{t,m}}(\epsilon)$$

$$= \int E(Z_{t+1,m}|Z_{t,m} = b(\epsilon), \ln I_t = \ell)dF_{\epsilon_{t,m}}(\epsilon)$$

Note that for each realization of $\epsilon_{t,m} = \epsilon$, we have $Z_{t,m} = b(\epsilon)$, where $b(\epsilon)$ is known given $\mu_{t,m}, \lambda_{t,m}$, and a are known. We identify the conditional expectation $E(Z_{t+1,m}|Z_{t,m} = b(\epsilon), \ln I_t = \ell)$ from the observed distribution of $Z_{t+1,m}$ and $Z_{t,m}$ measures. Because the distribution of measurement errors $F_{\epsilon_{t,m}}(\epsilon)$ is identified, we identify $E(Z_{t+1,m}|\ln\theta_t = a, \ln I_t = \ell)$.

■

Example 4 *Consider the case where* $\epsilon_{t,m} \sim N(0, \sigma^2_{t,m})$ $\forall t$. *We identify* $\sigma^2_{t,m}$ *from* $V(Z_{t,m}) = \lambda^2_{t,m} V(\ln\theta_t) + V(\epsilon_{t,m})$ *since we have already identified* $V(\ln\theta_t)$ *and* $\lambda_{t,m}$. *The idea of the proof of Lemma 1 is that the value of the current latent skills* ($\ln\theta_t = a$) *comes both from observable measure* ($Z_{t,m}$) *and unobservable measurement error* ($\epsilon_{t,m}$). *Since we identify the distribution of the unobservable, we are able to integrate out each possible realization of that unobservable random variable. Indeed, if* ϵ *takes value 0, because we are fixing* $\ln\theta_t$ *to be equal to* a, *this implies that* $Z_{t,m}$ *would equal:*

$$Z_{t,m} = \lambda_{t,m} \cdot a + \mu_{t,m} = b(0)$$

where both $\lambda_{t,m}$ *and* $\mu_{t,m}$ *are known. Hence weight* $E(Z_{t+1,m}|Z_{t,m} = b(0), \ln I_t = \ell)$ *with the likelihood of the event that* ϵ *takes the value of zero. Because* $\epsilon_{t,m} \sim N(0, \sigma^2_{t,m})$, *we have that the marginal density of the measurement error is*

$$f_{\epsilon_{t,m}}(\epsilon) = \frac{1}{\sigma_{t,m}\sqrt{2\pi}} e^{-\frac{\epsilon^2}{2\sigma^2_{t,m}}}$$

and

$$\int E(Z_{t+1,m}|Z_{t,m} = b(\epsilon), \ln I_t = \ell)\, f_{\epsilon_{t,m}}(\epsilon) d\,\epsilon$$

Because $\epsilon_{t,m}$ *is a continuous random variable, we integrate over all the values to find* $E(Z_{t+1,m}|\ln\theta_t = a, \ln I_t = \ell)$. *This approach would be similar in the case where investment is also a latent variable. In this case, we would integrate over the support of the measurement error terms of both variables. by* $F_{\epsilon_{t,m}}(\epsilon) = pr(\epsilon_{t,m} \leq \epsilon)$. *Define* $\widetilde{Z}_{t,m} = \ln\theta_t + \frac{\epsilon_{t,m}}{\lambda_{t,m}}$ *and its characteristic function* $\varphi_{\widetilde{Z}_{t,m}}(x) = E\left[e^{ix\left(\frac{Z_{t,m}-\mu_{t,m}}{\lambda_{t,m}}\right)}\right]$. *Define* $\varphi_{\ln\theta_t}(x) = E\left[e^{ix\ln\theta_t}\right]$ *to be the characteristic function of* $\ln\theta_t$. *Given the independence between* $\epsilon_{t,m}$ *and* $\ln\theta_t$ *(Assumption 1), we can rewrite the characteristic function for* $\frac{\epsilon_{t,m}}{\lambda_{t,m}}$ *to be:*

A.1.2 Proof of Theorem 1

Proof.

Given $G_t(\theta_t, I_t)$ and the measurement parameters for period t, $\mu_{t,m}$ and $\lambda_{t,m}$, are known, we use Lemma 1 to identify $E(\ln \theta_{t+1} | \ln \theta_t = a, \ln I_t = \ell)$ from $E(Z_{t+1,m} | \ln \theta_t = a, \ln I_t = \ell)$ for any $a \in \mathbb{R}$ and $\ell \in \mathbb{R}$. We then use the following transformation:

$$\frac{E(Z_{t+1,m}| \ln \theta = a_1, \ln I_t = \ell_1) - E(Z_{t+1,m}| \ln \theta = a_2, \ln I_t = \ell_2)}{E(Z_{t+1,m}| \ln \theta = a_3, \ln I_t = \ell_3) - E(Z_{t+1,m}| \ln \theta = a_2, \ln I_t = \ell_2)} = \frac{\ln f_t(e^{a_1}, e^{\ell_1}) - \ln f_t(e^{a_2}, e^{\ell_2})}{\ln f_t(e^{a_3}, e^{\ell_3}) - \ln f_t(e^{a_2}, e^{\ell_2})}$$

Because the function f_t satisfies the known location and scale definition, then for the points (a_2, ℓ_2) and (a_3, ℓ_3) the function evaluated at those points, $f_t(e^{a_2}, e^{\ell_2})$ and $f_t(e^{a_3}, e^{\ell_3})$, where $f_t(e^{a_2}, e^{\ell_2}) \neq f_t(e^{a_3}, e^{\ell_3})$, is known. Call these known points, $f_t(e^{a_2}, e^{\ell_2}) = \alpha_2$ and $f_t(e^{a_3}, e^{\ell_3}) = \alpha_3$.

$$\frac{E(Z_{t+1,m}| \ln \theta_t = a_1, \ln I_t = \ell_1) - E(Z_{t+1,m}| \ln \theta_t = a_2, \ln I_t = \ell_2)}{E(Z_{t+1,m}| \ln \theta_t = a_3, \ln I_t = \ell_3) - E(Z_{t+1,m}| \ln \theta_t = a_2, \ln I_t = \ell_2)} = \frac{\ln f_t(e^{a_1}, e^{\ell_1}) - \alpha_2}{\alpha_3 - \alpha_2}$$

We identify the function $\ln f_t(\theta_t, I_t)$ over its support by varying $a_1 \in \mathbb{R}$ and $\ell_1 \in \mathbb{R}$. We cannot of course use this transformation to identify the function at the point (a_2, ℓ_2), but the function evaluated at this point $f_t(e^{a_2}, e^{\ell_2})$ is already known by Definition 1. ■

A.1.3 Derivation of Example with CES Technology (Example 2)

$$\Delta_1 = \frac{\ln f_0(a_1, 0) - \ln f_0(1, 1)}{\ln f_0(e^1, e^1) - \ln f_0(1, 1)}$$

$$\Delta_1 = \frac{\ln(\gamma_0\, a_1) - 0}{\ln(e^1) - 0}$$

$$\Delta_1 = \frac{\ln(\gamma_0\, a_1)}{1}$$

$$e^{\Delta_1} = \gamma_0\, a_1$$

$$\gamma_0 = \frac{e^{\Delta_1}}{a_1}$$

Once we have γ_0, we can use the same ratio as before taking $a_1 \neq \{0, 1\}$, $a_3 \neq 0$, $\ell_1 = 1$, $a_2 = a_4 = \ell_2 = \ell_4$=1 and taking the limit $\ell_3 \to 0$ we have:

$$\Delta_2 = \frac{\ln f_0(a_1, 1) - \ln f_0(1, 1)}{\ln f_0(a_3, 0) - \ln f_0(1, 1)}$$

$$\Delta_2 = \frac{\ln f_0(a_1, 1) - 0}{\ln f_0(a_3, 0) - 0}$$

$$\Delta_2 = \frac{\ln f_0(a_1, 1)}{\ln f_0(a_3, 0)}$$

$$\Delta_2 = \frac{\ln(\gamma_0\, a_1^{\phi_0} + 1 - \gamma_0)}{\ln(\gamma_0\, a_3)}$$

$$\ln(\gamma_0\, a_3)\Delta_2 = \ln(\gamma_0\, a_1^{\phi_0} + 1 - \gamma_0)$$

$$(\gamma_0\, a_3)^{\Delta_2} = \gamma_0\, a_1^{\phi_0} + 1 - \gamma_0$$

$$(a_1)^{\phi_0} = \frac{(\gamma_0\, a_3)^{\Delta_2} - 1 + \gamma_0}{\gamma_0}$$

$$\phi_0 \ln(a_1) = \ln\left(\frac{(\gamma_0\, a_3)^{\Delta_2} - 1 + \gamma_0}{\gamma_0}\right)$$

$$\phi_0 = \frac{\ln\left(\frac{(\gamma_0\, a_3)^{\Delta_2} - 1 + \gamma_0}{\gamma_0}\right)}{\ln(a_1)}$$

A.1.4 Technologies and Output Elasticities

One rationale for the choice of a technology specification with non-constant returns to scale is the flexibility this specification offers with respect to the implied output elasticity. We consider the output elasticity with respect to investment defined as

$$\epsilon_I \equiv \frac{\partial \ln \theta_{t+1}}{\partial \ln I_t}$$

This elasticity is key to quantifying the effects of policy interventions.

In the general CES case, with technology given by

$$\theta_{t+1} = \left[\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t}\right]^{\frac{\psi_t}{\phi_t}},$$

the output elasticity is given by

$$\epsilon_I = \frac{\psi_t}{\phi_t}\left[\gamma \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t}\right]^{\frac{\psi_t}{\phi_t}-1} \phi(1-\gamma_t) I_t^{\phi_t - 1} \cdot \frac{I_t}{\left[\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t}\right]^{\frac{\psi_t}{\phi_t}}}$$

$$= \frac{\psi_t (1-\gamma_t) I_t^{\phi_t}}{\gamma_t \theta_t^{\phi_t} + (1-\gamma_t) I_t^{\phi_t}} \in [0, \infty)$$

In the special case of constant returns to scale (CRS), $\psi_t = 1$, and $\epsilon_I \in (0,1)$. CRS implies this elasticity is bounded from above by 1. The general, non-constant returns to scale, case allows a larger than unit elastic response.

Similarly, the general translog technology,

$$\ln \theta_{t+1} = \alpha_{1t} \ln \theta_t + \alpha_{2t} \ln I_t + \alpha_{3t} \ln \theta_t \ln I_t$$

with elasticity

$$\epsilon_I = \alpha_{1t} + \alpha_{3t} \ln \theta_t$$

also allows general higher than unit elastic elasticities.

The main insight we want to underline is that the CES technology with constant return to scale restricts the output elasticity to be between 0 and 1: a one percent change in investment leads to a less than one percent change in next period skills. This prediction is independent of data, hence it can potentially be very restrictive in the context of child development and skills formation.

A.2 Additional Tables and Figures

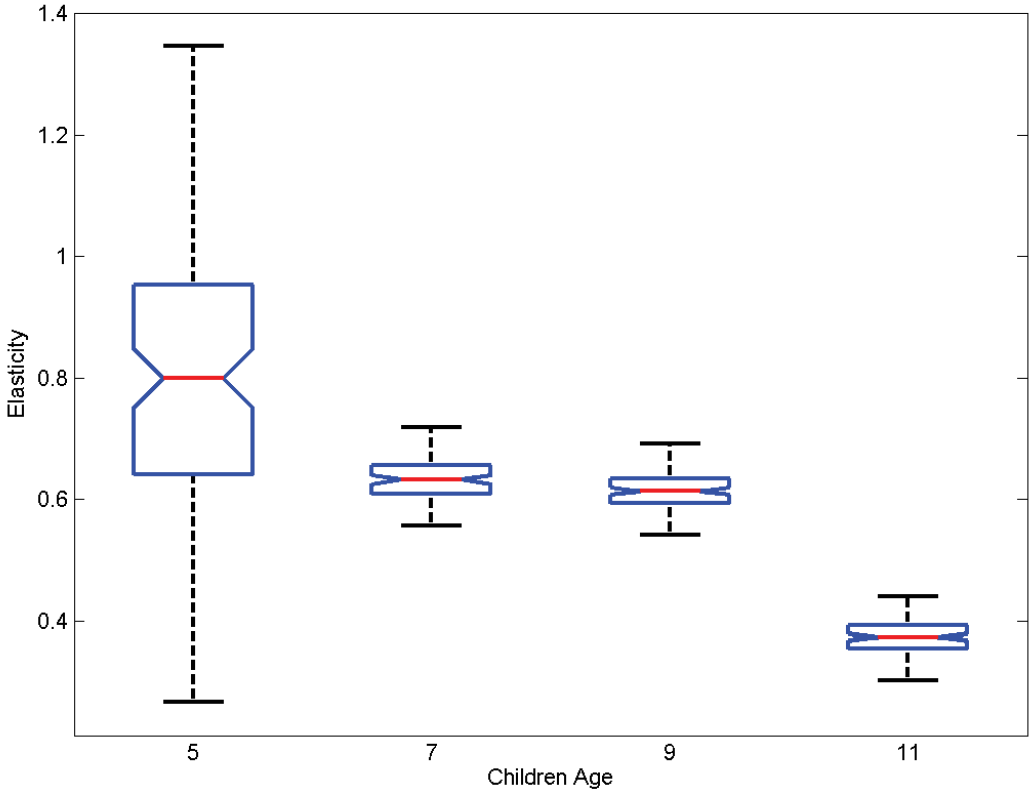
A.2.1 Additional Tables for Model 1 Corrected for Measurement Error

Table A.1: Estimates for Income Process

Constant	0.377
	(0.013)
Log Family Income t-1	0.753
	(0.008)
Variance Innovation	0.579
	(0.008)

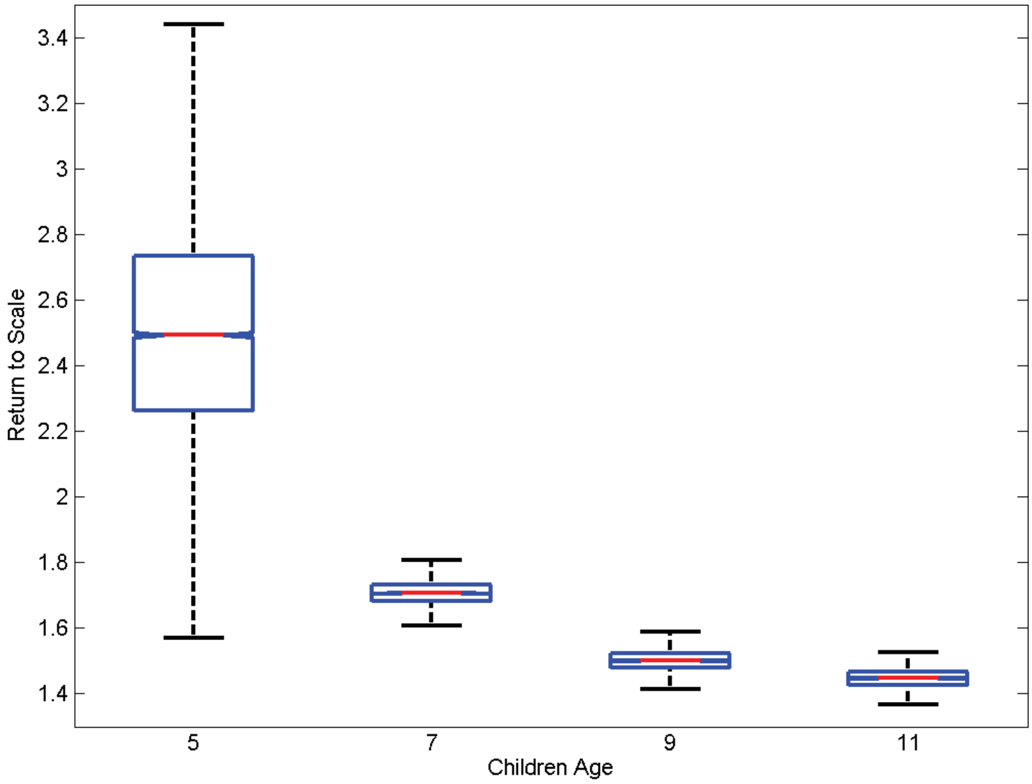
Notes: This table shows the estimates for the income process. The dependent variable is log family income at time t. Log Family Income $t-1$ is log family income two years prior. Standard errors in parenthesis are computed using a cluster bootstrap.

