Policies for Early Childhood Skills Formation: Accounting for Parental Choices and Noncognitive Skills

Iacopo Morchio*†

April 2018

Abstract

What are the returns in terms of children's skills development to child allowance policies? Answering this question requires a theory of the tradeoffs faced by households, as well as a realistic technology of skills formation. I build a model of parental choices which embeds the technology of cognitive and noncognitive skills formation estimated by Cunha, Heckman and Schennach (2010). The features of the technology, combined with the model, help to account quantitatively for important empirical regularities on skills formation and time allocation choices of households. Accounting for noncognitive skills implies higher policy returns than previously estimated in the literature.

Keywords: Childcare, Cognitive Skills, Noncognitive Skills, Inequality, Household Production, Time Use.

JEL Classification Numbers: D13, J13, J24.

*University of Vienna. Email: iacopo.morchio@univie.ac.at

†I am particularly indebted to Andrés Erosa, Javier Fernández Blanco and Juan José Dolado for their advice on an early version of this paper. I also benefited from comments by Pedro Gomes, Nezih Guner, Matthias Kredler, Christopher Rauh, José Victor Rios-Rull, Loris Rubini, Manuel Santos, Kjetil Storesletten, Ludo Visschers Yikai Wang and participants to the Madrid Macro Workshop 2017 and ESEM 2017. All errors are my own.

1 Introduction

Heterogeneity at age 20 has been shown to be one of the most important determinants of lifetime inequality.[1] Much of this heterogeneity builds up during childhood, which is known to be a crucial phase for skills development.[2] A large empirical literature estimates the returns of different policy experiments in terms of improvements in the skills of children. Most of these programs were focused on a specific subgroup, or on a small number of children. Much less is known about the returns of widespread policy interventions, for instance what the impact of introducing a universal child allowance in the US would be, or what is the optimal way to distribute such allowance according to observable characteristics, such as income or the age of a child.

Answering this question requires a theory of the tradeoffs faced by households, as well as a realistic technology of skills formation. Also, the answer depends crucially on the features of the skills formation process. The recent empirical literature on the technology of skills formation[3] emphasizes the importance of accounting for multiple skills in order to correctly estimate the returns to parental investment. In particular, accounting for noncognitive skills and their feedback to cognitive skills has been shown to be key. On the other hand, the structural literature has emphasized the importance of the tradeoffs faced by households who invest time and resources in their offspring.[4] The goal of this paper is to combine a technology of skills formation that accounts for both cognitive and noncognitive skills with a model in which households face these tradeoffs.

To this end, I introduce the estimates of the technology of cognitive and noncognitive skills formation proposed by Cunha, Heckman and Schennach (2010) in an heterogeneous agents decision theoretic model of parental investment choices and skills development, to account for the endogenous response of parents to changes in policies. The skills formation technology specifies a relationship such that the future skills of a child are a function of the

[1]See Huggett, Ventura and Yaron (2011); Lee and Seshadri (forthcoming); Guvenen and Kuruscu (2009); Keane and Wolpin (1996).

[2]The empirical evidence dates back to the Perry Preschool Project (1962) and the Coleman Report (1966); see for instance Heckman, Malofeeva, Pinto and Savelyev (2010b) and Heckman, Moon, Pinto, Savelyev and Yavitz (2010a), and also the Head Start and Early Head Start programs.

[3]See Cunha and Heckman (2007); Cunha, Heckman and Schennach (2010); Helmers and Patnam (2011). See also Agostinelli and Wiswall (2017) for recent developments.

[4]See Bernal and Keane (2010) (2011); Del Boca, Flinn and Wiswall (2014); Griffen (accepted); Brilli (2012); Youderian (2016); Yum (2018); Daruich (2017); Lee and Seshadri (forthcoming)

child's own current skill endowments, parental skills and parental investment. I take the parametrization of the technology of skills formation from the results of the paper by Cunha et al. (2010), while I estimate parental preferences and a production function of parental investment using data on time use and skills development from the US. What differentiates this paper from other structural and reduced-form work is the combination of a structural model with a careful treatment of the process of both cognitive and noncognitive skills formation. I show that accounting for noncognitive skills is crucial, as it implies three times higher policy returns than if they are neglected.

In the model, households are heterogeneous in wages, cognitive and noncognitive skills, and each household has one offspring who draws initial cognitive and noncognitive skills at birth. In each period, parents have to choose how much to consume and work, and how much time and money to spend in developing their offspring's skills. There are two key tradeoffs: one between child care time, work and leisure, the other between goods invested in the offspring and consumption. Time and goods combine to form parental investment. Finally, the offspring's skills are also subject to random shocks. As a result, the model generates heterogeneity in investment across parents, determined by joint heterogeneity in all initial conditions and luck.

I calibrate the model using data from the American Time Use Survey (ATUS) and the Children of the National Longitudinal Survey of the Youth (CNLSY/79). I find that the model replicates the intergenerational correlations in test scores observed in the data well, and that it matches the positive relationship between education and child care time found in the data. This result follows from the parametrization of the technology of skills formation, which implies that more educated parents are more productive at raising skillful children. Despite the fact that they face higher wages, their higher productivity offsets the higher opportunity cost of time. Consequently, more educated parents invest more time in their offspring's skills.

Taking the parametrization of the technology of skills formation as given allows me to focus on estimating parental preferences and the production function of parental investment, while reducing the degrees of freedom of the model. Using this technology also allows me to account for both cognitive and noncognitive skills, of both parents and offspring. However, this level of detail comes at a cost: in order to remain tractable, I have to abstract from the dynamic nature of asset accumulation. I still allow households to be heterogeneous in income, but I abstract from idiosyncratic productivity shocks and borrowing constraints as sources of inequality in parental investment. I argue that this implies that the estimates of returns to policy that I simulate are a lower bound of true, "underlying" policy returns. I discuss this in more detail in the model subsection.

Finally, I use the model to simulate the impact of several policies. First, I find that the introduction of a universal child allowance policy worth approximately 5% of median household income in all periods (inspired by the German Kindergeld scheme) increases cognitive skills at age 14 by 3% of a standard deviation and noncognitive skills at age 14 by more than 3.5%. This increase is higher for low income households and households where parents have higher cognitive skills. The increase in skills is driven by an increase in time invested by parents in child care. Even though more than 90% of the transfer is consumed, it allows parents to reduce labor supply and increase time invested in the offspring. Interestingly, the transfer can reduce the intergenerational correlation of income, but does not influence the intergenerational correlation of skills. This is because while the transfer can help parents who earn low wages, it cannot change the fact that more skilled parents are still more productive at raising skillful children.

A natural question is whether a government can do better than with a universal allowance, by making transfers depend on observable characteristics. To answer this question, I compute several optimal policies (in an average skill-maximizing sense) that are allowed to depend on the age of the child and the income of parents, and find that they substantially improve over a flat transfer, targeting more effectively households with younger children (whose development is most affected by parental investment) and poor households (which exhibit higher returns to policy changes). I find the best of this class of policies to be a scheme that directly subsidizes expenditure in children's skills development: such policy increases average cognitive skills by more than 6% of a standard deviation and average noncognitive skills by 9%. I find that constrained optimal policies that depend only on the age of the child and household income can produce three-quarters of the gains generated by more flexible policies that are allowed to vary with parental and the offspring's skills as well.

As mentioned above, I find that accounting for noncognitive skills increases dramatically the impact of policies. In a counterfactual exercise, I simulate the impact of the same policies in a model featuring only cognitive skills. The restricted model implies that all policies have a negligible impact on children's skills, which increase by about one-fourth of what the two-skills model implies. This result demostrates the importance of accounting for noncognitive skills to correctly estimate policy returns.[5]

This paper contributes to the literature that builds models for the analysis of policies designed to influence and promote early skills formation. In particular, this paper bridges the prevalently empirical literature that estimates the technology of skills formation (Cunha and Heckman 2007; Cunha, Heckman and Schennach 2010; Helmers and Patnam 2011; Todd and Wolpin 2007; Hanushek and Woessmann 2008.) with the structural literature that uses models to estimate early childhood policy returns (see Bernal and Keane 2010 2011; Caucutt and Lochner 2012; Del Boca, Flinn and Wiswall 2014; Griffen accepted; Brilli 2012; Youderian 2016; Yum 2018; Daruich 2017; Lee and Seshadri forthcoming). None of the structural papers cited features noncognitive skills. I choose the technology estimated in Cunha et al. (2010) in order to account for both skills, which the authors show to be important for correctly estimating the returns to investment. Conversely, I show that introducing noncognitive skills in a model of parental choices substantially increases returns to policies. When I account only for cognitive skills, my results are compatible with Del Boca et al. (2014), who find that policies are ineffective at influencing skills development.

The papers by Daruich (2017), Lee and Seshadri (forthcoming), Youderian (2016), Yum (2018) develop macroeconomic models of human capital formation to understand how policies would influence the accumulation of human capital and the intergenerational persistence of earnings across generations. Rather than focusing on general human capital, this paper focuses on a subset of human capital accumulation (cognitive and noncognitive skills) at the advantage of having a directly observable data equivalent for skills and a more data-driven technology of skills formation.

The paper is organized as follows. Section 2 describes the technology of skills formation and outlines the model. Section 3 discusses the data used and the identification of the model. Section 4 discusses quantitative results and external validation. Section 5 outlines the policy experiments and describes their impact. Section 6 concludes.

2 The model

The model features four key ingredients: the technology of skills formation, the investment formation technology, time-allocation choices of households and ex-ante heterogeneity in

[5] A similar result is mirrored in Cunha et al. (2010), who perform this experiment in a reduced-form fashion rather than in a model of parental investment.

skills for parents and offsprings.

Since several results are a direct consequence of returns to investment, the next subsection summarizes the main features of the technology of skills formation estimated by Cunha, Heckman and Schennach (2010) (CHS from now on) and the main mechanisms behind differentials in returns.

2.1 The technology of skills formation

CHS estimate the technology of skills formation assuming there exist two different developmental stages, $j = \{1, 2\}$, which correspond to *early childhood* (ages 0-6) and *later childhood* (ages 7-14) respectively. Early human capital of the child is assumed to be a two dimensional, time varying vector of skills; the latter are of type $k = \{C, N\}$, respectively cognitive and noncognitive skills.

Parents provide three different inputs for child development: their cognitive and noncognitive skills $s_{C,P}, s_{N,P}$, which are assumed to be time-invariant, and parental investment I_t. Parental skills are assumed to be those of the mother.

The technology of skills formation has the common Constant Elasticity of Substitution form

$$s_{k,t+1} = [\gamma_{j,k,1} s_{C,t}^{\phi_{j,k}} + \gamma_{j,k,2} s_{N,t}^{\phi_{j,k}} + \gamma_{j,k,3} I_t^{\phi_{j,k}} + \gamma_{j,k,4} s_{C,P}^{\phi_{j,k}} + \gamma_{j,k,5} s_{N,P}^{\phi_{j,k}}]^{1/\phi_{j,k}} , \qquad (1)$$

which states that next period's skills $s_{k,t+1}$ are a function of investment I_t, offspring's cognitive and noncognitive skills $\{s_{C,t}, s_{N,t}\}$ at time t and parental cognitive and noncognitive skills $\{s_{C,P}, s_{N,P}\}$. Notice that there are two stages j but many time periods t, which belong to one of the two stages. In the work of CHS, periods t are two years long and stages 1 and 2 correspond to ages 0-6 and 7-14, respectively. All parameters $\gamma_{j,k,i}$ and $\phi_{j,k}$ vary across developmental stages $j = \{1, 2\}$ and across skills $k = \{C, N\}$. The parameter $\phi_{j,k} \in (-\infty, 1]$ is crucial, because it determines the elasticity of substitution $1/(1 - \phi_{j,k})$ between inputs.

Cunha, Heckman and Schennach estimate the technology of skills formation under a number of alternative assumptions (household-specific heterogeneity and endogeneity of investment); their findings are robust to the alternatives. I summarize below the findings that drive results in the present paper.

1. **Self-Productivity**: skills exhibit self-productivity in the sense that $\gamma_{j,C,1} > 0, \gamma_{j,N,2} > 0$ for $j = \{1, 2\}$; higher initial skills lead on average to higher skills later on. Also,

early investment produces long-lasting effects because increasing skills at the beginning affects all the subsequent skill development.

2. **Cross-Productivity**: skills positively contribute to each other, in the sense that $\gamma_{j,C,2} > 0, \gamma_{j,N,1} > 0$ for $j = \{1, 2\}$. Higher cognitive skills increase noncognitive skills, and viceversa.

3. **Efficiency**: in the first stage, investment is more productive than in the second stage, for both cognitive and noncognitive skills; that is, $\gamma_{1,k,3} > \gamma_{2,k,3}$ for $k = \{C, N\}$.

4. **Complementarity**: in the first stage of cognitive skills development, the elasticity of substitution between inputs is roughly four times larger than in the second stage; this means that, during early childhood, parental investment can make up for adverse initial conditions (i.e. below-median initial cognitive endowments) and for low parental skills. During later childhood, instead, inputs become strongly complementary, so that increasing skills in this phase becomes extremely costly. Noncognitive skills, instead, exhibit roughly the same elasticity of substitution across stages.

The features of the technology, along with its estimated parametrization (available in Table XII in the Appendix), produce a number of derived results that give insights on how parental investment should behave if households knew the technology of skills formation. First of all, in the first stage it is easier to increase cognitive skills; the amount of investment required to increase skills by 1 % of a standard deviation is lower in the first stage with respect to the second, as Figure 1 shows. Noncognitive skills, instead, do not exhibit such a clear pattern for the productivity of investment.

Given that returns to investment are larger in the first stage, if parents care more for cognitive rather than noncognitive skills of their offspring, we should expect investment to be higher in early childhood rather than in later childhood.

Another feature of the technology is that investment in the second and first stage are strongly complementary: this happens because first stage investment enters second-stage skills production through the self-productivity of future periods' skills. Hence, the more investment is performed today, the more it is required tomorrow, even only to keep skills constant.

Figure 2 shows how much investment is required in order to keep skills constant in the second stage, after investing x units in the first stage, for a median household.

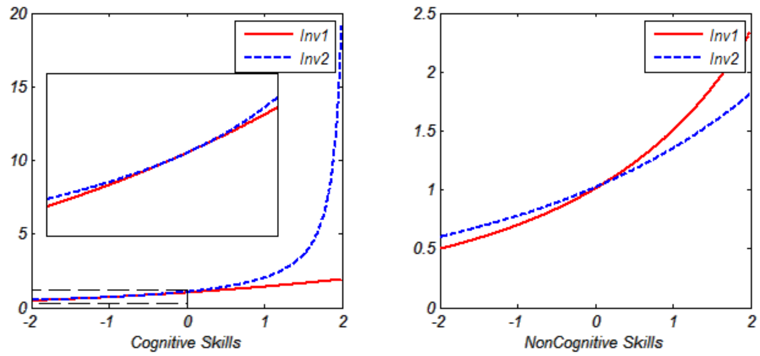

Figure 1: Amount of investment required to increase skills by 1 % of a standard deviation, by level of log standardized initial skills, in the first stage (red line) and second stage (blue line); parental skills fixed at the median. Graph includes magnification for lower-than-median initial cognitive skills.