Figure A.1: Distribution of Elasticity of Next Period Skills with respect to Investment by Age



Notes: This figure shows the box plot for the elasticity of next period skills with respect to investment by different ages in the estimated Model 1 controlling for measurement error. The box plot is constructed as follow: the "central box" represents the central 50% of the data. Its lower and upper boundary lines are at the 25^{th} and 75^{th} quantile of the data. The central line indicates the median of the data while the two extreme lines (the top and the bottom ones) represents the 5^{th} and 95^{th} percentiles.

Figure A.2: Distribution of Technology Return to Scale by Age



Notes: This figure shows the box plot for the technology return to scale by different ages in the estimated Model 1 controlling for measurement error. The box plot is constructed as follow: the "central box" represents the central 50% of the data. Its lower and upper boundary lines are at the 25^{th} and 75^{th} quantile of the data. The central line indicates the median of the data while the two extreme lines (the top and the bottom ones) represents the 5^{th} and 95^{th} percentiles.

A.2.2 Descriptive Statistics

Table A.2: Children's Skills Measures

Measures	Range Values	Age Range	Scoring Order
(The Peabody Individual Achievement Test):			
Math	0-84	5-14	Positive
Recognition	0-84	5-14	Positive
Comprehensive	0-84	5-14	Positive

Notes: This table shows the features of children cognitive measures. The first column indicate each type of children skills measure we use to estimate our model. The second column shows the minimum and maximum value that each measure takes. The third column shows the minimum and maximum children age at which each measure is available. The last column indicates whether the measure is ordered positively (the higher score tend to reveal higher skills) or negatively (the lower score tend to reveal higher skills).

Table A.3: Mothers Cognitive Skills Measures

Measures	Range Values	Scoring Order
Arithmetics	0-30	Positive
Word Knowledge	0-35	Positive
Paragraph Composition	0-15	Positive
Numeric Operations	0-50	Positive
Coding Speed	0-84	Positive
Math Knowledge	0-25	Positive

Notes: This table shows the features of mother cognitive measures. The first column indicate each type of mother cognitive skills measure we use to estimate our model. The second column shows the minimum and maximum value that each measure takes. The last column indicates whether the measure is ordered positively (the higher score tend to reveal higher skills) or negatively (the lower score tend to reveal higher skills).

Table A.4: Mothers Noncognitive Skills Measures

Type of variables	Range Values	Label	Scoring Order
Mother Noncognitive Measures			
(Rosenberg indexes): I am a person of worth I have a number of good qualities I am able to do things as well as most other people I take a positive attitude toward myself	1-4	1= Strongly agree 2= Agree 3=Disagree 4=Strongly disagree	Negative
I am inclined to feel that I am a failure I felt I do not have much to be proud of I wish I could have more respect for myself I certainly feel useless at times At times I think I am no good at all	1-4	1= Strongly agree 2= Agree 3=Disagree 4=Strongly disagree	Positive
(Rotter Indexes): Rotter 1 (Life is in control or not)	1-4	1= In Control and closer to my opinion 2= In control but slightly closer to my opinion 3= Not in control but slightly closer to my opinion 4= Not in control and closer to my opinion	Negative
Rotter 2 (Plans work vs Matter of Luck)	1-4	1= Plans work and closer to my opinion 2= Plans work but slightly closer to my opinion 3= Matter of Luck but slightly closer to my opinion 4= Matter of Luck and closer to my opinion	Negative
Rotter 3 (Luck not a factor vs Flip a coin)	1-4	1= Luck not a factor and closer to my opinion 2=Luck not a factor but slightly closer to my opinion 3= Flip a coin but slightly closer to my opinion 4= Flip a coin and closer to my opinion	Negative
Rotter 4 (Luck big role vs Luck no role)	1-4	1= Luck big role and closer to my opinion 2=Luck big role but slightly closer to my opinion 3= Luck no role but slightly closer to my opinion 4= Luck no role and closer to my opinion	Positive

Notes: This table shows the features of mother noncognitive measures. The first column indicate each type of mother cognitive skills measure we use to estimate our model. The second column shows the minimum and maximum value that each measure takes. The third column shows the type of answers associated with each measure value. The last column indicates whether the measure is ordered positively (the higher score tend to reveal higher skills) or negatively (the lower score tend to reveal higher skills).

Table A.5: Descriptive Statistics about Children's Cognitive Skills Measures

Measures	Mean	Std	Min	Max	Number of Values
Age 5-6					
PIAT Math	11.858	4.278	0.000	37.000	32.000
PIAT Recognition	12.864	5.048	0.000	57.000	35.000
PIAT Comprehensive	12.770	4.930	0.000	49.000	35.000
Age 7-8					
PIAT Math	23.016	8.681	0.000	74.000	58.000
PIAT Recognition	25.748	8.774	0.000	80.000	67.000
PIAT Comprehensive	24.099	8.142	0.000	69.000	60.000
Age 9-10					
PIAT Math	38.720	10.832	0.000	84.000	71.000
PIAT Recognition	40.825	11.487	0.000	84.000	76.000
PIAT Comprehensive	37.540	10.231	0.000	78.000	64.000
Age 11-12					
PIAT Math	48.184	10.543	0.000	84.000	78.000
PIAT Recognition	51.079	13.278	0.000	84.000	74.000
PIAT Comprehensive	45.732	11.272	0.000	84.000	72.000
Age 13-14					
PIAT Math	53.767	11.387	0.000	84.000	78.000
PIAT Recognition	58.670	14.262	0.000	84.000	74.000
PIAT Comprehensive	51.015	12.229	0.000	84.000	74.000

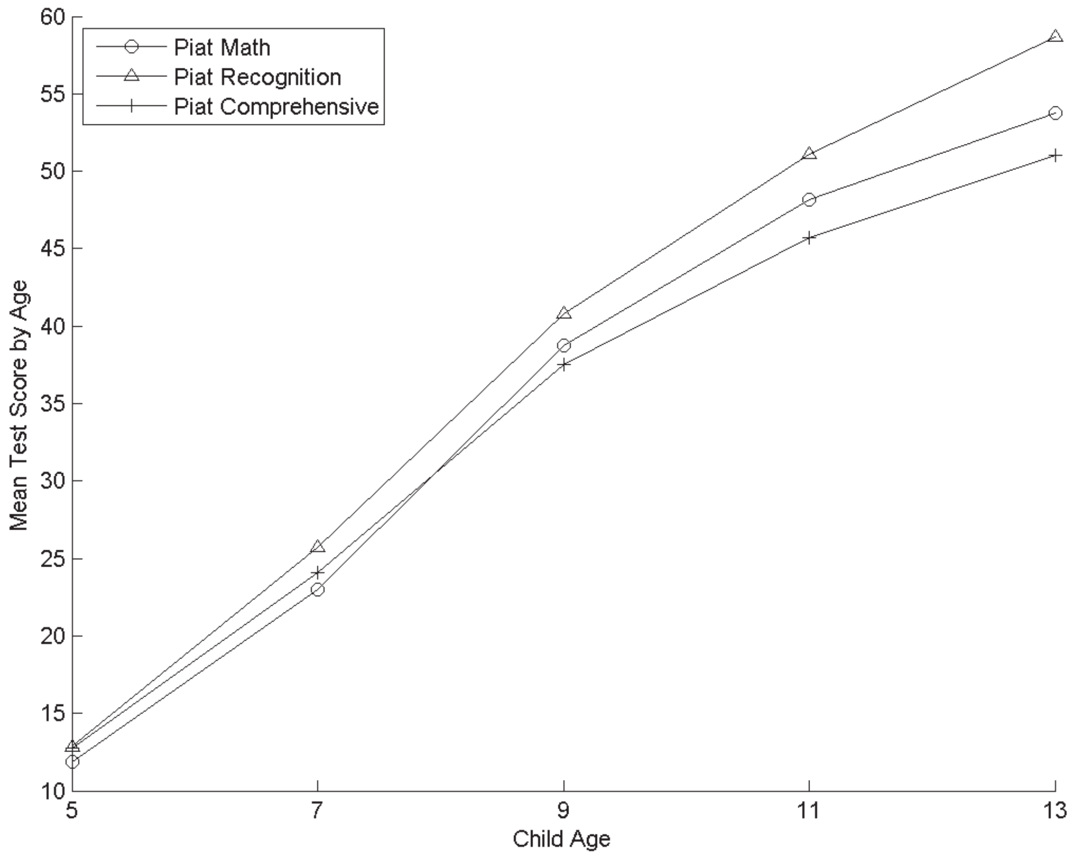
Notes: This table shows main sample statistics of children cognitive skills measures by children age.

Table A.6: Descriptive Statistics of Mother Cognitive and Noncognitive Skills Measures

Mother Cognitive Skills					
Measures	Mean	Std	Min	Max	Number of Values
Mom's Arithmetic Reasoning Test Score	13.946	6.603	0.000	30.000	31.000
Mom's Word Knowledge Test Score	21.773	8.562	0.000	35.000	36.000
Mom's Paragraph Composition Test Score	9.620	3.778	0.000	15.000	16.000
Mom's Numerical Operations Test Score	31.044	11.831	0.000	50.000	51.000
Mom's Coding Speed Test Score	42.953	17.468	0.000	84.000	85.000
Mom's Mathematical Knowledge Test Score	10.853	5.867	0.000	25.000	26.000
Mother Non Cognitive Skills					
Mom's Self-Esteem: "I am a person of worth"	2.461	0.549	0.000	3.000	4.000
Mom's Self-Esteem: " I have good qualities"	2.338	0.539	0.000	3.000	4.000
Mom's Self-Esteem: "I am a failure"	3.379	0.618	1.000	4.000	4.000
Mom's Self-Esteem: "I am as capable as others"	2.291	0.567	0.000	3.000	4.000
Mom's Self-Esteem: "I have nothing to be proud of"	3.360	0.669	1.000	4.000	4.000
Mom's Self-Esteem: "I have a positive attitude"	2.183	0.619	0.000	3.000	4.000
Mom's Self-Esteem: "I wish I had more self-respect"	2.796	0.817	1.000	4.000	4.000
Mom's Self-Esteem: "I feel useless at times"	2.650	0.770	1.000	4.000	4.000
Mom's Self-Esteem: "I sometimes think I am no good"	3.039	0.802	1.000	4.000	4.000
Mom's Rotter Score:"I have no control"	2.863	1.058	1.000	4.000	4.000
Mom's Rotter Score: "I make no plans for the future"	2.386	1.192	1.000	4.000	4.000
Mom's Rotter Score: "Luck is big factor in life"	3.205	0.856	1.000	4.000	4.000
Mom's Rotter Score: "Luck plays big role in my life"	2.594	1.024	1.000	4.000	4.000

Notes: This table shows main sample statistics of mother cognitive skills measures.

Figure A.3: Descriptive Statistics: Mean of PIATs over the Childhood



Notes: This figure shows the mean Piat Math, Recognition and Comprehensive test scores by age. The x-axis shows children age. Child age of 5 is age 5-6, 7 is age 7-8, and so on.

A.2.3 Measurement Parameter Estimates

Table A.7: Measurement Parameter Estimates for Children's Cognitive Measures

Measures	μ	λ	Signal	Noise
Age 5-6				
PIAT Math	11.858	1.000	0.270	0.730
PIAT Recognition	12.864	2.238	0.972	0.028
PIAT Comprehensive	12.770	2.159	0.948	0.052
Age 7-8				
PIAT Math	11.858	1.000	0.757	0.243
PIAT Recognition	15.592	0.906	0.608	0.392
PIAT Comprehensive	15.014	0.802	0.554	0.446
Age 9-10				
PIAT Math	11.858	1.000	0.779	0.221
PIAT Recognition	10.297	1.136	0.894	0.106
PIAT Comprehensive	12.273	0.936	0.765	0.235
Age 11-12				
PIAT Math	11.858	1.000	0.803	0.197
PIAT Recognition	2.107	1.347	0.918	0.082
PIAT Comprehensive	6.129	1.089	0.833	0.167
Age 13-14				
PIAT Math	11.858	1.000	0.927	0.073
PIAT Recognition	8.556	1.195	0.845	0.155
PIAT Comprehensive	9.041	1.002	0.806	0.194

Notes: This table shows the measurement error parameters and associated statistics for children cognitive measures. The first two columns shows the measurement parameters (μ and λ) while the last two columns shows the signal and noise variance decomposition for the children cognitive measures.

Table A.8: Measurement Parameter Estimates for Mother Cognitive and Noncognitive Measures

Measures	μ	λ	Signal	Noise
Mother Cognitive Skills				
Mom‘s Arithmetic Reasoning Test Score	13.946	1.000	0.692	0.308
Mom‘s Word Knowledge Test Score	21.773	1.345	0.745	0.255
Mom‘s Paragraph Composition Test Score	9.620	0.584	0.722	0.278
Mom‘s Numerical Operations Test Score	31.044	1.720	0.638	0.362
Mom‘s Coding Speed Test Score	42.953	2.308	0.527	0.473
Mom‘s Mathematical Knowledge Test Score	10.853	0.854	0.639	0.361
Mother Non Cognitive Skills				
Mom‘s Self-Esteem: ”I am a person of worth”	2.461	1.000	0.152	0.848
Mom‘s Self-Esteem: ” I have good qualities”	2.338	1.263	0.252	0.748
Mom‘s Self-Esteem: ”I am a failure”	3.379	1.612	0.311	0.689
Mom‘s Self-Esteem: ”I am as capable as others”	2.291	1.127	0.181	0.819
Mom‘s Self-Esteem: ”I have nothing to be proud of”	3.360	1.746	0.312	0.688
Mom‘s Self-Esteem: ”I have a positive attitude”	2.183	1.474	0.260	0.740
Mom‘s Self-Esteem: ”I wish I had more self-respect”	2.796	2.080	0.297	0.703
Mom‘s Self-Esteem: ”I feel useless at times”	2.650	1.861	0.268	0.732
Mom‘s Self-Esteem: ”I sometimes think I am no good”	3.039	2.096	0.313	0.687
Mom‘s Rotter Score:”I have no control”	2.461	1.000	0.092	0.908
Mom‘s Rotter Score: ”I make no plans for the future”	2.338	1.263	0.140	0.860
Mom‘s Rotter Score: ”Luck is big factor in life”	3.379	1.612	0.118	0.882
Mom‘s Rotter Score: ”Luck plays big role in my life”	2.291	1.127	0.044	0.956

Notes: This table shows the measurement error parameters and associated statistics for mother cognitive and noncognitive measures. The first two columns shows the measurement parameters (μ and λ) while the last two columns shows the signal and noise variance decomposition for the mother measures.