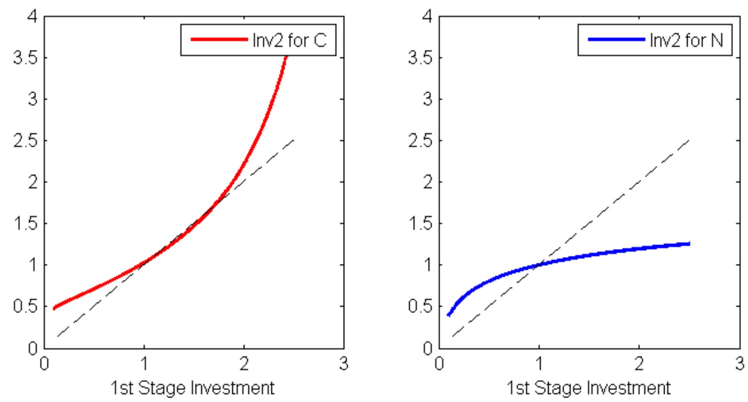

Figure 2: Amount of second-stage investment required to maintain skills constant, by initial investment, against 45° line; child's initial skills and parental skills fixed at the median.

The natural consequence of these two features is that we expect investment to be "smoothed" across phases, on average; moreover, household groups who invest more in the first stage will, on average, invest more also in the second stage.

The final feature I discuss here is that high-skilled parents are more productive at raising skillful children; Figure 3 summarizes this feature of the technology.

For instance, when a mother's cognitive skills are one standard deviation above the median, the first-stage gains in the child's cognitive skills are higher by 10 % with respect to what the median mother would produce. In general, higher parental skills yield to higher offspring's skills; and these gains are larger during early childhood than later childhood.

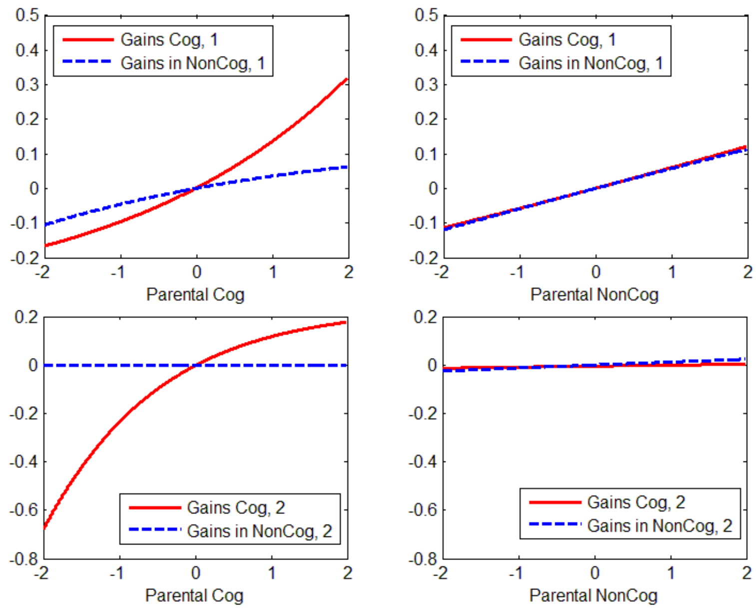

Figure 3: Gain in skills (as fraction of a standard deviation) by log standardized parental skills and by developmental stage; initial child's skills fixed at the median.

2.2 Investment in Children

The technology of skills formation estimated by Cunha, Heckman and Schennach allows to quantify returns from investment in children; in order to link these returns to the patterns of child care time, a mapping between observables and the abstract concept of "investment" is required. CHS identify investment from a large number of measurements which include whether the child has access to education goods (such as theatres, museums, musical shows, books, musical instruments), the number of specific toys children own and others. In the present paper, I make the assumption that investment at time t can be expressed as

$$I_t = A(\alpha_t x_t^{\omega} + (1-\alpha_t) e_t^{\omega})^{1/\omega}, \tag{2}$$

where x_t is primary child care time spent by the household with its offspring, e_t is the amount of goods spent in child care, which are a stand-in for all goods that might be relevant for a child's skills development, α_t determine the relative weight of each input, which are allowed to vary over the age of the child, and ω determines the degree of complementarity/substitutability between time and goods.

2.3 The Model

Finally, a decision theoretic model that embeds the two previous ingredients is developed, in order to rationalize the observed cross-sectional patterns of investment in children. The model is a parental choice model in the spirit of Becker and Tomes (1976) (1979) and Cunha and Heckman (2007). Differently from the aforementioned models, I include a time tradeoff, consider investment as a combination of child care time and goods and I assume parents to care for the "quality" of their child in every period.

Households face a time allocation problem and a resource allocation problem. Time is limited and must be allocated among work, time invested in children and leisure; resources come from labor income and must be allocated between consumption and goods invested in children.

2.4 Environment and Timing

The economy is populated by a continuum of households of measure 1, each of which has one offspring as in Cunha and Heckman (2007). Time is discrete (indexed by t) and there are $T+K$ time periods, $t = 1, 2, ...T_1, ..., T+K$ where periods $1, ..., T_1$ belong to early childhood, $T_1 + 1, ...T$ belong to late childhood and periods from $T+1$ to $T+K$ denote periods in which households cannot invest in their child anymore.

In the first period, households are endowed with cognitive skills $s_{C,P}$ and noncognitive skills $s_{N,P}$, which are assumed to be time-invariant; these will be referred to as "parental skills". Every child is also endowed with initial skills $\{s_{C,1},\ s_{N,1}\}$, which will be referred to as "offspring's skills". Households are heterogeneous in initial conditions $\left\{s_{C,1}, s_{N,1}, s_{C,P}, s_{N,P}\right\}$ and, as is standard in the literature, are assumed to have full knowledge of them.

In the economy there exists only one good, which is used as the numeraire; such good can be indifferently consumed or used for investment in children.

2.5 Preferences and Choices

In periods $1, ..., T+K$, households decide how to allocate one unit of time into working n_t, time with their offspring x_t and leisure, how much good c_t to consume and how much to spend in goods e_t for their offspring. The household gets labor income $w(s_{C,P}, s_{N,P}, \epsilon_P^w)\ n_t$, where the wage is a function of parental skills plus additional household-specific heterogeneity ϵ_P^w, and n_t is the amount of time spent working in period t[6]. Since parental skills $\left\{s_{C,P}, s_{N,P}\right\}$ and household-specific heterogeneity ϵ_P^w are fixed over time, the wage faced by the household is fixed as well.

Following Restuccia and Urrutia (2004), there is no financial asset that allows redistribution of resources between time periods.[7] Hence the budget constraint of an household can be written as

$$c_t + e_t \leq w(s_{C,P}, s_{N,P}, \epsilon_P^w) n_t \quad \text{for } t = 1, ..., T+1 \tag{3}$$

Since there are no income shocks, borrowing constraints across periods are not generating large inefficiencies. As Restuccia and Urrutia (2004) argue, this is likely to imply that policy returns estimated in the model are a lower bound of true, underlying policy returns, because income shocks would give more scope for redistribution.

The consumption good c_t gives CRRA utility $\frac{c_t^{1-\theta}}{1-\theta}$. Leisure gives utility $\frac{\zeta(1-n_t-x_t)^{1-\sigma}}{1-\sigma}$; households discount future outcomes at the common rate β.

Finally, using the terminology of Becker and Tomes (1976), at each period the "quality" of children in terms of both cognitive skills $s_{C,t}$ and noncognitive skills $s_{N,t}$ maps into parental utility $W(s_{C,t}, s_{N,t})$, weakly increasing in both arguments and strictly concave (that is, $W_{s_{C,t}} \geq 0$, $W_{s_{N,t}} \geq 0$ and the Hessian of W is negative definite for every level of skills).[8]

[6] The underlying assumption is that time of the father and of the mother are perfect substitutes. One might think that some activities can exclusively be performed by the mother, i.e. breastfeeding; however, empirical evidence suggests that although children of single mothers are at a disadvantage, such disadvantage is too small to be reconciled with strong complementarity between fathers' and mothers' time. For instance Carlson and Corcoran (2001) show that the difference between children in cognitive scores of single-parent households with intact households is statistically insignificant after controlling for income and Army Force Qualification Test score of the mother.

[7] In a sense, this is a very strong form of borrowing constraint; however, inside each period resources can be freely moved in time. The assumption is simplistic but allows to take into account long-term constraints in resources in the simplest way.

[8] Such assumption is standard in models of parental choices, see for instance Del Boca et al. (2014) and Brilli (2012). Other models assume that parents only care for the future continuation value of their children. However, in the technology of skills formation I assume, investment depreciates over time. In a robustness check, I show that if parents get utility only from the final quality of their children, this implies that investment is increasing over time, because investment in the first periods depreciates. Such a pattern is completely counterfactual, as the data show that time invested in children is decreasing in the age of the child.

2.6 Investment in children

Offspring's skills evolve according to the two-stage production function described in the first subsection:

$$s_{k,t+1} = [\gamma_{j,k,1} s_{C,t}^{\phi_{j,k}} + \gamma_{j,k,2} s_{N,t}^{\phi_{j,k}} + \gamma_{j,k,3} I_t^{\phi_{j,k}} + \gamma_{j,k,4} s_{C,P}^{\phi_{j,k}} + \gamma_{j,k,5} s_{N,P}^{\phi_{j,k}}]^{1/\phi_{j,k}} exp(\eta_{j,k})\,, \tag{4}$$

for time $t = 1, ...T_1, ..., T$, stages $j = \{1, 2\}$ and skills $k = \{C, N\}$. The technology exhibits constant returns to scale, that is, $\sum_{i=1}^{5} \gamma_{j,k,i} = 1$, for $j = \{1, 2\}$ and $k = \{C, N\}$. Investment I_t is given by the combination of time x_t and goods e_t described in equation 2. Finally, shocks $\eta_{j,k}$ are assumed to be independently normally distributed and realize at the end of the period, hence households have to form expectations on next period's skills of the child for all possible realizations of the shocks.

2.7 Dynamic Problem

The state of each household at time t can be described by the current period's skills of her offspring plus the additional income variability ϵ_P^w, where the household-specific characteristics $s_{C,P}, s_{N,P}, \epsilon_P^w$ are constant while offspring's skills $s_{C,t}, s_{N,t}$ evolve over time. Hence the problem of a household in period t which belongs to developmental stage $j \in \{1, 2\}$ can be written as follows:

$$\begin{aligned} V_t(s_{C,t}, s_{N,t}, s_{C,P}, s_{N,P}, \epsilon_P^w) = \max_{c_t, e_t, n_t, x_t} & \frac{c_t^{1-\theta}}{1-\theta} + \zeta \frac{(1 - n_t - x_t)^{1-\sigma}}{1-\sigma} + W(s_{C,t}, s_{N,t}) + \\ & \beta \mathbb{E}\Big[V_{t+1}(s_{C,t+1}, s_{N,t+1}, s_{C,P}, s_{N,P}, \epsilon_P^w) \Big] \end{aligned}$$

$$
\begin{aligned}
\text{s.t. } c_t + e_t &\leq w(s_{C,P}, s_{N,P}, \epsilon_P^w) n_t \\
I_t &= A(\alpha x_t^\omega + (1-\alpha) e_t^\omega)^{1/\omega} \\
0 \leq n_t + x_t &\leq 1, \quad n_t, x_t \geq 0 \\
s_{C,t+1} &= [\gamma_{j,C,1} s_{C,t}^{\phi_{j,C}} + \gamma_{j,C,2} s_{N,t}^{\phi_{j,C}} + \gamma_{j,C,3} I_t^{\phi_{j,C}} + \gamma_{j,C,4} s_{C,P}^{\phi_{j,C}} + \gamma_{j,C,5} s_{N,P}^{\phi_{j,C}}]^{1/\phi_{j,C}} \exp(\eta_{j,C}) \\
s_{N,t+1} &= [\gamma_{j,N,1} s_{C,t}^{\phi_{j,N}} + \gamma_{j,N,2} s_{N,t}^{\phi_{j,N}} + \gamma_{j,N,3} I_t^{\phi_{j,N}} + \gamma_{j,N,4} s_{N,P}^{\phi_{j,N}} + \gamma_{j,N,5} s_{N,P}^{\phi_{j,N}}]^{1/\phi_{j,N}} \exp(\eta_{j,N}) \\
\eta_{j,C} &\sim \mathcal{N}(0, \sigma^2_{\eta_{j,C}}), \quad \eta_{j,N} \sim \mathcal{N}(0, \sigma^2_{\eta_{j,N}}).
\end{aligned}
$$

At the end of skills development, parents still get utility from the quality of their children, but cannot influence them anymore. Their maximization problem in periods $T+1, ..., T+K$ becomes

$$
\begin{aligned}
V_t(s_{C,T+1}, s_{N,T+1}, s_{C,P}, s_{N,P}, \epsilon_P^w) = \max_{c_t, e_t, n_t, x_t} & \frac{c_t^{1-\theta}}{1-\theta} + \zeta \frac{(1-n_t-x_t)^{1-\sigma}}{1-\sigma} + \\
& W(s_{C,T+1}, s_{N,T+1}) + \\
& \beta \mathbb{E}\Big[V_{t+1}(s_{C,T+1}, s_{N,T+1}, s_{C,P}, s_{N,P}, \epsilon_P^w)\Big]
\end{aligned}
$$

$$
\begin{aligned}
\text{s.t. } c_t + e_t &\leq w(s_{C,P}, s_{N,P}, \epsilon_P^w) n_t \\
0 \leq n_t + x_t &\leq 1, \quad n_t, x_t \geq 0
\end{aligned}
$$

while the maximization problem in the last period $T+K$ is

$$
\begin{aligned}
V_{T+K}(s_{C,T+1}, s_{N,T+1}, s_{C,P}, s_{N,P}, \epsilon_P^w) = \max_{c_{T+K}, e_{T+K}, n_{T+K}, x_{T+K}} & \frac{c_{T+K}^{1-\theta}}{1-\theta} + \zeta \frac{(1-n_{T+K}-x_{T+K})^{1-\sigma}}{1-\sigma} + \\
& W(s_{C,T+1}, s_{N,T+1})
\end{aligned}
$$

$$
\begin{aligned}
&\text{s.t. } c_{T+K} + e_{T+K} \leq w(s_{C,P}, s_{N,P}, \epsilon_P^w) n_{T+K} \\
&0 \leq n_{T+K} + x_{T+K} \leq 1, \quad n_{T+K}, x_{T+K} \geq 0
\end{aligned}
$$

2.8 Economic Mechanisms and tradeoffs

First order conditions imply that households trade off consumption and leisure following the equation

$$(1 - n_t - x_t) = \left(\frac{\zeta c_t^{\theta}}{w} \right)^{1/\sigma} \tag{5}$$

Notice that the labor choice cannot be separated from the choice of time with the offspring, so that an increase in the future value of investment in the offspring will yield, *ceteris paribus*, to a variation in labor time.

Taking derivative with respect to x_t yields

$$\zeta(1 - n_t - x_t)^{-\sigma} = \mu_t A \alpha x_t^{\omega-1} \big(\alpha x_t^{\omega} + (1-\alpha) e_t^{\omega} \big)^{\frac{1-\omega}{\omega}} , \tag{6}$$

where μ_t is the multiplier associated to the investment equation at time t, which encompasses the combination of the productivity of investment and the shadow value that each agent attributes to her offspring's future skills.

Equation 6 states that in an interior optimum, the marginal value of leisure must equal the marginal value of investment in the offspring. Substituting equation 5 inside 6 yields

$$w c_t^{-\theta} = \mu_t A \alpha x_t^{\omega-1} \big(\alpha x_t^{\omega} + (1-\alpha) e_t^{\omega} \big)^{\frac{1-\omega}{\omega}} , \tag{7}$$

which makes clear the tradeoff faced by agents between consumption and investment; working time (hence consumption) must be traded off with time invested in the offspring.