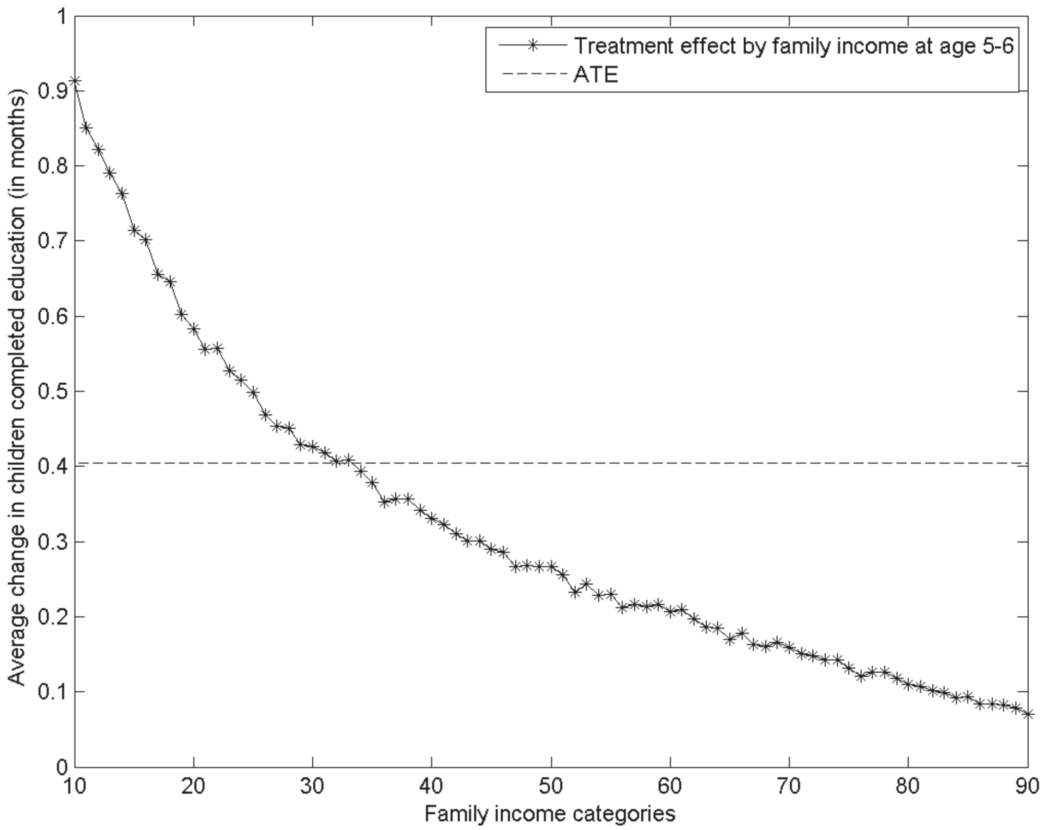
A.2.4 Estimates and Results for Model 2 with Measurement Error Corrected Estimator

Table A.9: Estimates for Investment (Model 2)

Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.230	0.069	0.068	0.065
	(0.059)	(0.021)	(0.029)	(0.030)
Log Mother Cognitive Skills	0.071	0.004	0.011	-0.005
	(0.022)	(0.009)	(0.014)	(0.012)
Log Mother Noncognitive Skills	0.359	0.711	0.660	0.678
	(0.131)	(0.059)	(0.084)	(0.084)
Log Family Income	0.341	0.217	0.261	0.262
	(0.076)	(0.054)	(0.072)	(0.082)
Variance Shocks	1.186	0.969	0.831	1.028
	(0.232)	(0.134)	(0.211)	(0.259)

Notes: This table shows the measurement error corrected estimates for the investment equation for Model 2. Each column shows the coefficients of the investment equation at the given ages. The dependent variable is investment in period t which is determined by the covariates at time t . For example, the first column shows the coefficients at age 5-6 for parental investments and child's skill and family income at age 5-6 as well. Standard errors in parenthesis are computed using a cluster bootstrap.

Figure A.4: Heterogeneity in Policy Effects by Age 5 Household Income (Outcome: Schooling at Age 23)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income for the estimated Model 2, controlling for measurement error. Each income category is defined as the people contained between n^{th} and the $n-1^{th}$ of the percentiles of the income distribution. For example, Income category 10 in the graph means the people who belong between the 9^{th} and 10^{th} percentile of the income distribution. In the estimated income distribution for our sample, income categories 10, 50, and 90 contain families with about $14,000, $45,000, and $145,000 of annual family income. This figure also plots the average effect over the income distribution.

A.2.5 *Estimates and Results without Measurement Error Correction (Model 1 and Model 2)*

Table A.10: Estimates for Investment (Model 1 and Model 2)

	Model 1 (Free Return to Scale Technology and TFP Dynamics)				Model 2 (Restricted Return to Scale Technology and No TFP Dynamics)			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.083	0.032	0.024	0.015	0.083	0.045	0.030	0.014
	(0.023)	(0.009)	(0.009)	(0.007)	(0.023)	(0.012)	(0.011)	(0.007)
Log Mother Cognitive Skills	0.082	0.010	0.010	-0.002	0.082	0.010	0.010	-0.002
	(0.019)	(0.011)	(0.014)	(0.011)	(0.019)	(0.011)	(0.014)	(0.011)
Log Mother Noncognitive Skills	0.248	0.454	0.442	0.553	0.248	0.448	0.440	0.553
	(0.093)	(0.073)	(0.098)	(0.074)	(0.093)	(0.073)	(0.098)	(0.074)
Log Family Income	0.587	0.504	0.524	0.434	0.587	0.498	0.521	0.435
	(0.074)	(0.070)	(0.095)	(0.077)	(0.074)	(0.069)	(0.095)	(0.078)
Variance Shocks	1.635	1.522	1.537	1.535	1.635	1.504	1.529	1.537
	(0.224)	(0.172)	(0.364)	(0.327)	(0.224)	(0.168)	(0.360)	(0.329)

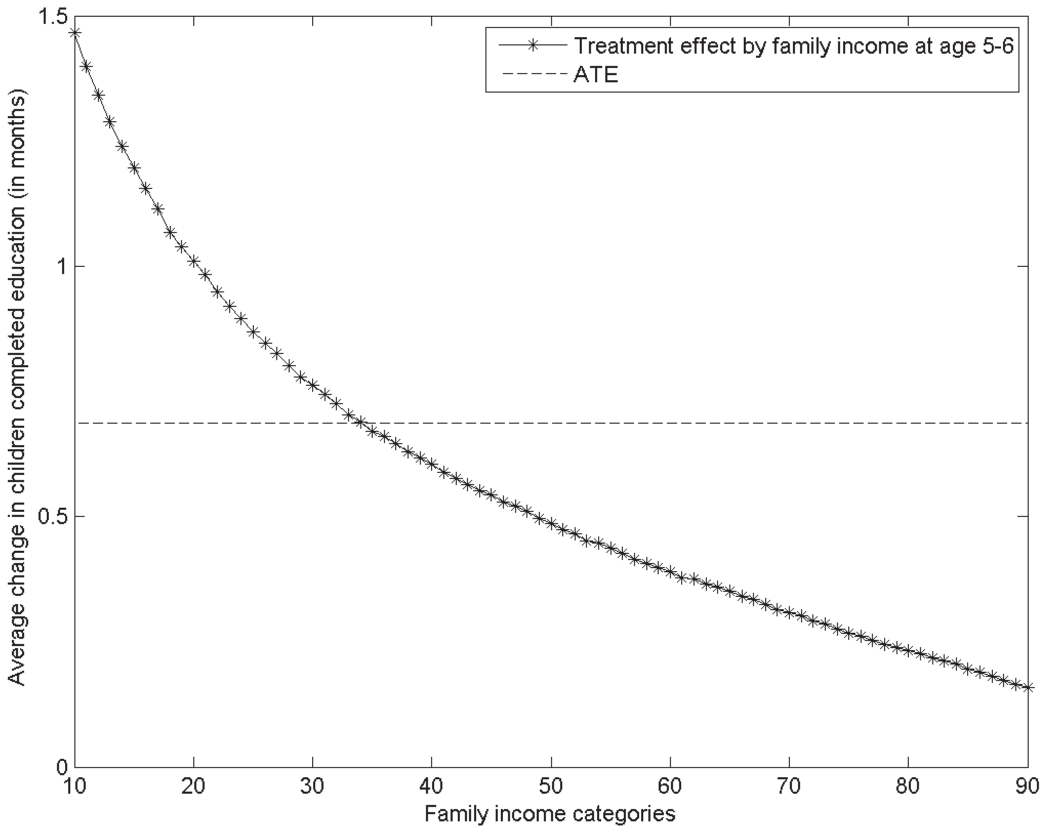
Notes: This table shows the estimates (not corrected for measurement error) for the investment equation for both Model 1 and Model 2. Each column shows the coefficients of the investment equation at the given ages. The dependent variable is investment in period t which is determined by the covariates at time t . For example, the first column shows the coefficients at age 5-6 for parental investments and child's skill and family income at age 5-6 as well. Standard errors in parenthesis are computed using a cluster bootstrap.

Table A.11: Estimates for Skill Technology (Model 1 and Model 2)

	Model 1 (Free Return to Scale Technology and TFP Dynamics)				Model 2 (Restricted Return to Scale Technology and No TFP Dynamics)			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.875 (0.057)	0.771 (0.022)	0.669 (0.017)	0.770 (0.018)	0.625 (0.047)	0.868 (0.039)	0.897 (0.039)	0.880 (0.052)
Log Investment	0.518 (0.089)	0.069 (0.066)	0.042 (0.061)	0.325 (0.099)	0.370 (0.045)	0.125 (0.038)	0.101 (0.039)	0.127 (0.052)
(Log Skills * Log Investment)	0.006 (0.012)	0.007 (0.003)	0.002 (0.002)	-0.006 (0.002)	0.005 (0.009)	0.008 (0.004)	0.002 (0.002)	-0.007 (0.003)
Return to scale	1.399 (0.098)	0.846 (0.072)	0.713 (0.063)	1.089 (0.096)	1.000 (-)	1.000 (-)	1.000 (-)	1.000 (-)
Variance shocks	7.490 (0.127)	7.673 (0.145)	6.716 (0.192)	7.382 (0.220)	5.354 (0.386)	6.155 (0.565)	7.211 (0.769)	9.092 (0.980)
Log TFP	12.789 (0.215)	18.491 (0.299)	18.477 (0.444)	14.011 (0.690)	0.000 (-)	0.000 (-)	0.000 (-)	0.000 (-)

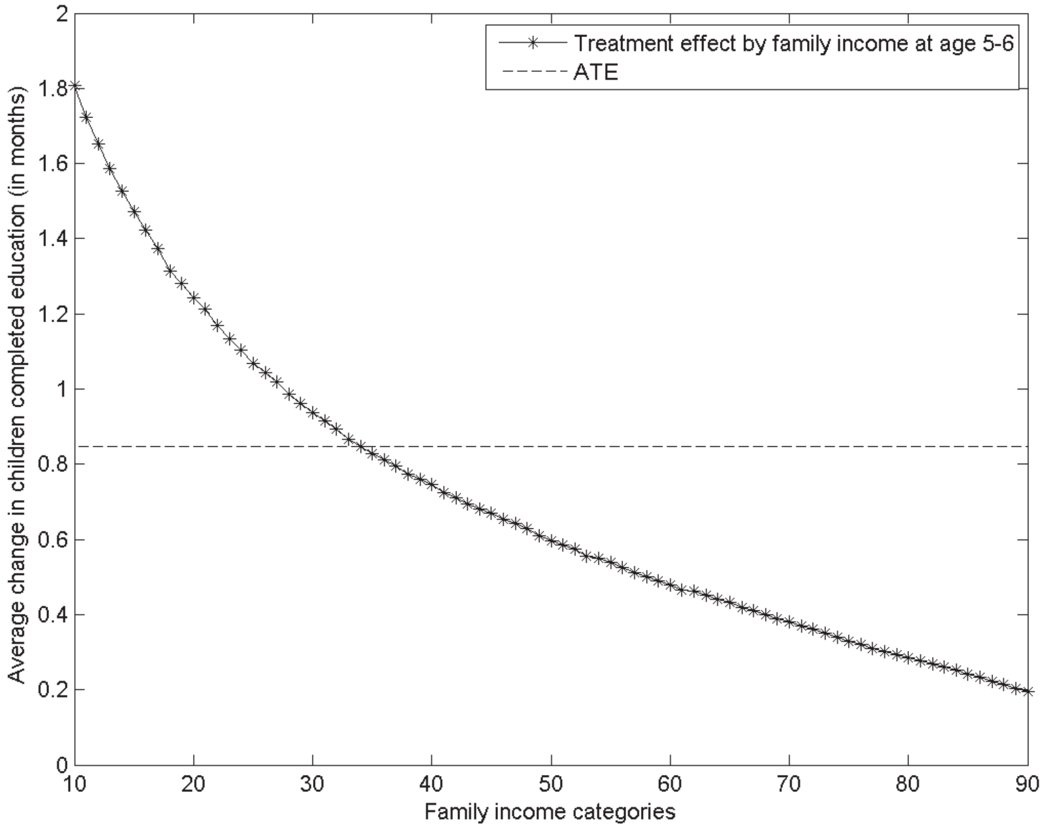
Notes: This table shows the estimates (not corrected for measurement error) for the technology of skills formation and the technology return to scale (i.e. the sum of the share parameters for each input) for not measurement error corrected estimates of both Model 1 and Model 2. Each column shows the coefficients of the technology of skills formations at the given age. The dependent variable is log skills in the next period $t+1$ while the covariates (inputs) are at time t. For example, the first column shows the coefficients for the skills inputs at age 5-6 which lead to log skills at age 7-8. Standard errors in parenthesis are computed using a cluster bootstrap.

Figure A.5: Heterogeneity in Policy Effects by Age 5 Household Income (Outcome: Schooling at Age 23, Model 1)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income for the estimated Model 1, not controlling for measurement error. Each income category is defined as the people contained between n^{th} and the $n-1^{th}$ of the percentiles of the income distribution. For example, Income category 10 in the graph means the people who belong between the 9^{th} and 10^{th} percentile of the income distribution. In the estimated income distribution for our sample, income categories 10, 50, and 90 contain families with about $14,000, $45,000, and $145,000 of annual family income. This figure also plots the average effect over the income distribution.

Figure A.6: Heterogeneity in Policy Effects by Age 5 Household Income (Outcome: Schooling at Age 23, Model 2)



Notes: This figure plots the heterogeneous effect of a $1,000 income transfer at age 5-6 on completed months of schooling by the percentile of initial (age 5-6) family income for the estimated Model 2, not controlling for measurement error. Each income category is defined as the people contained between n^{th} and the $n-1^{th}$ of the percentiles of the income distribution. For example, Income category 10 in the graph means the people who belong between the 9^{th} and 10^{th} percentile of the income distribution. In the estimated income distribution for our sample, income categories 10, 50, and 90 contain families with about $14,000, $45,000, and $145,000 of annual family income. This figure also plots the average effect over the income distribution.