In an interior solution, investment goods are positively related to the wage w of the agent and to total time spent with the offspring x_t, depending on the complementarity parameter ω.

$$e_t = \left[\left(\frac{1-\alpha}{\alpha} \right) w \right]^{\frac{1}{1-\omega}} x_t \tag{8}$$

The solution is always interior in the baseline model; however, results are easily extendable to the case in which there exists a government that transfers resources to households, so that the latter might work zero hours if such transfers are large enough. If $n_t = 0$, investment goods satisfy instead the equation

$$e_t = \left[\frac{(1-\alpha)}{\alpha} \, \zeta (1 - x_t)^{-\sigma} (z_t - e_t)^{\theta} \right]^{\frac{1}{1-\omega}} x_t \, , \tag{9}$$

where z_t represents lump sum transfers from the government. Equation 8 is a particular case of equation 9, in which equation 5 is used to simplify the marginal utility of leisure and of consumption.

Notice that equations 8 and 9 give the solution for e_t even in the boundary case in which child care time x_t is equal to zero; this can occur if the marginal product of investment is finite for $I_t = 0$. Since the technology of skills formation is of the CES form, this happens if the complementarity parameter ϕ is lower than zero; for $\phi > 0$, the solution for x_t is always interior.

To clarify: in the simpler case in which the parameter ω tends to zero, the investment formation technology becomes Cobb-Douglas and in an interior solution investment becomes equal to

$$I_t = Ax_t \left[w\left(\frac{1-\alpha}{\alpha}\right) \right]^{1-\alpha}, \tag{10}$$

so that both the wage and time spent with children matter for the evolution of skills.

Clearly, in the periods $T+1, ..., T+K$, since households cannot influence their child anymore, maximization implies $x_t, e_t = 0$; n_t still satisfies equation 5.

If the share of time in the investment function is larger than the share of goods ($\alpha > 1/2$), two main features of the model arise from the FOCs. First, households will spend more time in child care in the most productive stage; second, depending on the parameters of the utility function, they may also choose to work fewer hours.

Proposition 1: *Assume $\omega \to 0$. Consider the multipliers μ_t associated to the constraint $I_t = Ax_t^{\alpha} e_t^{1-\alpha}$ as the function $\mu_t = \mu_t(K_t, S_t, x_t)$, where $\frac{\partial \mu_t}{\partial K_t} > 0$.*

- *Suppose that $\alpha > 1/2$; then we have that $\frac{\partial x_t}{\partial K_t} > 0$, that is, agents respond to increased productivity in investment by increasing time invested in the offspring.*

- *If $\alpha > 1/2$, preferences satisfy Balanced Growth Path ($\theta = 1$) and we have that $\sigma \in [0,1], \zeta > \frac{1-\alpha}{\alpha\sigma}$, then $\frac{\partial n_t}{\partial K_t} < 0$, that is agents respond to higher productivity in investment by decreasing hours of work.*

3 Data and Estimation

I use data from the Children of the National Longitudinal Survey of the Youth 1979 (CNLSY/79) and from surveys of work hours and child care time carried out from 2003 to 2008. The data from the CNLSY/79 is the same as in Cunha, Heckman and Schennach (2010). This choice is motivated by the fact that, by choosing to introduce their technology in the model, the model has to be consistent with the patterns found in the data used to estimate the technology itself. The dataset of CHS is a collection of variables regarding 2207 firstborn white children from the CNLSY/79 sample. Children in the dataset have been assessed every 2 years, along with their mothers, starting in 1986. Assessments start at birth and end at age 14; they include several measures of cognitive achievement, such as the PIAT mathematics and reading comprehension tests, and measures of noncognitive achievement such as temperamental scores. For very early ages (0-2), the best predictors of future tests are measured; for instance, for estimating cognitive skills at birth, the authors use gestation length, birth weight and motor-social development.

I obtain part of the estimation targets from the estimation of "skills factors" from assesments of children and mothers. Following CHS, the statistical tool employed is factor analysis; the idea is that a set $[Z_1, ..Z_i, ..Z_M]$ of variables, such as tests of mathematical and reading abilities, are error-contaminated measurements of the underlying cognitive and noncognitive abilities $\{s_C, s_N\}$ of an individual. Then, each measurement i is assumed to be related to the unobservable skills of individual j at time t according to

$$Z_{i,j,t} = \alpha_{i,t} + \beta_{i,t} log(s_{C,j,t}) + \epsilon_{i,j,t} \,, \tag{11}$$

so that the underlying latent variables $s_{C,j,t}, s_{N,j,t}$ can be identified from the covariance between measurements up to the normalization of one of the coefficients $\beta_{i,t}$.[9] In this study, the latent variables are simply obtained by taking the first principal factor of several different measurements for cognitive skills and noncognitive skills, taken in the same year; the underlying identifying assumption is that, for two measurements i, j of the same child such that $i \neq j$, $\mathbb{COV}(\epsilon_{i,t}, \epsilon_{j,t}) = 0$. Cunha, Heckman and Schennach identify the factors within the estimation procedure of the technology; I choose a different strategy because of simplicity and transparency, but in principle I could use the same factors as targets for the model. Part of the correlation matrix between offspring's skills and parental skills at the end of early childhood will be taken as targets of the model. For consistency in the use of the technology, I estimate the factors following closely the choice of variables described in Cunha, Heckman

[9]See Cunha, Heckman, Schennach (2010) for a discussion of the application of such methodology in the context of skills formation.

and Schennach (2010).[10]

In line with Cunha, Heckman and Schennach (2010) I consider parental skills to be the mother's skills, skills of the offspring at the end of early childhood to be the latter's skills at ages 5-6 and final skills to be offspring's skills at age 13-14. Such skills have a data counterpart in the factors calculated from tests. In the Appendix I present the estimated correlation matrix for offspring's skills and parental skills, at the end of early childhood (Table XIV) and at ages 13-14 (Table XV); I choose part of these to be targets for calibration.

Targets and stylized facts for child care time are calculated on the 2003 to 2008 waves of the American Time Use Survey. Specifically, I use data from Ramey and Ramey (2010), which is the merge of several surveys of time uses from 1965 to 2008. I choose to target averages from the 2003 to the 2008 waves of the American Time Use Survey (ATUS) because of the larger sample size and narrower time window, which makes the data easier to compare across years. I discuss in detail the choice of data and the estimation sample in the Appendix.

3.1 Calibration

Two of the main ingredients of the model are taken from the paper of Cunha, Heckman and Schennach (2010): the parametrization of the technology of skills formation and the initial distribution of skills at birth. Parameters of the technology are reported in Table XII, including the variance of the shocks to skills; the authors estimate several versions of the technology under different sets of assumptions, such as the existence of unobserved heterogeneity across households or endogeneity of investment; I choose the latter estimates as they already correct for endogeneity of the investment function, making it more suitable for inclusion in a decision theoretic model.

In line with CHS, cognitive and noncognitive skills at birth and parental skills are assumed to be jointly lognormally distributed with mean zero and covariance matrix Σ. The parametrization of the covariance matrix is taken from the appendix of the paper of CHS and available in Table XI in the Appendix to this paper, along with the correlation matrix, to allow an easier interpretation.

The technology has been estimated on two-years-long intervals, hence I set the time span of the model so that one period corresponds to two years. Periods 1,2,3 correspond

[10] Table XIII in Appendix C provides basic statistics on the variables in the dataset for ages 5-6 and 13-14, showing that they match closely the results by Cunha et al. (2010).

to early childhood, from when a child is born to when he is 6; periods 4,5,6,7 correspond to late childhood so that skills development is assumed to end at age 14, and periods 8-15 correspond to the final periods, so that a household's life lasts 30 years since the birth of the child. Summarizing, I set $T_1 = 3$, $T = 7$ and $T + K = 15$.

The discount factor β is set to 0.92, which is equivalent to 0.96 at the yearly level, a standard value in the macroeconomics literature.

Following Osuna and Rios Rull (2003), I set the time endowment of households to be 200 hours per week, excluding sleep and personal maintenance.

The baseline value of risk aversion θ is set to 1, which implies that utility of consumption is given by $\log(c_t)$. As Erosa, Fuster and Kambourov (2012) argue, values of θ different from 1 imply different hours of work across lifetime income groups, while the data show that individuals with different levels of lifetime income tend to work roughly the same amount of hours.[11]

The curvature of leisure σ (at the micro level) is relevant for the analysis as it determines the substitution between investment in children and leisure. I set the baseline value in order to obtain a plausible elasticity of labor supply at the household level; a wide literature attempts to estimate such elasticity. For instance Pistaferri (2003) argues for a value near 0.7, while Erosa, Fuster and Kambourov (2012) use a value of about 0.6. Keane (2011) also shows that estimates of labor supply elasticity vary widely in the literature, and the average of the estimates is around 0.85. I target the value of 0.6, which lies in the upper range of the micro estimates of the early literature, and well within the range surveyed by Keane (2011). The baseline value of σ is set to 3.5, which implies a Frisch labor supply elasticity $\eta^{\lambda} = 0.58$ in the first stage for the median household, 0.53 in the second stage.[12] I use as a reference point the median household because the model generates a distribution of elasticities of labor supply, due to the fact that households choose jointly different levels of child care time and of labor supply.

The parameter ζ targets the average hours of work of married households aged 25-44. A well-known stylized fact is that per-capita hours of work in the US have not moved much in the last 50 years; As the average hours per week in 2000 were around 41 for married men and 25 for married women in the 25-34 and 35-44 years old age groups (without children or

[11] I perform sensitivity analysis with the values of 0.8 and 1.2. Results do not change sensibly depending on θ.

[12] I have performed several sensitivity checks for elasticities in the range of $(0.25, 1)$ and found that they have little impact on results.

with a child older than 6), this gives 66 hours per household which means $66/200 = 33\%$ of available time per week[13].

The shares of time α_t in the investment function gauge the relative importance of time vs goods in the investment function. Since $1 - \alpha_t$ roughly translates to the share of goods invested in the offspring, I choose the sequence of α_t that matches the shares of income spent in child care, education and health of the child as reported by the USDA in 2011, by age of the child, for married, single-child households.[14] I restrict the sequence of α_t to have a third-degree polynomial shape, so that I use 4 parameters to match 5 shares (ages 0-2, 3-5, 6-8, 9-11, 12-14).

The scale A of the investment function implies a normalization of final skills; the chosen normalization is that the logarithm of final cognitive skills generated by the model has mean zero, i.e. the average skills of mothers and offsprings are equal at the end of the developmental process. One possible concern is that the psychometric literature reports that several countries have experienced massive increases (up to 1.2 standard deviations) in cognitive test scores in the last 50 years, the so called "Flynn effect" (see Flynn 2009). However, whether this represents a generalized increase in cognitive abilities or, instead, an increase in the ability to perform well in tests, is a matter of debate. To answer this possible critique, I perform a robustness check in which the scale A targets another possible normalization. Results do not change very much depending on the chosen normalization.

Estimating the complementarity between time and goods ω directly is difficult because it requires to observe time allocation choices, household expenditures in children and parental investment at the same time. I take an indirect inference approach instead. Equation 8 describes a relationship between expenditures in the offspring and child care time in the model. Solving for child care time, I can write

$$x_t = \left[(\frac{1-\alpha}{\alpha}) w \right]^{\frac{1}{\omega-1}} e_t \,,$$

[13]These numbers are taken from McGrattan and Rogerson (2004), which give data on hours worked of individuals from 1950 to 2000.

[14]Source: "Expenditures on Children by Families, 2011 (Miscellaneous Publication Number 1528-2011)", USDA.

and taking logs we obtain:

$$\log x_t = \frac{1}{\omega - 1}\log\big(\frac{1-\alpha}{\alpha}\big) + \frac{1}{\omega - 1}\log w + \log e_t \tag{12}$$

Equation 12 implies that it is possible to obtain the coefficient $\frac{1}{\omega-1}$ by regressing child care time x_t on the wage faced by households w. However, e_t is unobservable and correlated with w_t by definition, creating an omitted variable problem that makes ω impossible to identify with the simple regression. I use the model-generated data to circumvent this issue. I perform a regression between child care time and log wages both on the ATUS data and on the model-generated data, omitting e_t in both regressions, and choose ω to make the coefficient estimated in the model as close as possible to that in the data.

Income at the household level depends on parental skills and on the labor supply choice; while the dataset provides the former, the latter are unobserved. Here too I use indirect inference to make the income process generated by the model consistent with the NLSY/79 data. I first run a Mincer regression between household log income and parental cognitive and noncognitive factors, uncovering Mincer returns for the two skills.[15]

$$\log Y_t = const + \beta_C s_{C,P} + \beta_N s_{N,P} + \Gamma controls + \epsilon_t\,. \tag{13}$$

Then, returns to skills for wages in the model are calibrated so that, when performing an analogous Mincer regression on model-generated data, the estimated coefficients match the data counterparts β_1 and β_2.

Controls include a full set of year dummies and a cubic polynomial in the age of the mother. To make the coefficients of the regression consistent with the scale of parental skills in the model, factors estimated from the data and the model are rescaled to have standard deviation equal to 1. Moreover, as the model has 2-years time periods, the regression is performed on the sum of 2 years of log income[16], which allows to get a better fit by diminishing the amount of income variance due to measurement error or temporary idiosyncratic shocks. Table I reports results from the estimation of a GLS random effects model on the dataset.

[15] I use only the mother's factors because the dataset does not provide information on the father's skills. However, using only mother's skills allows to account for the correlation between fathers and mothers while economizing state variables in the model.

[16] The concept resembles the idea of "permanent income"; in order to get better estimates, one could sum income of more years. However, the NLSY data does not provide enough consecutive observations of income, a shortcoming that would make the estimation sample smaller and the estimates less precise.

	(1) Log Family Income
Norm. Mom Cog Fac	0.288***
	(0.017)
Norm. Mom NonCog Fac	0.074***
	(0.017)
Constant	0.568*
	(0.332)
Observations	5943
R^2 - Within	0.037
R^2 - Between	0.333
R^2 - Overall	0.243

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table I: Mincer regression: Log Household Income in the NLSY/79 as a function of Log Cognitive and Noncognitive Factors of the mother; years 1979-2004, household with child older than 6 present. Controls include a full set of year dummies and a cubic polynomial in age (omitted). Source: data extract from NLSY79/CNSLY79 by Cunha, Heckman and Schennach (2010).

Household-specific heterogeneity in wages ϵ_P^w is assumed to be normally distributed, independent of skills and it is calibrated to have the same variance as that observed in the fixed component of the residual of the Mincer regression ϵ_t estimated above.[17]

Finally, the choice of the utility function W is somewhat more difficult, as many adult outcomes may contribute to the value parents attribute to investment in the offspring; the chosen functional form for utility given by "quality" of children is

$$W(s_{C,t}, s_{N,t}) = \chi \frac{\left(s_{C,t}^{\psi} s_{N,t}^{1-\psi}\right)^{1-\xi}}{1-\xi}.$$

χ gives the relative importance of the future of the offspring with respect to consumption of the household and leisure. The higher χ, the higher will be the value of the offspring's skills for households, the higher will be investment. Hence, this parameter targets the average amount of time parents spent in early primary child care in 2003-2008, from when the offspring is born to when he/she is five years old; I target the average of early child care

[17] I exclude the variance of the random component in order to filter out measurement error and idiosyncratic income shocks, which the model does not feature.

time as observed in the American Time Use Survey of 2003-2008 for married parents, which is around 13% of the available endowment.[18]

ψ determines the relative importance of cognitive skills w.r.t. noncognitive skills and ξ encompasses the risk aversion in the future of the offspring. The first parameter is identified by targeting the ratio of time spent by parents in early (ages 0-6) versus late (ages 7-14) child care time. The intuition is that, since early childhood is crucial for the development of cognitive skills, we should observe that parents spend relatively more time in child care during that phase if they care more for cognitive skills, rather than noncognitive ones. Instead, if parents care more for noncognitive skills, we should observe a flatter profile over the child's age. Hence, ψ can be identified by targeting how steep the age profile of child care time is in the ATUS data.