A.2.6 Skills measures in CNLSY79

Measures for Cognitive Skills

- **Peabody Picture Vocabulary Test**

  The Peabody Picture Vocabulary Test, revised edition (PPVT-R) "measures an individual's receptive (hearing) vocabulary for Standard American English and provides, at the same time, a quick estimate of verbal ability or scholastic aptitude" (see Dunn and Dunn, 1981). The PPVT was designed for use with individuals aged 2 to 40 years. The English language version of the PPVT-R consists of 175 vocabulary items of generally increasing difficulty. The child listens to a word uttered by the interviewer and then selects one of four pictures that best describes the word's meaning. The PPVT-R has been administered, with some exceptions, to NLSY79 children between the ages of 3-18 years of age until 1994, when children 15 and older moved into the Young Adult survey. In the current survey round, the PPVT was administered to children aged 4-5 and 10-11 years of age, as well as to some children with no previous valid PPVT score.

  The first item, or starting point, is determined based on the child's PPVT age. Starting at an age-specific level of difficulty is intended to reduce the number of items that are too easy or too difficult, in order to minimize boredom or frustration. The suggested starting points for each age can be found in the PPVT manual (see Dunn and Dunn, 1981).

  Testing begins with the starting point and proceeds forward until the child makes an incorrect response. If the child has made 8 or more correct responses before the first error, a "basal" is established. The basal is defined as the last item in the highest series of 8 consecutive correct answers. Once the basal is established, testing proceeds forwards, until the child makes six errors in eight consecutive items. If, however, the child gives an incorrect response before 8 consecutive correct answers have been made, testing proceeds backwards, beginning at the item just before the starting point, until 8 consecutive correct responses have been made. If a child does not make eight consecutive responses even after administering all of the items, he or she is given a basal of one. If a child has more than one series of 8 consecutive correct answers, the highest basal is used to compute the raw score.

  A "ceiling" is established when a child incorrectly identifies six of eight consecutive items. The ceiling is defined as the last item in the lowest series of eight consecutive items with six incorrect responses. If more than one ceiling is identified, the lowest ceiling is used to compute the raw score. The assessment is complete once both a basal and a ceiling have been established. The ceiling is set to 175 if the child never makes six errors in eight consecutive items.

  A child's raw score is the number of correct answers below the ceiling. Note that all answers below the highest basal are counted as correct, even if the child answered some of these items incorrectly. The raw score can be calculated by subtracting the number of errors between the highest basal and lowest ceiling from the item number of the lowest ceiling.

- **The Peabody Individual Achievement Test (PIAT): Math**

The PIAT Mathematics assessment protocol used in the field is described in the documentation for the Child Supplement (available on the Questionnaires page). This subscale measures a child's attainment in mathematics as taught in mainstream education. It consists of 84 multiple-choice items of increasing difficulty. It begins with such early skills as recognizing numerals and progresses to measuring advanced concepts in geometry and trigonometry. The child looks at each problem on an easel page and then chooses an answer by pointing to or naming one of four answer options.

Administration of this assessment is relatively straightforward. Children enter the assessment at an age-appropriate item (although this is not essential to the scoring) and establish a "basal" by attaining five consecutive correct responses. If no basal is achieved then a basal of "1" is assigned (see PPVT). A "ceiling" is reached when five of seven items are answered incorrectly. The non-normalized raw score is equivalent to the ceiling item minus the number of incorrect responses between the basal and the ceiling scores.

- **The Peabody Individual Achievement Test (PIAT): Reading Recognition**

  The Peabody Individual Achievement Test (PIAT) Reading Recognition subtest, one of five in the PIAT series, measures word recognition and pronunciation ability, essential components of reading achievement. Children read a word silently, then say it aloud. PIAT Reading Recognition contains 84 items, each with four options, which increase in difficulty from preschool to high school levels. Skills assessed include matching letters, naming names, and reading single words aloud.

  The only difference in the implementation procedures between the PIAT Mathematics and PIAT Reading Recognition assessments is that the entry point into the Reading Recognition assessment is based on the child's score in the Mathematics assessment, although entering at the correct point is not essential to the scoring.

  The scoring decisions and procedures are identical to those described for the PIAT Mathematics assessment.

- **The Peabody Individual Achievement Test (PIAT): Reading Comprehension**

  The Peabody Individual Achievement Test (PIAT) Reading Comprehension subtest measures a child's ability to derive meaning from sentences that are read silently. For each of 66 items of increasing difficulty, the child silently reads a sentence once and then selects one of four pictures that best portrays the meaning of the sentence.

  Children who score less than 19 on Reading Recognition are assigned their Reading Recognition score as their Reading Comprehension score. If they score at least 19 on the Reading Recognition assessment, their Reading Recognition score determines the entry point to Reading Comprehension. Entering at the correct location is, however, not essential to the scoring.

Basals and ceilings on PIAT Reading Comprehension and an overall nonnormed raw score are determined in a manner identical to the other PIAT procedures. The only difference is that children for whom a basal could not be computed (but who otherwise completed the comprehension assessment) are automatically assigned a basal of 19. Administration instructions can be found in the assessment section of the Child Supplement.

A.3 Alternative Measures

One of the characteristics of the data used to study child development is the rich variety skill measures. The previous sections considered identification where the skill measures are in a "raw" form: each measure is a linear function of the latent log skill. This measurement system, while commonly assumed in the prior literature, is in some respects a "best case."

In this section, we consider alternative forms of measures and re-examine whether we can identify the same types of production technologies using these alternative measures. We consider four classes of measures which are frequently encountered empirically: (i) *age-standardized* measures where the raw measures are transformed ex post to have mean 0 and standard deviation 1 for the sample at hand; (ii) *relative* measures where the measures reflect not the level of a child's skill but the child's skill relative to the population mean; (iii) *ordinal* measures which provide a discrete ranking of children's skills; and iv) *censored* measures where the measures are truncated with a "floor" (finite minimum value) and/or a "ceiling" (finite maximum value). For each type of measure, we discuss which of our prior identification results still hold, if any, and what auxiliary assumptions would be sufficient to restore our identification results.

A.3.1 Age-Standardized Measures

Age-standardized measures are defined as the following transformation of raw measures $Z_{t,m}$:

$$Z^S_{t,m} = \frac{Z_{t,m} - E(Z_{t,m})}{V(Z_{t,m})^{1/2}}. \tag{A.1}$$

By construction, these measures are mean 0 and standard deviation 1 for all child ages.

Our main identification result using standardized measures (Theorem 1) continues to hold if the technology of skill formation has known scale and location functions (KLS, Definition 1). To show this, we can re-write the standardized measures as a linear function of the latent variable:

$$Z^S_{t,m} = \mu^S_{t,m} + \lambda^S_{t,m} \ln\theta_t + \epsilon^S_{t,m}$$

where the measurement parameters and measurement error are

$$\mu^S_{t,m} = -\lambda^S_{t,m}(V(\ln\theta_t)) \cdot E(\ln\theta_t)$$

$$\lambda^S_{t,m} = \frac{\lambda_{t,m}}{V(Z_{t,m})^{1/2}} = \frac{\lambda_{t,m}}{(\lambda^2_{t,m}V(\ln\theta_t) + V(\epsilon_t))^{1/2}}$$

$$\epsilon^S_{t,m} = \frac{\epsilon_{t,m}}{V(Z_{t,m})^{1/2}} = \frac{\epsilon_{t,m}}{(\lambda^2_{t,m}V(\ln\theta_t) + V(\epsilon_t))^{1/2}}$$

These expressions show that the standardized measurement parameters are linear functions of the underlying moments of the latent skill distribution. [1] The reason for the invariance of our identification result to the use of standardized or raw measures is that any measurement parameters are "transformed away" as shown in Lemma 1. More generally, identification of the KLS production technologies is invariant to any increasing linear transformation of the original raw measures, say $Z'_{t,m} = a + bZ_{t,m}$ for $a \in R$ and $b \in R^+$. [2]

However, the use of age-standardized measures may not be cost free in the sense that age-standardized measures, which are constructed to be age-stationary in their

[1] It is important to recognize that the use of standardized measures does not necessarily imply that any particular restriction on the underlying latent variables such as $E(\ln\theta_t) = 0$ or $V(\ln\theta_t) = 1$. The standardizations are necessarily in terms of the observed measures, not the unobserved latent variables.

[2] One caveat deserves mention. Recall that because the initial conditions are normalized to a particular measure, using standardized rather than raw measures can affect the normalized location and scale of the latent skills, and in general affect the values of the production parameters which are identified up to the normalized initial period measure.

first and second moments, contain no information about skill dynamics in these moments. For example, standardizing *age-invariant* measures, as defined in the previous section, so that the mean and variance of these measures is equal at all ages, would essentially "throw away" information regarding the average skill development of children across ages. This loss of information prevents the identification of the broader classes of technology of skills formation discussed above, the unknown Total Factor Productivity (TFP) functions (as in equation 14) or unknown scale functions (as in equation 15).

To see this point, recall that the identification of TFP or scaling parameter are based on additional information of the dynamics of measurement parameters. In the case of raw measures, those parameters are fully free parameters. On the other hand, when we use standardized measures, the new measurement parameters ($\mu^S_{t,m}$ and $\lambda^S_{t,m}$) are no longer free parameters but functions of the moments of the latent distribution. Hence, restricting the dynamics of the measurement parameters in this case (imposing Assumption 2 and Assumption 3) is equivalent to restricting the dynamics of the latent skills, and can restrict the possible classes of technologies. While age-standardizing measures may provide some descriptive value, in the context of identifying dynamic production technologies, there is simply no point to transforming the measures in this way and throwing away potentially important identifying information.

A.3.2 Relative Measures

Some of the proxies used to measure children outcomes come from surveys where observers (often mothers, fathers, or other caregivers) provide assessments of the child. It can be plausible then that these observers are actually evaluating the child with respect to their perceptions of the average in the population. We call this type of measure a *relative measures.* In this case, these measures can be written as:

$$Z^R_{t,m} = \mu^R_{t,m} + \lambda^R_{t,m}(\ln\theta_t - E(\ln\theta_t)) + \epsilon^R_{t,m}. \tag{A.2}$$

where $(\ln\theta_t - E(\ln\theta_t)$ is the latent variable being measured by $Z^R_{t,m}$, which we model as the deviation of the actual level of the child's skill $\ln\theta_t$ relative to the mean value in the population $E(\ln\theta_t)$. Relative measures are not ordinal ranking measures (which we discuss below) but a continuous measure of skills relative to the population mean. As with the age-standardized measures, the relative measures are an increasing linear function of the underlying latent variable, and therefore the main identification result in Theorem 1 continues to hold as the measurement parameters are "transformed away."

A.3.3 Ordinal Measures

We define ordinal measures the measures which are based on children rankings: this child has higher skills than another child. Let's assume that we observe in data children's skill rank. Let $Z_t = \{1, 2, \ldots, J\}$ be the child's human capital rank, with 1 highest level, and J lowest level. The observer (or us forming ranks from test scores) forms rank according to this ordinal model:

$$Z^O_{t,m} = \begin{cases} J \text{ if } \lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} < \kappa_{J,t,m} \\ J-1 \text{ if } \kappa_{J,t,m} < \lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} < \kappa_{J-1,t,m} \\ \vdots \\ 2 \text{ if } \kappa_{3,t,m} < \lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} < \kappa_{2,t,m} \\ 1 \text{ if } \lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} > \kappa_{2,t,m} \end{cases} \tag{A.3}$$

where the $\kappa_2, \ldots, \kappa_J$, with $\kappa_2 > \kappa_3, \ldots, \kappa_J$, are measurement parameters which provide the mapping from latent skills $\ln \theta_t$ and measurement error $\epsilon^O_{t,m}$ to the observed ordinal ranking values $Z^O_{t,m}$. The probability a child is ranked first ($j = 1$) is then

$$\begin{aligned} pr(Z^O_{t,m} = 1) &= pr(\lambda^O_{t,m} \ln \theta_t + \epsilon^O_{t,m} > \kappa_{2,t,m}) \\ &= F_\epsilon(\lambda^O_{t,m} \ln \theta_t - \kappa_{2,t,m}) \end{aligned}$$

where F_ϵ is the distribution function for the measurement error $\epsilon^O_{t,m}$.

With ordinal ranking measures the non-parametric identification result no longer holds. There is no longer a one-to-one mapping between a child's latent skills θ_t and expected measures, as multiple values of θ_t are consistent with a child having a certain rank. Without additional assumptions beyond Assumption 1 (independence of measures), ordinal measures of skills do not allow non-parametric identification of the continuous skill production function.

If the researcher were to assume a particular known distribution for the measurement errors F_ϵ, then under this assumption for an ordinal measure of $t+1$ skills we would have:

$$F_{\epsilon}^{-1}(pr(Z^{O}_{t+1,m} = 1 | \ln \theta_t, \ln I_t)) = \lambda_{t+1,m} f_t(I_t, \theta_t) - \kappa_{2,t+1,m}$$

where $pr(Z^{O}_{t+1,m} = 1 | \ln \theta_t, \ln I_t)$ is the probability the child receives rank 1 at age $t+1$ given inputs θ_t, I_t at age t. This expression shows that with a known distribution for measurement errors, we can then apply Theorem 1 to identify a KLS technology $f_t(I_t, \theta_t)$ up to this assumed distribution.

A.3.4 Censored Measures

Censored measures are defined as

$$Z^C_{t,m} = \begin{cases} \overline{Z} \text{ if } Z_{t,m} \geq \overline{Z} \\ Z_{t,m} \text{ if } \underline{Z} < Z_{t,m} < \overline{Z} \\ \underline{Z} \text{ if } Z_{t,m} < \underline{Z} \end{cases} \tag{A.4}$$

where $Z_{t,m} = \mu_{t,m} + \lambda_{t,m} \ln \theta_t + \epsilon_{t,m}$ is the "latent" measure, and $\overline{Z}$ ("ceiling") and $\underline{Z}$ ("floor"), with $\overline{Z} > \underline{Z}$, are the truncation points. Censoring occurs, for example, when a test score used as the measure has a maximum score (answering all questions correctly) and a minimum score (say answering none of the questions correctly). In practice, researchers can ascertain whether censoring is an important issue empirically by investigating what proportion of the sample actually has measured skills at the floor or ceiling points of the measure. Because censored measures do not have full support, the non-parametric identification result of Theorem 1 appears no longer to hold. As with the ordinal measures, auxiliary assumptions could be used to achieve identification up these additional assumptions (for a complete analyze of the problem, see Wang et al. 2009, Koedel and Betts, 2010)

A.4 Monte Carlo Exercise for Model 1 and Measurement Error Correction

We implement a Monte Carlo exercise to examine the properties of our estimator. The true data generating process is assumed to be the estimated (measurement error corrected) Model 1 with some additional parametric assumptions about the measurement error process. In order to simulate the dataset, we use the both the estimated measurement parameters and the joint distribution of children skills and investments. In addition, we assume that all the measurement noises are Normally distributed. [3] We generate a simulated longitudinal dataset of 10,000 children ranging from age 5-6 to age 13-14. In particular, the Monte Carlo analysis is performed estimating the model on 200 simulated data sets. In the following tables we show the mean estimates over the 200 estimates of the coefficients.

We focus only on estimates of skills technology, investment process and children's skills measurement parameters. Tables A.12-A.14 show true and mean estimated parameters. Overall, the estimator is able to recover the true parameters with minimal bias.

[3]We assume that the standard deviation of the error terms for all the skills measures are 0.5 (children and mothers) while we fix to 0.1 the standard deviation of the error terms for all the investment measures.

Table A.12: Monte Carlo Estimates for Investment Process

	True Parameters				Monte Carlo Estimates			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	0.230	0.027	0.020	0.018	0.249	0.026	0.020	0.018
Log Mother Cognitive Skills	0.071	0.004	0.012	-0.005	0.077	0.002	0.008	-0.011
Log Mother Noncognitive Skills	0.359	0.742	0.694	0.712	0.322	0.748	0.700	0.700
Log Family Income	0.341	0.227	0.274	0.275	0.352	0.224	0.272	0.292
Variance Shocks	1.186	1.019	0.868	1.087	1.263	0.993	0.827	1.103

Notes: This table shows the both the true estimates (reported also in Table 3) and the mean Monte Carlo estimates for the investment equation. Each column shows the coefficients of the investment equation at the given ages. The dependent variable is investment in period t which is determined by the covariates at time t . For example, the first column shows the coefficients at age 5-6 for parental investments and child's skill and family income at age 5-6 as well.

Table A.13: Monte Carlo Estimates for Skill Technology

	True Parameters				Monte Carlo Estimates			
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 5-6	Age 7-8	Age 9-10	Age 11-12
Log Skills	1.966	1.086	0.897	1.065	1.955	1.091	0.897	1.071
Log Investment	0.799	0.695	0.713	0.252	0.759	0.700	0.839	0.502
(Log Skills * Log Investment)	-0.105	-0.005	-0.003	0.003	-0.092	-0.005	-0.005	-0.002
Return to scale	2.660	1.776	1.606	1.320	2.623	1.786	1.731	1.571
Variance shocks	5.612	4.519	3.585	4.019	5.613	4.520	3.586	4.018
Log TFP	13.067	14.747	11.881	2.927	13.060	14.689	11.801	2.594

Notes: This table shows the both the true estimates (reported also in Table 4) and the mean Monte Carlo estimates for the technology of skills formation. Each column shows the coefficients of the technology of skills formations at the given age. The dependent variable is log skills in the next period $t+1$ while the covariates (inputs) are at time t. For example, the first column shows the coefficients for the skills inputs at age 5-6 which lead to log skills at age 7-8.

Table A.14: Monte Carlo Estimates for Measurement Parameters

	True Constant (μ)					Monte Carlo Constant (μ) Estimates				
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 13-14	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 13-14
PIAT Math	11.858	11.858	11.858	11.858	11.858	11.858	11.858	11.858	11.858	11.858
PIAT Recognition	12.864	15.592	10.297	2.107	8.556	12.864	15.592	10.298	2.110	8.555
PIAT Comprehensive	12.770	15.014	12.273	6.129	9.041	12.770	15.013	12.270	6.132	9.040

	True Factor Loadings (λ)					Monte Carlo Factor Loadings (λ) Estimates				
Parameter	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 13-14	Age 5-6	Age 7-8	Age 9-10	Age 11-12	Age 13-14
PIAT Math	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
PIAT Recognition	2.238	0.906	1.136	1.347	1.195	2.238	0.905	1.136	1.347	1.196
PIAT Comprehensive	2.159	0.802	0.936	1.089	1.002	2.159	0.802	0.936	1.089	1.002

Notes: This table shows the both the true estimates (reported also in Table A.7) and the mean Monte Carlo estimates for the measurement parameters of children skills measures equation. Each column shows the parameters at the given ages for each test score.

Chapter B

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

B.1 Data Appendix

The Add Health database was designed to study the impact of the social environment (i.e. friends, family, neighborhood and school) on adolescents' behavior in the United States of America. Add Health's original sample comprises students of 132 representative schools in the United States. There are 90,118 students, ranging between grades 7 and 12 in the 1994–1995 school year (Wave I). A subsample of students (20,745) was selected for an additional home interview (in-home). The home interview includes new questions for the children and a questionnaire for one of their parents. Information related to school quality and area of residence makes the data set attractive. Those surveyed were interviewed again in 1995–96 (Wave II), 2001–2002 (Wave III), and 2007–2008 (Wave IV). The data set includes specific information on family background, students' school grades and their scores in the Add Health Picture Vocabulary Test (AHPVT – a revised version of the Peabody Picture Vocabulary Test [PPVT]), as well as information about children's peers.

A main source of information which makes the Add Health data set particularly attractive for achieving the objective of this project is the friendship nomination. During the first two waves, children were asked, both during the in-home and in-school interviews, to nominate their best five male and best five female friends. This detailed information helped me to reconstruct the structure of friendship for every child in the sample by simply matching their identifier.

B.1.1 Measures for Children's Skills and Parental Investments

The Peabody Picture Vocabulary Test (PPVT) was developed in 1959 as a test of receptive vocabulary and is oriented to give an estimation of verbal ability and school aptitude. More generally, it intends to provide a measure of intelligence. The test is age standardized, and performance does not depend on the reading ability of the test-taker. The test-giver reads a word, and the test-taker selects the image she thinks best fits the meaning of the word from among four simple illustrations.

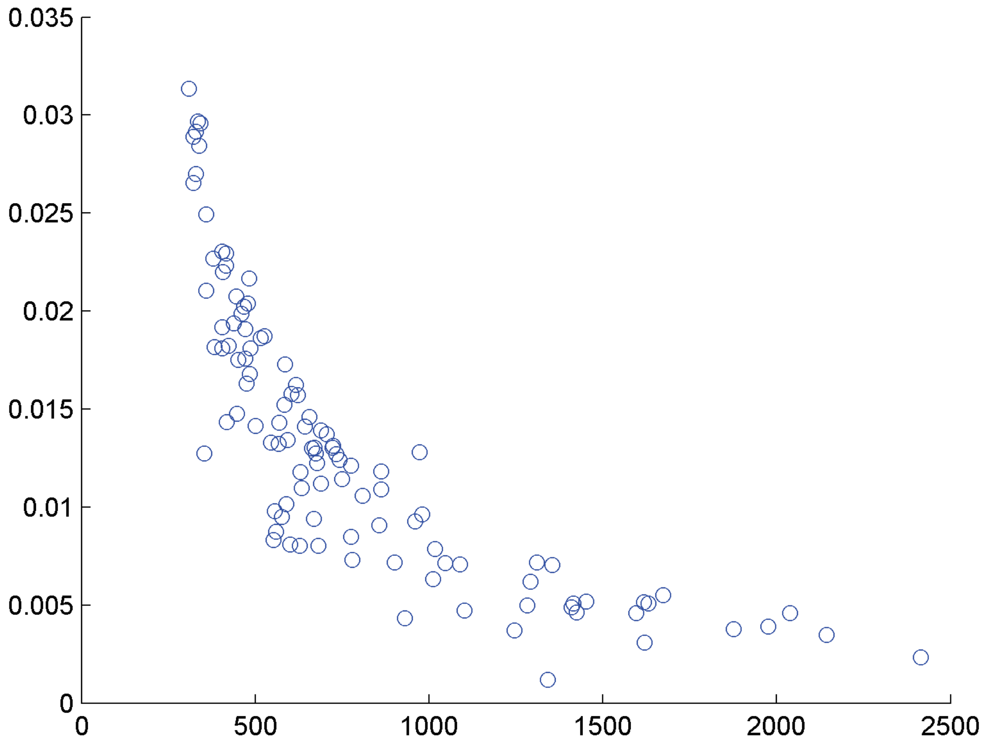
In the Add Health dataset, a computerized and shorter version of the Peabody Picture Vocabulary Test, the AHPVT, has been implemented. The AHPVT includes half of the questions of the original PPVT (every other item in the original sequence was selected for use). Add Health provides two versions of the AHPVT: raw and standardized test scores. The standardized version has a mean of 100 and a standard deviation of 15 and is standardized by age (for further technical details about the AHPVT, see Halpern, 2000).

Add Health provides information about the AHPVT test scores for Waves I, III and IV. During Wave I, respondents are between 11 and 21 years old. I consider this piece of information as one of the measures of children's latent skills. Additionally, I consider children's grades at school for both Wave I and Wave II. Analyzing all these multiple measurements, I am able to combine both the cross-sectional as well as the longitudinal information about child development.

During the in-home interview in Wave I and Wave II, children provided information about activities they had engaged in with their parents during the previous four weeks. These activities include: going shopping, sport activities, going to a movie/museum/concert or sport event, talking about personal problems or school, or working on a project for school. There are a total of nine activities each child can do with their parents. These types of activities provide information about the level of parental engagement with their child. During Wave I, a parent, preferably the resident mother, of each adolescent respondent was interviewed (in-home interview). The parent questionnaire included a question about the achieved level of education for the respondent, which is considered as one proxy for the mother's skills. During the same survey, the respondent provided information about total family income during the previous calendar year.

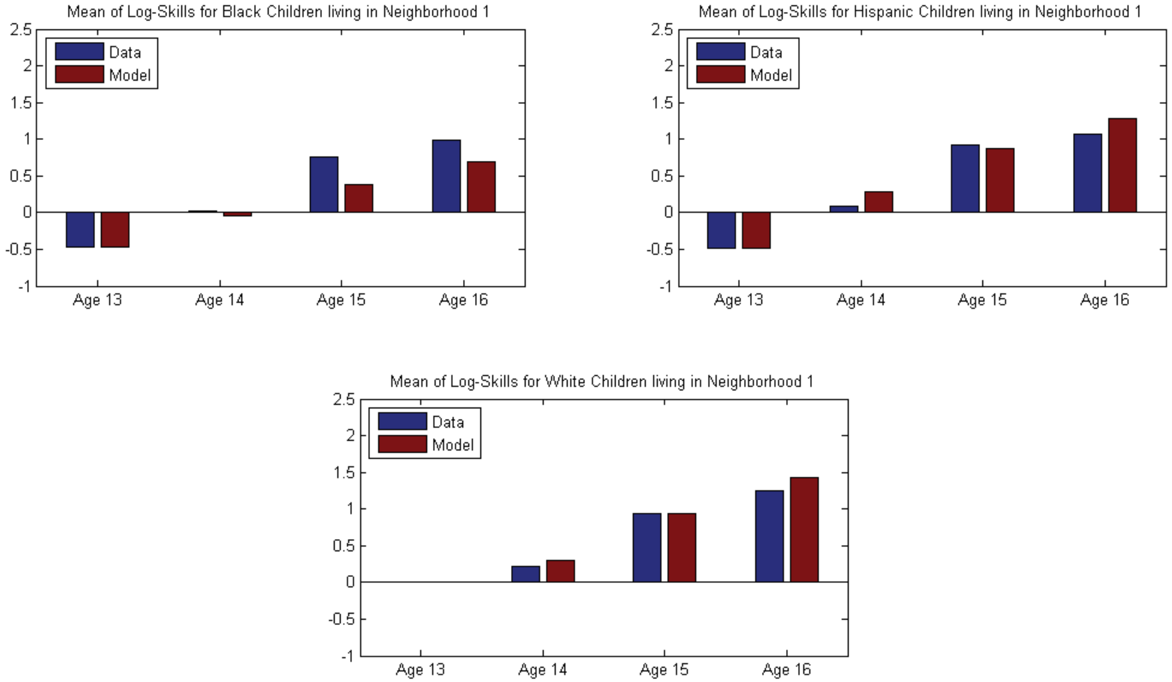
B.2 Additional Figures and Tables

Figure B.1: Probability of the Friendships and School's Size



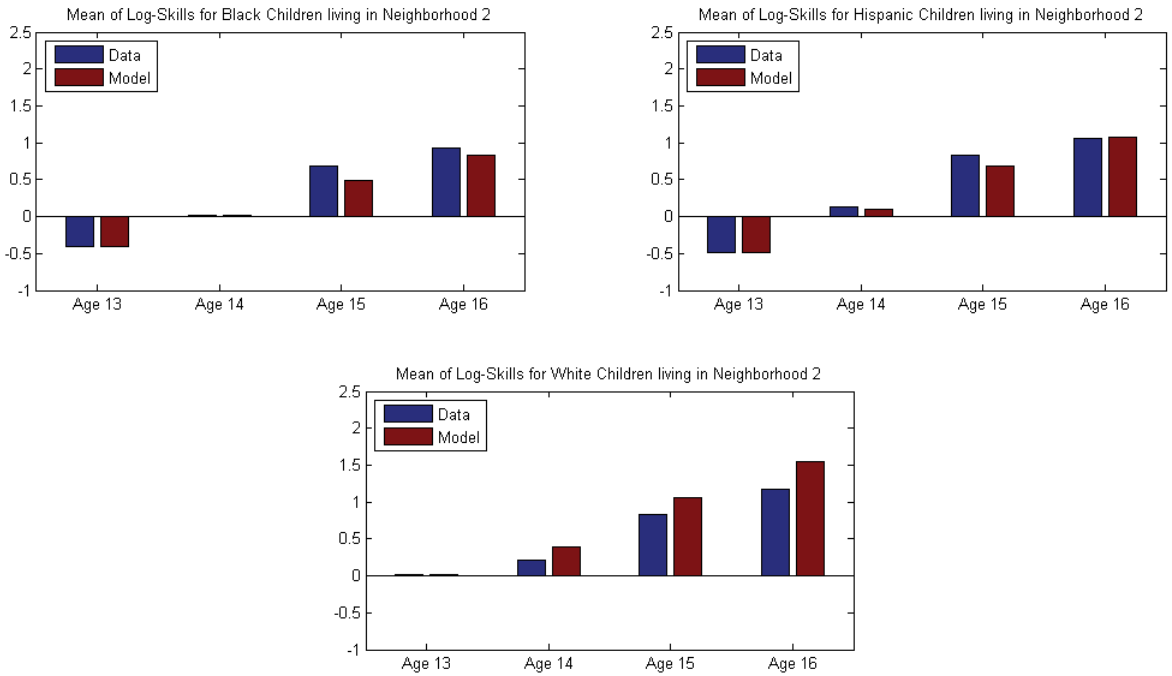
Notes: This figure shows the probability, for each school, that a child becomes friend with another child as function of school's size.