Finally, a lower risk aversion in the quality of the child (parameter ξ) implies that high-skilled parents have stronger incentives to invest, because marginal returns are higher than those of consumption; hence, a lower risk aversion increases the intergenerational persistence of skills. Following this intuition, I use ξ to match the correlation between parental skills and offspring's skills at the end of early childhood.

Tables II (externally set parameters) and III (endogenously determined) summarize the proposed calibration; the model has no trouble in matching the targets very closely.

The calibration shows first of all that goods by themselves are relatively unimportant for child development: the calibrated values of α_t (Figure 4) are always larger than 0.8, suggesting that the pattern of investment is mainly explained by the pattern of child care time. The relative importance of goods increases in later childhood but the role of time remains prominent.

Second, the calibrated value of $\omega = -0.185$ shows that time and goods are complements, so that it's not easy for high-income households to substitute child care time with goods and viceversa.

Third, $\psi = 0.74$ suggests that cognitive skills are relatively more important to households

[18]The number is calculated by taking the sum of average hours per week spent by mothers and of average hours per week by fathers, and dividing it by the stock of available hours per week, assumed to be 200 hours. This target may be different depending on the definition of child care time / time with children that are considered as investment; see for instance Ramey and Ramey (2010), Bianchi (2000), Sandberg and Hofferth (2001), Del Boca, Flinn and Wiswall (2014)

[19]See McGrattan and Rogerson (2004).

[20]See McGrattan and Rogerson (2004).

[21]See Restuccia and Urrutia (2004).

Parameter	Value
Technology of Skills Formation	Cunha, Heckman, Schennach (2010) (see Table XII in the Appendix)
Covariance Matrix of Initial Conditions	Cunha, Heckman, Schennach (2010) (see Table XI in the Appendix)
Duration of One Period	2 years
β	0.92
θ	1
σ	3.5

Table II: Calibration of parameters/functional forms set exogenously.

Parameter	Value	Target	Data	Model
Preferences				
ζ	0.49	Hours worked	0.330[a]	0.336
χ	0.55	Avg. Time in Child Care when child < 6	0.134[b]	0.136
ψ	0.74	Ratio Early/Late Time	1.767[c]	1.853
ξ	0.49	Correlation($s_{C,4}, s_{C,P}$)	0.284[c]	0.261
Income Equation				
$\beta_{C,\text{model}}$	0.44	Mincer returns to $s_{C,P}$	0.288	0.289
$\beta_{N,\text{model}}$	0.54	Mincer returns to $s_{N,P}$	0.074	0.074
Variance(ϵ_P^w)	0.47	Var. of Mincer Residuals	0.218	0.222
Investment				
A	10.348	Mean($s_{C,3}$) = 0		
α_1	0.855	Share of income spent in child, ages 0-2	0.046[c]	0.048
ω	-0.185	Indirect Inference, Time on Wage	0.135	0.140

[a] See McGrattan and Rogerson (2004).
[b] Source: author's calculations on 2003-2008 ATUS.
[c] Source: author's calculations on CNLSY/79.

Table III: Calibration of parameters endogenously determined; targets, data moments and simulated moments.

than noncognitive skills; as I will show, this implies that early childhood investment is substantially higher than later investment. As the productivity of investment for cognitive skills is higher during early childhood, households concentrate their efforts there.

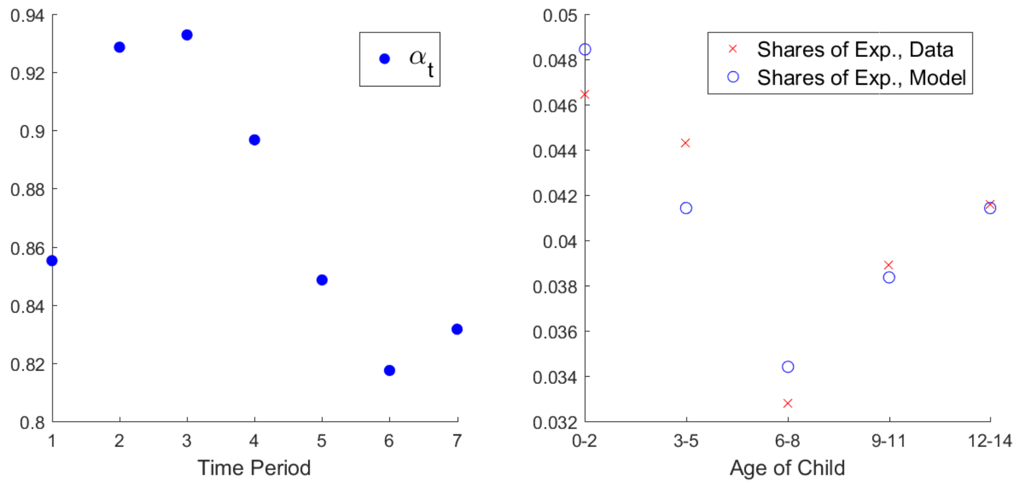

Figure 4: Left panel: Shares of time α_t in investment function by model period. Right panel: shares of income spent in children, model vs data.

Finally, $\xi = 0.49$ suggests that risk aversion in the "quality" of children is lower than risk aversion in consumption; in fact, sensitivity analysis on the risk aversion for consumption θ shows that calibrated values of ξ are always lower than θ.

4 Results and Discussion

I solve the model and externally validate it by exploring how the model performs in replicating non-calibrated stylized facts of child care time and intergenerational persistence. The model presents some computational challenges, as there are 5 continuous state variables and the dynamic problem changes at all periods due to finite life and different technologies for different phases of childhood.[19] I use polynomial approximation of the value function to solve the model. For this particular application, polynomial approximation gives very precise and reliable results; I discuss in detail the algorithm in the Appendix.

In order to generate college/noncollege differentials in the model, I first estimate a probit model of the probability of a mother having completed college depending on her cognitive and noncognitive skills, using the CNLSY/79 data. Then, I use the estimated probit to

[19]In principle I could discretize state variables and rely on interpolation techniques between grid nodes. However, this alternative approach is much more computationally intensive, even with very rough approximations. Moreover, I find it to yield large approximation errors.

partition households in the model economy into those with a college-educated mother and households with a less-than-college-educated mother.

	(1) Dep. Var: College Prob.
Norm. Mom Cog Fac	1.060***
	(0.060)
Norm. Mom NonCog Fac	0.047
	(0.044)
Constant	-1.160***
	(0.055)
Observations	1581
Pseudo-R2	0.278

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table IV: Probit model: being a college-educated mother as a function of her Cognitive and Noncognitive skills; sample includes all mothers aged > 24. Source: data extract from NLSY79/CNSLY79 by Cunha, Heckman and Schennach (2010).