Figure B.2: Mean of Children Skills by Race (Low-Income Neighborhood)



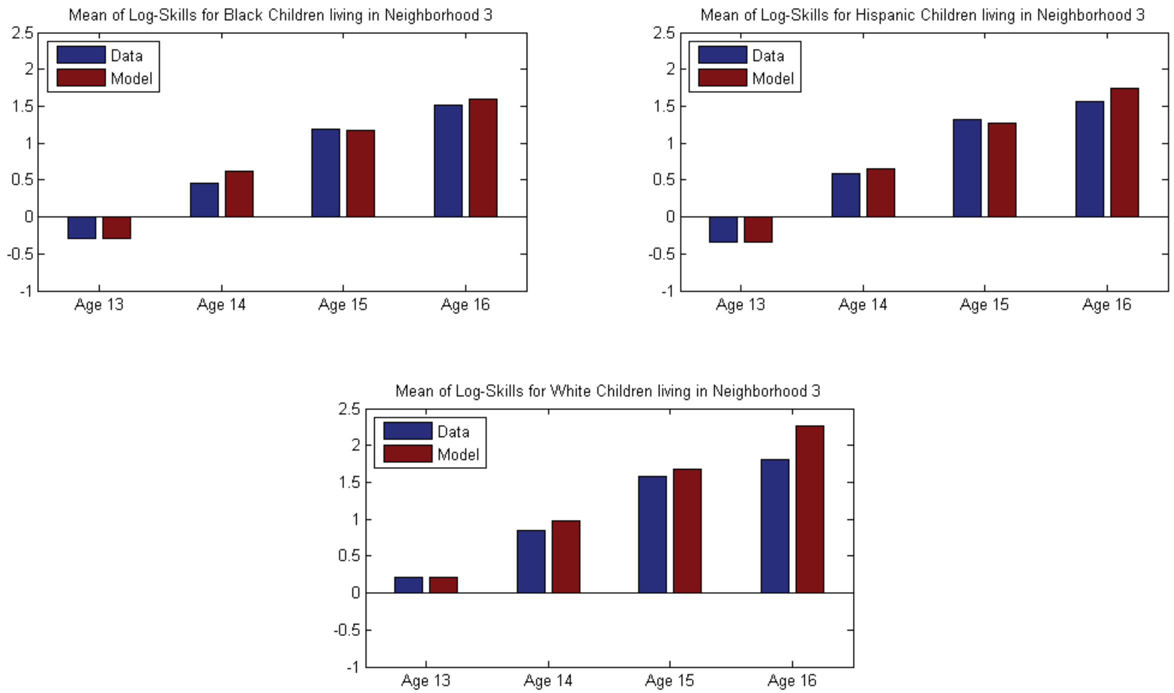
Notes: This figure shows the sample fit for the mean of children's skills for low-quality (low-income) neighborhood by race.

Figure B.3: Mean of Children Skills by Race (Medium-Income Neighborhood)



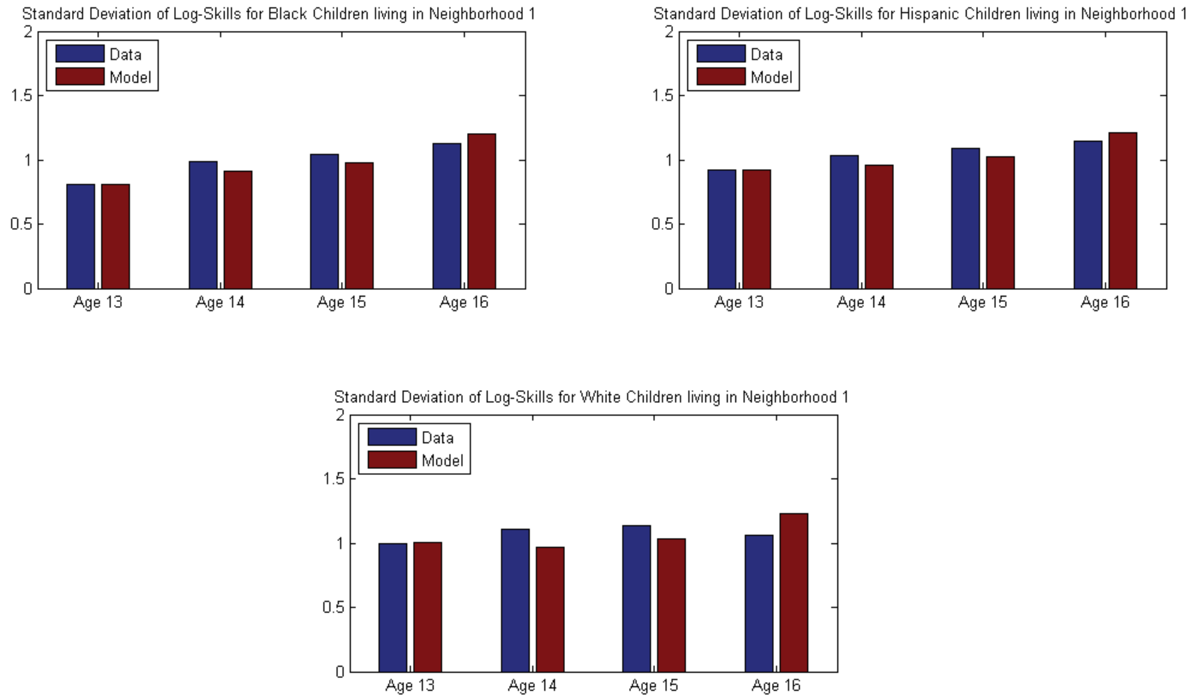
Notes: This figure shows the sample fit for the mean of children's skills for medium-quality (medium-income) neighborhood by race.

Figure B.4: Mean of Children Skills by Race (High-Income Neighborhood)



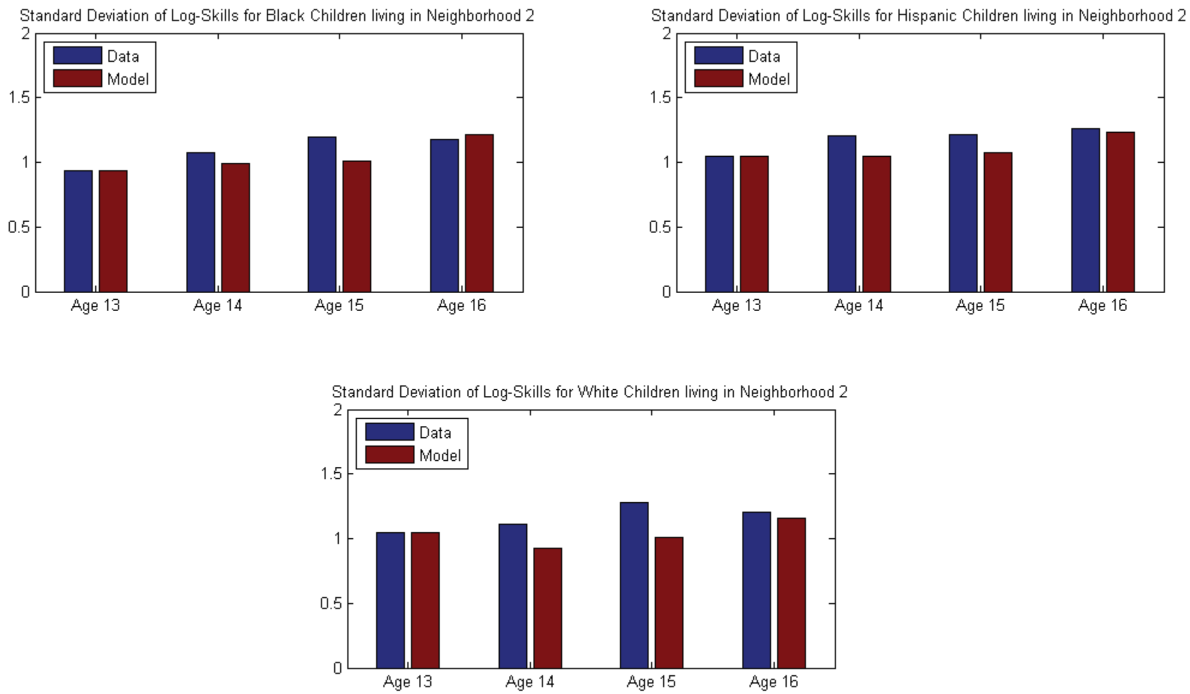
Notes: This figure shows the sample fit for the mean of children's skills for high-quality (high-income) neighborhood by race.

Figure B.5: Standard Deviation of Children Skills by Race (Low-Income Neighborhood)



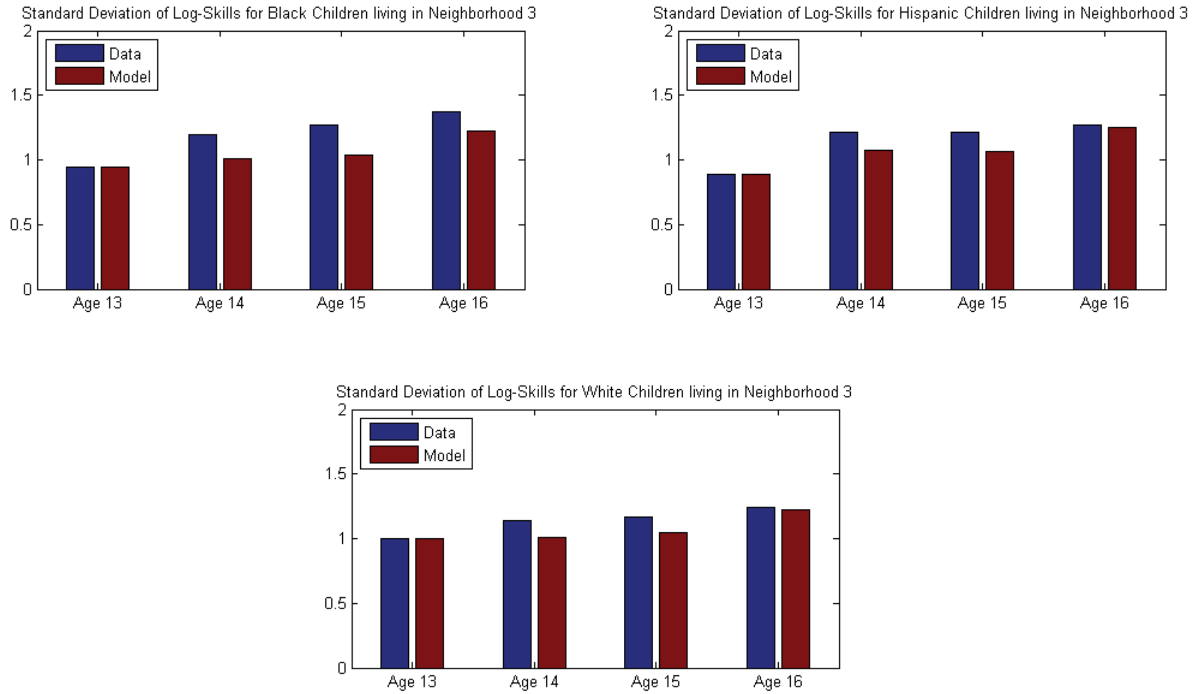
Notes: This figure shows the sample fit for the standard deviation of children's skills for low-quality (low-income) neighborhood by race.

Figure B.6: Standard Deviation of Children Skills by Race (Medium-Income Neighborhood)



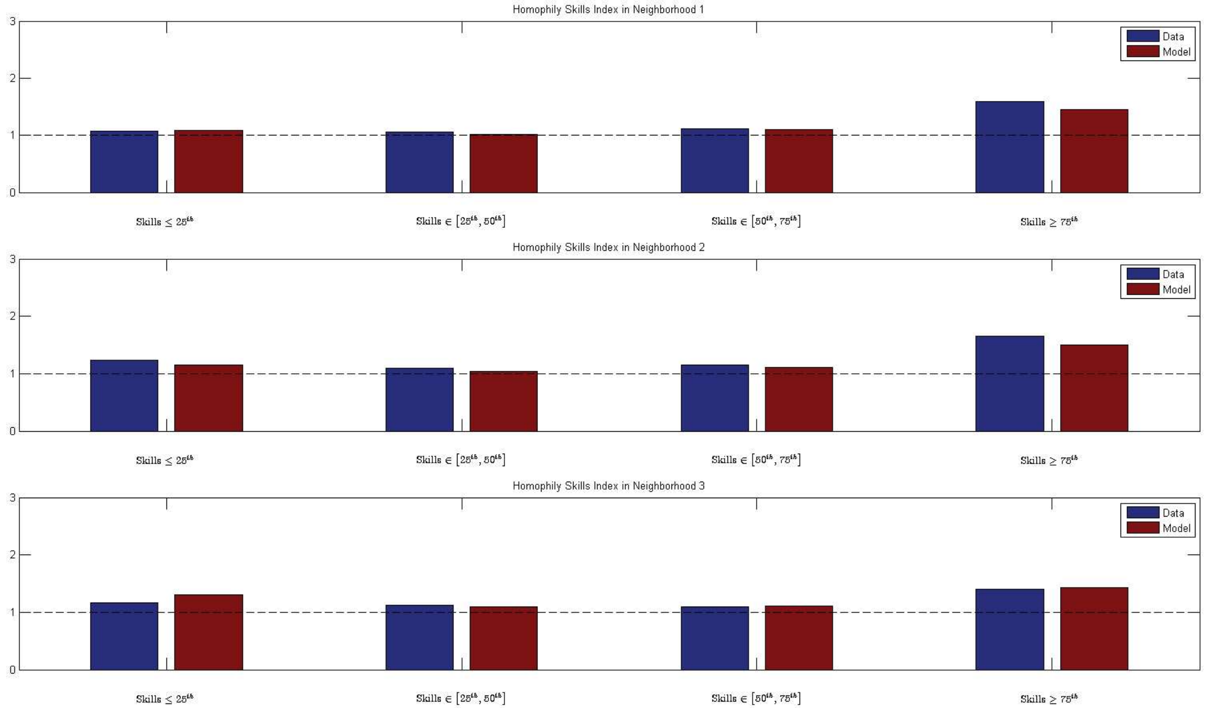
Notes: This figure shows the sample fit for the standard deviation of children's skills for medium-quality (medium-income) neighborhood by race.

Figure B.7: Standard Deviation of Children Skills by Race (High-Income Neighborhood)



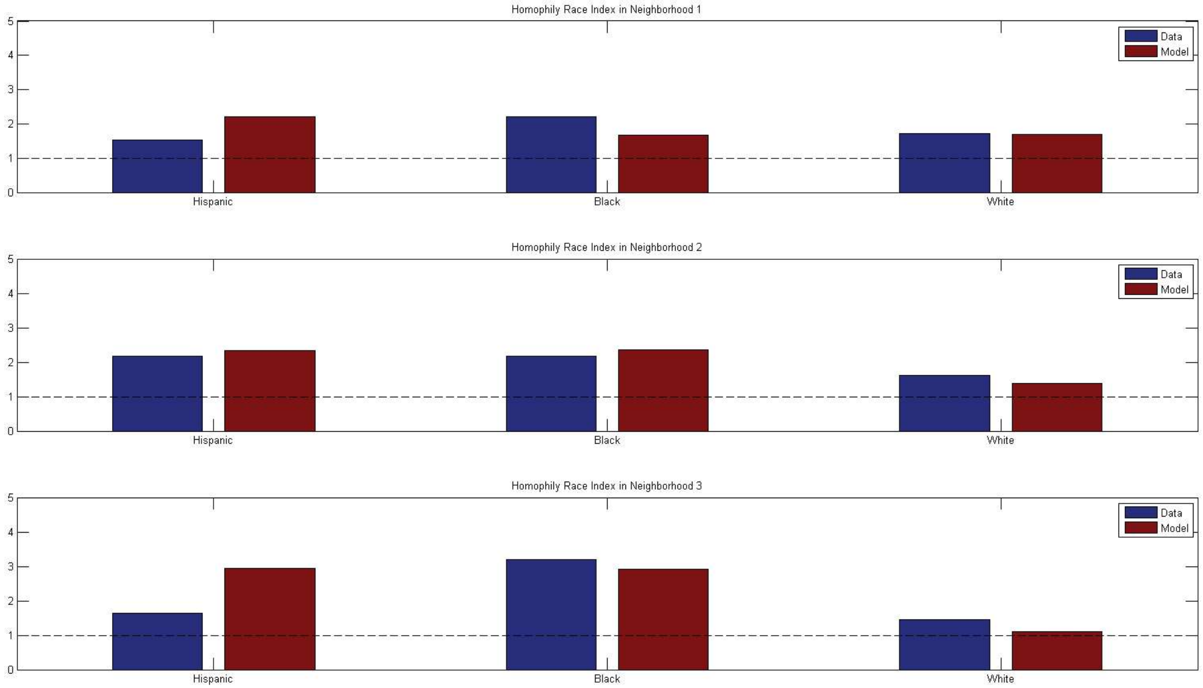
Notes: This figure shows the sample fit for the standard deviation of children's skills for high-quality (high-income) neighborhood by race.

Figure B.8: Index for Skills Homophily by Neighborhoods



Notes: This figure shows the sample fit for the homophily index for skills by by neighborhood-quality type (low-median-high income).

Figure B.9: Index for Race Homophily by Neighborhoods



Notes: This figure shows the sample fit for the homophily index for race by by neighborhood-quality type (low-median-high income).

Table B.1: Sample Statistics by Neighborhood Types

Neighborhood's Type	(1)	(2)	(3)
Panel A: Fraction of Families by Race			
Fraction Black (%)	31.74	14.00	5.36
Fraction Hispanic (%)	26.77	14.50	10.86
Fraction White (%)	41.50	71.50	83.78
Panel B: Family Income (in 1994 $)			
	Mean Family Income		
Family Income	29,637	43,458	57,512
	Mean Family Income by Race		
Black Family	26,917	41,778	46,592
Hispanic Family	26,167	33,749	49,940
White Family	34,046	45,632	59,224
Panel C: Children PPVT Achievements			
	Mean PPVT by Race		
Black Children	57.02	62.37	64.39
Hispanic Children	58.06	60.33	64.51
White Children	63.87	66.46	69.98
	PPVT Gaps between Neighborhoods		
Black Children	66 (% SD)	18 (% SD)	-
Hispanic Children	58 (% SD)	37 (% SD)	-
White Children	55 (% SD)	32 (% SD)	-
	PPVT Gaps within Neighborhoods		
Black Children	61 (% SD)	37 (% SD)	50(% SD)
Hispanic Children	52 (% SD)	55 (% SD)	49(% SD)
White Children	-	-	-

Notes: This table reports descriptive statistics by by neighborhood-quality type (low-medium-high income).

Table B.2: Sample Fit for Auxiliary Coefficients (Investments)

	Dependent Variable Fraction (%) of Invested Parental Time			
	Measurement Error Adjusted	Measurement Error Adjusted and Instrumental Variables (IV)	Measurement Error Adjusted	Measurement Error Adjusted and Instrumental Variables (IV)
	Data		Model	
Child Skills (Log)	2.660	2.120	2.152	3.124
	(0.316)	(0.668)		
	[2.041,3.280]	[0.810,3.430]		
Peers' Skills (Log)	-1.441	0.720	-1.272	0.895
	(0.650)	(0.354)		
	[-2.715,-0.167]	[0.026,1.414]		
		First Stage		First Stage
$Z_{1,i,t}$ (Minorities Children)		-0.104		-0.127
		(0.052)		
		[-0.206,-0.002]		
$Z_{2,i,t}$ (White Children)		0.082		0.105
		(0.037)		
		[0.009,0.155]		

Notes: This table shows the sample fit for auxiliary regression coefficients for models in (3.2). Columns 1 and 2 shows the same estimated coefficients as in Table 3.2. Both standard errors in parenthesis and the 95% confidence interval in square brackets are computed using a cluster bootstrap.

Data source: National Longitudinal Survey of Adolescent Health (Add Health).

Table B.3: Sample Fit for Auxiliary Coefficients (Dynamics of Skills)

	Data	Model
Investments (Log)	0.408	0.343
	(0.197)	
	[0.021,0.794]	
Child's Skills (Log)	0.750	0.760
	(0.238)	
	[0.283,1.216]	
Peer's Skills (Log)	0.366	0.167
	(0.167)	
	[0.038,0.693]	

Notes: This table shows the sample fit for auxiliary regression coefficients of the auxiliary model for the dynamics of children's skills. The dependent variable is the next period (t+1) child's log-skills. The covariates in the table are: log-investments, current child's log-skills, current peers' log-skills. The regression also include as controls: age fixed effects, race, last period family's income, mother's skills, school's fixed effects. Both standard errors in parenthesis and the 95% confidence interval in square brackets are computed using a cluster bootstrap.

Data source: National Longitudinal Survey of Adolescent Health (Add Health).

Table B.4: Counterfactual Effects on Receiving Children by Race (moved children are 5% of the receiving cohort)

Panel A: Effects on Children's Log-Skills (Mean)

	Black			Hispanic			White	
	Baseline	Counterfactual (Equilibrium)		Baseline	Counterfactual (Equilibrium)		Baseline	Counterfactual (Equilibrium)
Age 13	-0.30	-0.30	Age 13	-0.34	-0.34	Age 13	0.22	0.22
Age 14	0.28	0.23	Age 14	0.45	0.41	Age 14	0.97	0.96
Age 15	0.96	0.87	Age 15	1.08	1.03	Age 15	1.66	1.64
Age 16	1.49	1.39	Age 16	1.59	1.53	Age 16	2.24	2.21

Panel B: Effects on Parent's Investment Decision (Mean)

	Black			Hispanic			White	
	Baseline	Counterfactual (Equilibrium)		Baseline	Counterfactual (Equilibrium)		Baseline	Counterfactual (Equilibrium)
Age 13	20.09	19.01	Age 13	21.97	20.71	Age 13	27.92	27.67
Age 14	18.85	17.94	Age 14	20.67	20.25	Age 14	28.02	27.85
Age 15	16.32	16.36	Age 15	16.69	16.63	Age 15	24.29	24.37

Notes: This table shows the counterfactual policy effects for receiving children (by race) when a fraction of moved children (5% of the receiving population) are moved into a high-income environment. For each subgroup, I compare the baseline results in skills and parental investments (Panel B) with the equilibrium counterfactual predictions.

B.3 The Latent Factor Models

I consider the following latent factor model for either the child's skills ($h_{i,t}$) or the mother's skills (m_i):

$$Z^h_{i,t,k} = \mu^h_{t,k} + \lambda^h_{t,k} \cdot \ln h_{i,t} + \epsilon^h_{i,t,k}$$

$$Z^m_{i,k} = \mu^m_{t,k} + \lambda^m_k \cdot \ln m_i + \epsilon^m_{i,k} \tag{B.1}$$

where μs and λs are, respectively, the location and scale (or loading factor) parameters for each considered measure k at any age t for each child i. The distribution of the latent factors is identified by exploiting multiple measures in the data. One commonly used condition for identification is the independence of the joint distribution of latent variables and measurement errors.

Assumption 4 *Measurement model assumptions:*

(i) $\epsilon^h_{i,t,k} \perp \epsilon^h_{i,t,k'}$ *and* $\epsilon^m_{i,k} \perp \epsilon^m_{i,k'}$ *for all* t *and* $k \neq k'$

(ii) $\epsilon^h_{i,t,k} \perp \epsilon^h_{i,t',k'}$ *and* $\epsilon^m_{i,k} \perp \epsilon^m_{i,k'}$ *for all* $t \neq t'$ *and all* k *and* k'

(iii) $\epsilon^h_{i,t,k} \perp h_{i,t'}$ *and* $\epsilon^m_{i,k} \perp m_i$ *for all* t *and* t' *and all* k

(iv) $\epsilon^h_{i,t,k} \perp \epsilon^h_{j,t,k}$ *and* $\epsilon^m_{i,k} \perp \epsilon^m_{j,k}$ *for all* k *and for any family* j *different from family* i

(v) $\epsilon^h_{i,t,k}$, $\epsilon^m_{i,k} \perp X_i$, *for all* t *and all* k

Assumption 4 (i) is that measurement errors are independent contemporaneously across measures. Assumption 4 (ii) is that measurement errors are independent over time. Assumption 4 (iii) is that measurement errors in any period are independent of the latent variables in any period. Assumption 4 (iv) that measurement errors are independent between observations. Finally, assumption 4 (v) assures that errors in measuring skills are independent of the observable characteristics of children and mothers. While these assumptions are strong, they are common in the current literature (see Cunha *et al.*, 2010; Agostinelli and Wiswall, 2016a). Assumptions (i)–(v), together with the specification in (B.1), guarantee the non-parametric identification of the latent distribution. However, in this work, I consider a parametric model of skill formation, hence each of the conditions in (i)–(v) can be relaxed to statements about the zero correlation between the considered variables, instead of the full independence conditions.

B.4 Proof of Theorem 4

In this proof, I consider the two-period case. The four-period case follows by the induction hypothesis. The proof is based on backward induction. The goal is to show that for each period t, I can compute the policy functions for both parents and children that solve the fixed point associated to the equilibrium conditions in (3.4.5). For the purpose of exposition, I employ only the endogenous state variables of the problem $(h, \overline{H})$ and define the technology of skill formation in (3.4) as $f(\cdot)$. [1] Finally, the proof is executed for the case of log utility ($\gamma_3 = \gamma_5 = 1$) and perfect substitution between parents and peers ($\alpha_4 = 1$). This parametric case is the harder case to prove an existence of a fixed point through Tarski's fixed point theorem, and this is because of the non-trivial preservation of monotonicity and supermodularity in the value function. Any other case where either utility function is more convex or the technology provides higher complementarity between endogenous inputs follows by construction. Hence, by proving that this case admits a fixed point, I prove that a fixed point exists for any other admissible parameterization of the model .

- Last period case ($t = T$):

  During the last period, children decide their own peer-solving problem (3.6) based on their current level of skills. Parents observe the realization of their child's network formation and then make their last investment decision. No equilibrium conditions here are necessary, since during the next period, no children's network is formed. In other words, during the last period, the equilibrium policy functions to solve the two (parent's and child's) decision problems. The associated last period value function for parent i is:

[1] For clarity and without loss of generality, I ignore all the exogenous variables which are irrelevant for the equilibrium analysis, such as the mother's skills and the family's characteristics.

$$V_T^P(h_{i,T}, \overline{H}_{i,T}) = \left(\left(\tau - I_T^*(h_{i,T}, \overline{H}_{i,T})\right) \cdot w_{i,T} + y_{i,T}\right)^{\gamma_1} + (\gamma_2 + \alpha_1)\, log(h_{i,T})$$
$$+ \beta\, \gamma_4 log\left(\alpha_3 I_T^*(h_{i,T}, \overline{H}_{i,T}) + (1 - \alpha_3)\overline{H}_{i,T}\right) \quad \text{(B.2)}$$

Lemma 2 *The value function $V_T^P(h_{i,T}, \overline{H}_{i,T})$ is monotone, increasing in both arguments and supermodular.*

Proof: It is easy to show that the policy function is monotone in both dimensions (monotone increasing in the first argument and decreasing in the second argument). Hence, the value function clearly has monotone increasing in the first argument. Additionally, because of the homothetic preferences, the overall peer effects H on children's skills is positive, e.g., the change (decrease) of the policy function due to higher H does not dominate the initial change in H. [2] *Hence, the value function is also increasing with the peer effects. To show that the value is supermodular, consider the derivative of V_T^P with respect to $h_{i,T}$:*

$$\frac{\partial V_T^P(h_{i,T}, \overline{H}_{i,T})}{\partial h_{i,T}} \equiv \frac{\gamma_2 + \alpha_1}{h_{i,T}} \quad \text{(B.3)}$$

where equation (B.3) is derived applying the principle of optimality. Given equation (B.3) is independent of $H_{i,T}$, it follows trivially that V_T^P is supermodular in $(h_{i,T}, \overline{H}_{i,T})$.

- First period case ($t = T - 1$):

In order to solve the parent's problem and child's problem differently (remember that parents take a child's decision regarding the next period's network formation as a given), let us define $\widetilde{h}_{i,t+1}$ as the skills that children care about once

[2] This can be proved through the comparative statics of the problem, using the concavity of utility over consumption and children's skills.

they decide upon their friends in the next period. Parents take $\widetilde{h}_{i,t+1}$ as a given – that is, they think the process of the formation of peer groups is independent of their investment decisions. In this case, the parent's problem is:

$$V^P_{T-1}(h_{i,T-1}, \overline{H}_{i,T-1}\,;\ \widetilde{h}_{i,T}) =$$

$$\max_{I_{i,T-1}\in[0\,,\,\tau]} \;((\tau - I_{i,T-1})\cdot w_{i,T-1} + y_{i,T-1})^{\gamma_1} + \gamma_2\, h_{i,T-1} +$$

$$\beta\, E\left[V^P_T\left(h_{i,T}\,(I_{i,T-1})\,, \overline{H}_{i,T}(\widetilde{h}_{i,T})\right)\middle|\widetilde{h}_{i,T}\right] \tag{B.4}$$

where $\overline{H}_{i,T}(\widetilde{h}_{i,T})$ is the stochastic mapping about the next period's peer effects $\overline{H}_{i,T}$ given $\widetilde{h}_{i,T}$. Given the empirical evidence on children's social interactions, I consider the case when this mapping is stochastically ordered in $\widetilde{h}_{i,T}$,

$$E\left[\overline{H}_{i,T}(\widetilde{h}_{i,T})\middle|\widetilde{h}_{i,T}\right] \leq E\left[\overline{H}_{i,T}(\widetilde{h}_{i,T}\text{'})\middle|\widetilde{h}_{i,T}\text{'}\right] \;\text{ if }\; \widetilde{h}_{i,T} < \widetilde{h}_{i,T}\text{'},$$

which means that the higher the child's skills, the higher the probability of becoming friends with higher-skilled children. The fix-point problem here comes from the equilibrium consistency conditions, which require that the endogenous network formation is consistent with the parental decisions about the child's development. In other words, in the equilibrium path, I am imposing that the optimal level of children's skills decided by the parents are the same level of skills governing the network-formation decision

$$\widetilde{h}_{i,T} = h^*_{i,T}.$$

To show that this fix point has a solution, I am applying standard results in the dynamic lattice programming literature.

Lemma 3 *Under Lemma 2, there exists a policy function $I_{T-1}(\cdot)$ which solves the equilibrium fix-point problem in (B.4).*

Proof: This result follows directly from Tarski's fixed point theorem. The continuation value is both increasing and supermodular in the two endogenous variables (individual endogenous children's skills and peers' skills). The stochastic process governing the network formation stochastically ascends with respect to the children's skills, and the choice set is a complete lattice.

The Markovian equilibrium of the model is defined as the sequence of the policy functions solves each of the two periods' equilibrium problems.

B.5 Indirect Inference

Part of the selected moments in the estimation procedure includes coefficients of auxiliary regressions. In particular, I consider separate auxiliary models to analyze the *parental investments* and the *dynamics of skills*. The first set of coefficients consider parental investments as the dependent variable in the following linear regression

$$I_{i,s,t} = \beta_0 + \beta_1 \ln h_{i,s,t} + \beta_2 \ln \overline{H}_{i,s,t} + X_i'\beta_3 + \beta_s + u_{i,s,t} \; , \tag{B.5}$$

where $I_{i,s,t}$ is the parental investment (as fraction of time) for parent of child i, in school s when she is t years old, which is recovered through a latent factor model (see Section 3.5.1) using data on parental engagement described in previous section. The child's skills are defined as $h_{i,s,t}$, while $\overline{H}_{i,s,t}$ is the mean of her peers' skills. X_i is a vector of child and parents' exogenous characteristics, which includes race, age fixed-effects, lagged family income and mother's skills. Finally, β_s is the school fixed effects. The coefficients of interests are related to how parental investments respond to changes in child skills (β_1) and peers' skills (β_2). In the estimation procedure, I include as targeted moments the parameters in (B.5), for both the school fixed-effects estimator, as well as for the IV estimator explained in Section 3.3.4.

Additionally, I consider an auxiliary model of skill dynamics, where the next period skills are the dependent variable of the following regression

$$\ln h_{i,s,t+1} = \beta_0^h + \beta_1^h \ln h_{i,s,t} + \beta_2^h \ln \overline{H}_{i,s,t} + \beta_3^h \ln I_{i,s,t} + X_i'\beta_4^h + \beta_s^h + u_{i,s,t}^h \; , \tag{B.6}$$

where $(\beta_0^h, \beta_1^h, \beta_2^h, \beta_3^h)$ are the specific auxiliary parameters I selected as targeting moments for estimation. They represent respectively the elasticity of the next period skills with respect to the stock of current skills, the peers' skills and the parental investments. X_i is a vector of child and parents' exogenous characteristics, which includes race, age fixed-effects, lagged family income and mother's skills. Finally, β_s is the school fixed effects. The estimated distribution of school fixed-effects are used in the estimation as the school quality distributions in the model.

Chapter C

SUPPLEMENTARY MATERIAL FOR CHAPTER 3

C.1 Additional Tables

Table C.1: First Stage Estimates – Full Set of Individual Controls

	Combined Math-Reading		Behavior Problems Index	
	ΔIncome	ΔHours Worked	ΔIncome	ΔHours Worked
	(1)	(2)	(3)	(4)
ΔEITC	1.026**	1.481***	1.101**	1.488***
	(0.488)	(0.282)	(0.482)	(0.280)
LabDemShocks	1.659***	0.322*	2.067***	0.245
	(0.395)	(0.186)	(0.405)	(0.178)
Male	0.185	-0.006	0.134	-0.012
	(0.279)	(0.119)	(0.288)	(0.110)
Age	-0.155**	-0.007	-0.109**	-0.020
	(0.064)	(0.028)	(0.052)	(0.024)
No siblings	0.053	0.024	-0.181	0.045
	(0.533)	(0.240)	(0.481)	(0.212)
Two or more sibling	0.079	-0.070	0.128	0.021
	(0.397)	(0.163)	(0.406)	(0.152)
Black	-2.728***	-0.441**	-2.624***	-0.393**
	(0.447)	(0.180)	(0.417)	(0.171)
Hispanic	-2.087***	-0.342*	-1.782***	-0.312*
	(0.525)	(0.205)	(0.522)	(0.189)
SW Chi-sq. (Under id)	13.21	14.40	21.89	20.57
P-value	0.00	0.00	0.00	0.00
SW F (Weak id)	13.19	14.38	21.86	20.54
P-value	0.00	0.00	0.00	0.00
KP (Weak id)	6.42	6.42	10.43	10.43
Observations	12,288	12,288	13,777	13,777

Notes: This table shows the estimates for both our first stage models. Dependent variable: Δ Income (columns 1 and 3), and ΔHours worked (columns 2 and 4). Columns (1) and (2) refer to the estimating sample used for the analysis of child cognitive development (combined Math-Reading test score). Columns (3) and (4) consider the estimating sample used for the analysis of child behavioral development (Behavior Problems Index, BPI). For each analysis, the two endogenous variables are: changes in income (ΔIncome) and changes in maternal hours worked (ΔHours). The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks (*LabDemShocks*). Income and the EITC are measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). One sibling is the reference category for child's number of siblings. White is the reference category for child's race. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table C.2: Changes in EITC Schedule, Labor Demand Shocks, and Spouse Labor Supply

	Combined Math-Reading	Behavior Problems Index
	ΔHours Worked Spouse	ΔHours Worked Spouse
	(1)	(2)
ΔEITC	0.402	0.788
	(0.661)	(0.644)
LabDemShocks	0.166	0.098
	(0.204)	(0.192)
Observations	7,726	8,845

Notes: This table shows the estimates for our analysis of changes in spouse labor supply. Dependent variable: ΔHours worked by the spouse. Column (1) refers to the estimating sample used for the analysis of child cognitive development (combined Math-Reading test score). Column (2) considers the estimating sample used for the analysis of child behavioral development (Behavior Problems Index, BPI). The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks ($LabDemShocks$). Income and the EITC are measured in $1,000 of year 2000 dollars. Hours worked by the spouse are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table C.3: Income, Hours Worked, and Child Test Scores – Full Set of Individual Controls

	Combined Math-Reading	
	OLS	IV
	(1)	(2)
ΔIncome	0.001*	0.044***
	(0.000)	(0.015)
ΔHours worked	0.000	-0.060**
	(0.001)	(0.024)
Male	0.024**	0.017
	(0.010)	(0.017)
Age	0.001	0.008*
	(0.003)	(0.005)
No siblings	-0.001	-0.006
	(0.020)	(0.032)
Two or more sibling	-0.026**	-0.028
	(0.012)	(0.022)
Black	-0.156***	-0.057
	(0.014)	(0.041)
Hispanic	-0.076***	-0.009
	(0.016)	(0.035)
Observations	12,288	12,288

Notes: This table shows the estimates for our analysis of child cognitive development. Dependent variable: Combined Math-Reading test score. Column (1) reports the OLS estimates. Column (2) shows the IV estimates. The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks (*LabDemShocks*). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). One sibling is the reference category for child's number of siblings. White is the reference category for child's race. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table C.4: Income, Hours Worked, and Child Behavior – Full Set of Individual Controls

	Behavior Problems Index	
	OLS	IV
	(1)	(2)
ΔIncome	0.000	0.013
	(0.000)	(0.009)
ΔHours worked	-0.001	-0.052**
	(0.001)	(0.022)
Male	-0.016	-0.018
	(0.011)	(0.013)
Age	0.010***	0.011***
	(0.003)	(0.003)
No siblings	0.026	0.027
	(0.020)	(0.023)
Two or more sibling	0.002	0.005
	(0.013)	(0.015)
Black	-0.008	0.010
	(0.015)	(0.028)
Hispanic	0.023	0.031
	(0.016)	(0.022)
Observations	13,777	13,777

Notes: This table shows the estimates for our analysis of child behavioral development. Dependent variable: Behavior Problems Index (BPI). Column (1) reports the OLS estimates. Column (2) shows the IV estimates. The two instrumental variables are: changes in EITC benefits (ΔEITC) and labor demand shocks (*LabDemShocks*). Income is measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). One sibling is the reference category for child's number of siblings. White is the reference category for child's race. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table C.5: Baseline Estimates with State Trends

	Combined Math-Reading		Behavior Problems Index	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
ΔIncome	0.001*	0.041***	0.000	0.008
	(0.000)	(0.010)	(0.000)	(0.006)
ΔHours worked	0.000	-0.056**	-0.001	-0.049**
	(0.001)	(0.022)	(0.001)	(0.020)
First Stage Tests (Income/Hours):				
SW Chi-sq. (Under id)		23.54/19.32		40.98/25.72
P-value		0.00/0.00		0.00/0.00
SW F (Weak id)		23.41/19.21		40.77/25.60
P-value		0.00/0.00		0.00/0.00
KP (Weak id)		10.30		15.38
Observations	12,288	12,288	13,777	13,777

Notes: This table shows the estimates for the analysis of cognitive and behavioral development with a state fixed effects specification. Dependent variable: Combined Math-Reading test score (columns 1–2), and Behavior Problems Index (BPI) (columns 3–4). Income and EITC are measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include state fixed effects and a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include state fixed effects, as well as controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table C.6: Baseline Estimates with Controls for School Financial and Economic Resources

	Combined Math-Reading		Behavior Problems Index	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
ΔIncome	0.001*	0.042***	0.000	0.012
	(0.000)	(0.014)	(0.000)	(0.009)
ΔHours worked	0.000	-0.057**	-0.001	-0.051**
	(0.001)	(0.023)	(0.001)	(0.022)
ΔTotal revenues (per pupil)	0.003	-0.002	0.002	0.006
	(0.012)	(0.020)	(0.013)	(0.015)
ΔTotal public expenditure (per pupil)	0.012	-0.016	0.021	-0.006
	(0.022)	(0.044)	(0.023)	(0.032)
First Stage Tests (Income/Hours):				
SW Chi-sq. (Under id)		14.58/15.60		23.45/21.03
P-value		0.00/0.00		0.00/0.00
SW F (Weak id)		14.56/15.58		23.41/21.00
P-value		0.00/0.00		0.00/0.00
KP (Weak id)		7.08		11.05
Observations	12,255	12,255	13,735	13,735

Notes: This table shows the estimates for the analysis of cognitive and behavioral development when we control for per-pupil school resources by state. Dependent variable: Combined Math-Reading test score (columns 1–2), and Behavior Problems Index (BPI) (columns 3–4). Family income, the EITC, the total revenues per pupil, and the total expenditure per pupil are measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. The total revenues per pupil are the total revenues from all sources divided by the fall membership as reported in the state finance file. The total current expenditure per pupil is the total current expenditure for public elementary and secondary education divided by the fall membership as reported in the state financial file. Data about revenues and expenditures are from the CDD National Public Education Financial Survey. All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table C.7: Common Pre-trends between Labor Demand Shocks and Child Development

	Combined Math-Reading (1)	Behavior Problems Index (2)
LabDemShocks $(t+1)$		
F-stat.	6.11	0.12
P-value	0.01	0.73
LabDemShocks $(t+2)$		
F-stat.	0.70	0.38
P-value	0.40	0.54
LabDemShocks $(t+3)$		
F-stat.	1.35	0.61
P-value	0.25	0.43
LabDemShocks $(t+1)$,$(t+2)$		
F-stat.	0.47	0.23
P-value	0.63	0.80
LabDemShocks $(t+1)$,$(t+2)$,$(t+3)$		
F-stat.	0.81	1.50
P-value	0.49	0.21

Notes: This table is based on the analysis of the effect of future labor demand shocks on current cognitive and behavioral development. The table shows the F-statistic and the relative significance of the coefficients for future labor demand shocks. In cases with multiple variables for future labor demand shocks, we jointly test the significance of labor demand shocks. Dependent variable: Combined Math-Reading test score (column 1), and Behavior Problems Index (BPI) (column 3). Each specification contains controls for EITC benefits (ΔEITC) and labor demand shocks (*LabDemShocks*). In addition, each model also contains variables for future labor demand shocks as explained in each panel header. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table C.8: Baseline Estimates Excluding Movers Across States

	Combined Math-Reading		Behavior Problems Index	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
ΔIncome	0.001**	0.052***	0.000	0.010
	(0.000)	(0.020)	(0.000)	(0.010)
ΔHours worked	0.000	-0.069**	-0.000	-0.053**
	(0.001)	(0.030)	(0.001)	(0.024)
Observations	11,707	11,707	13,087	13,087

Notes: This table shows the estimates for the analysis of cognitive and behavioral development once we exclude observations with changes in state of residence in two consecutive periods. Dependent variable: Combined Math-Reading test score (columns 1–2), and Behavior Problems Index (BPI) (columns 3–4). Income and EITC are measured in $1,000 of year 2000 dollars. Hours worked are yearly hours and expressed in hundreds. All models include a third order Taylor polynomial expansion of predicted income as a control function (see equation 4.7). All models also include controls for child's age, gender, race, and number of siblings. Standard errors are clustered at the family level and reported in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Table C.9: Time Allocation to Child Care, Mother's Hours Worked, and Family Income

	Physical Child Care (1)	Help with Homework (2)	Read & Play (3)	Other Child Care (4)	Total Child Care (5)
Income	-0.001	0.003	-0.001	0.002	0.002
	(0.002)	(0.002)	(0.003)	(0.004)	(0.006)
Hours worked (per week)	-0.007**	-0.014***	-0.011***	-0.028***	-0.060***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.008)
Observations	3,183	3,183	3,183	3,183	3,183

Notes: This table shows the OLS estimates for the analysis of parental time investment in the child using data from the American Time Use Survey (ATUS) and the American Heritage Time Use Survey (AHTUS). Dependent variable: Physical Child Care (column 1), Help with Homework (column 2), Read and Play (column 3), Other Child Care (column 4), and Total Child Care (column 5). Time investment is measured in hours per week. Income is measured in $1,000 of year 2003 dollars. Hours worked are weekly hours worked. All models include controls for single-head household, mother's age, child's age, mother's education, number of siblings. All models also include year fixed effects. Standard errors are in parentheses. *, **, *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

C.2 The Child Development Supplement

In Table C.10 we show the variables construction process used to analyze the Child Development Supplement (CDS). We focus on the first wave of the CDS, the so-called CDS-I, collected in 1997.

Table C.10: CDS – Variables Construction

	Original Definition (1)	Original Answers (2)	Variable Definition (3)
Encourage hobbies	Family encourages hobbies	Yes, No	Yes=1
Physical affection	Shown physical affection (times past week)	1-350	1-350
Parenting class pre-birth	Take parenting classes before child's birth	Yes, No	Yes=1
Parenting class	Never take parenting classes	Yes, No	No=1
Never cared by others	Child's age when first cared by others	0-10	Never=1
Use of rules	Family with lots of rules or not very many rules	Lots, Not many	Lots=1
How often...			
Control who the child is with	Control which children your child spend time with	N, S, SM, O, VO	O, VO=1
Control activities after school	Control how child spends time after school	N, S, SM, O, VO	O, VO=1
Set homework time	Set a time for homework	N, S, SM, O, VO	O, VO=1
Reaction to grades lower than expected:			
Contact faculty	Contact teacher/principal	U, SU, NS, SL, L	SL, L=1
Closer eye on activities	Closer eye on child's activities	U, SU, NS, SL, L	L=1
Lecture a child	Lecture the child	U, SU, NS, SL, L	SL, L=1
Tell child to work harder	Tell the child to spend more time on homework	U, SU, NS, SL, L	L=1
Help with schoolwork	Increase time helping the child with schoolwork	U, SU, NS, SL, L	L=1
Full home	Full home scale	7-27	7-27
Cognitive stimulation	Cognitive stimulation subscale	2-14	2-14
Emotional support	Emotional support subscale	2-14	2-14
Parental warmth	Parental warmth subscale	1-5	1-5
Time diaries (in seconds)			
School	Student, attending classes	0-86,400	0-86,400
TV	TV use	0-86,400	0-86,400
Electronic games	Electronic video games use	0-86,400	0-86,400
Art, sculpture	Art, arts and crafts,	0-86,400	0-86,400
Books	Reading or looking at books	0-86,400	0-86,400
Visiting others, socializing	Socializing with people outside own household	0-86,400	0-86,400

Note: This table shows variable definitions from the CDS-I data set used in Section 4.5.4. In the table the following abbreviations are used: (i) N: Never, S: Seldom, SM: Sometimes, O: Often, VO: Very often; (ii) U: Not at all likely, SU: Somewhat unlikely, NS: Not sure how likely, SL: Somewhat likely, L: likely. Refer to the text and the CDS-I User Guide Supplement for further details about the original and the constructed variables.