The results of my simulations are summarized in Table V. First, I look at how time with children varies across parental education and across developmental stages. It is well known that parental time in child care is higher when a child is ages 0 to 6. Moreover, Ramey and Ramey (2010) and Guryan, Hurst and Kearney (2008) show that higher educated parents spend more time with their children, especially during early childhood. The first two sections of Table V report a comparison of key moments generated by the model with the data in the American Time Use Survey 2003-2008; conditional means are calculated on the data of Ramey and Ramey (2010) for parents aged 25-54. The means I use are obtained as predictions from a regression that controls for state and year fixed effects, a polynomial in age, ethnicity and the number of children in the household. Unfortunately the ATUS does not have information on the extent of assortative mating, i.e. how often college-educated mothers are married to college-educated fathers. I overcome this issue by weighting fathers with the degree of assortative mating in the US in 2007, as estimated by Eika, Mogstad and Zafar (2017).[20]

[20]Define p as the probability that a college-educated mother has a college-educated husband. In practice,

	Model	Data
Early (0-6) Time in Child Care		
College	15.0	14.9
Noncollege	13.2	13.0
Δ, College/NonCollege	12.9%	14.1%
Late (7-14) Time in Child Care		
College	8.3	7.5
Noncollege	7.1	6.8
Δ, College/NonCollege	17.2%	10.6%
Hours of Work		
Early (0-6)	29.6	30.8
Final (14)	33.6	33.0
% Change during Early C.	-12.0%	-6.8%
Intergenerational Correlations		
$\rho(\theta_{C,4}, \theta_{C,P})$	0.26 (calibr.)	0.28
$\rho(\theta_{C,T+1}, \theta_{C,P})$	0.40	0.43
$\rho(\theta_{N,4}, \theta_{N,P})$	0.15	0.20
$\rho(\theta_{N,T+1}, \theta_{N,P})$	0.14	0.14
$\rho(\theta_{C,T+1}$,HH Income)	0.27	0.29
$\rho(\theta_{N,T+1}$,HH Income)	0.17	0.22

Table V: External Validation, Data vs Model. Summary statistics for Time invested in children and Intergenerational Correlations. Data on time use are the author's calculations on the 2003-2008 American Time Use Survey, on married parents aged 25-44 of children aged 0-6 for early time and 7-14 for late time. Numbers are obtained by summing average primary child care time of mothers and average time of fathers, and dividing by the assumed time endowment of 200 hours. The degree of assortative mating is as in Eika, Mogstad and Zafar (2017) for the US. All observations are weighted as recommended by the ATUS. Work hours data are taken from McGrattan and Rogerson (2004) and are calculated as the sum of average working hours for married males plus average working hours for married females. Intergenerational Correlations are the author's calculations on CNLSY/79 data.

---

child care time for a household with a college-educated mother is calculated as:

$$\begin{aligned}\text{Avg. College Child Care Time} = &\frac{1}{2}\text{Avg. College Mother's Time} \\ &+\frac{1}{2}\Big[p(\text{Avg. College Father's Time}) \\ &+(1-p)(\text{Noncollege Father's Time})\Big]\end{aligned}$$

The model predicts that both during early and late childhood, college-educated parents spend more time with their children, as is found in the data. The model explains more than 90% of the college/noncollege difference in early child care time, while it overshoots the difference in late child care time found in the ATUS. Also, the model generates a decline in work hours among parents of very young children, again consistently with the data.

The reason behind the large difference in early and late child care time across all education groups is that households care for the cognitive development of their offsprings; since cognitive skills can be boosted primarily during early childhood, due to a combination of higher productivity of investment and higher elasticity of substitution between inputs, households direct most of their efforts during this stage. Moreover, the high elasticity of substitution makes returns to investment high for all households, so that even low-skilled parents prefer to invest in early childhood rather than later. However, in the later phase inputs are strongly complementary: this makes households invest more "smoothly" than what they would do if the later childhood elasticity were the same. This happens because households anticipate that they will have to sustain the skills of their children also in the later stage, when it will be unlikely and costly to increase them.

Finally, the model is successful in replicating the intergenerational correlations between the offspring's test scores and parental ones, and between the offspring's test scores and household income, that are found in the CNLSY/79. This means that simulated parental investment is consistent with the data.

Two main mechanisms drive investment; higher parental skills grant higher productivity and higher income. Goods are complementary to time: hence higher income families have a comparative advantage towards investing in children. Since marginal returns to skills are higher than those to consumption, college-educated households choose to invest more time in their children, despite their higher opportunity cost.

Notice that even higher income parents invest more time in their children, even after controlling for skills; the latter comparison highlights the role of complementarity alone, separated by the impact of skills.

and similarly for noncollege-educated mothers. p is the ratio between the probability of "being a college-educated woman with a college-educated man" and the overall fraction of college-educated women, found in Table C.1 of the Appendix of Eika, Mogstad and Zafar (2017).

4.1 Alternative Models

I present here a discussion of the key assumptions and of the performance of the model under alternatives; most of the main results under alternative assumptions are presented in Table VI.

	Baseline	(1) Only Cog Skills	(2) All ϕ = 0	(3) 2-Periods Model	(4) Final Pref.	(5) Flynn Effect
Share of time α (avg.)	0.873	0.116	0.879	0.963	0.893	0.878
Ratio Early/Late Time	1.853	1.395	1.866	2.474	1.144	1.801
Δ College/NonColl. (early)	13.1%	-5.9%	10.5%	19.5%	15.6%	14.6%
Δ College/NonColl. (late)	17.3%	11.3%	14.5%	11.0%	15.1%	15.1%
Ratio Early/Late Work Hours	88.0%	100.6%	88.6%	82.3%	100.1%	88.7%
$\rho(s_{C,T+1}, s_{C,P})$	0.398	0.917	0.315	0.284	0.397	0.529
$\rho(s_{N,T+1}, s_{N,P})$	0.143	–	0.149	0.144	0.131	0.152
$\rho(s_{C,T+1}, s_{N,T+1})$	0.227	–	0.213	0.149	0.197	0.196
$\rho(s_{N,T+1}, \text{HH Income})$	0.165	–	0.157	0.139	0.151	0.127

Table VI: Alternative Models: effects of changes in assumptions on simulated data. All models have been recalibrated on the same loss function as the baseline under the different assumptions.

4.2 A two-periods model

One might think that, once the technology has been identified for early childhood and late childhood, most results can be obtained also with a simpler two-periods model in which there is one period of early childhood, $t = 1$ and one period of late childhood $t = 2$, in the spirit of the policy experiment of Cunha, Heckman and Schennach (2010). However, given the technology of skills formation, having many periods of child development reinforces the correlation between parents and children every period. To see why, consider a very simple model of child development, in which future skills only depend on the child's own skills and parental skills in a linear fashion:

$$s_{t+1} = \gamma s_t + (1 - \gamma) s_P,$$

where $\gamma \in (0, 1)$. It follows that at period $t + k$,

$$s_{t+k} = \gamma^k s_t + (1 - \gamma^k) s_P$$

so that parental skills account for a larger fraction of the child's skills at $t + k, \forall k > 1$.

Simulation results (column 3 of Table VI) suggest that a two-periods-only model can roughly account for the patterns of child care time, but fails in delivering intergenerational correlations that are quantitatively consistent with the CNLSY/79, because such correlations require more periods to build up.

4.3 The role of preferences

In order to study the role of assumptions on the household's preferences, I compute a version of the model in which parents care for the skills of their children only in the terminal period (column 4 of Table VI). While such a model can be successfully calibrated to the data and is still capable of delivering quantitatively consistent intergenerational correlations of skills, other of its implications are highly counterfactual. Due to depreciation of skills in the technology, which is particularly important in early childhood, it predicts a flat pattern of child care time over the life cycle, which is completely inconsistent with the data. Also, it predicts that hours of work rise slightly during early childhood.

4.4 Equal elasticities across stages

Some of the parameters $\phi_{j,k}$ in the estimation of the technology of skills formation by CHS have a larger standard error: since these govern the elasticity of substitution between inputs, results may crucially depend on them. In order to investigate the robustness of results to values of these elasticities, I run a counterfactual exercise in which all elasticities are set equal to one (that is, the technology of skills formation becomes Cobb-Douglas). Results are shown in column 2 of Table VI. The fit of the model remains good on most dimensions of the data, although this model cannot generate a large enough intergenerational correlation of cognitive skills. Finally, simulations show that policy returns change only slightly when elasticities are forced to be equal to one.[21]

4.5 A Model with Cognitive Skills only

CHS show that the two-skills assumption is extremely relevant for policy analysis; when they estimate a cognitive-skills-only production function, policy prescriptions move investment from disadvantaged children to advantaged ones. To investigate how a one-skill model would perform on the data, I calibrate a version of the model in which the two-skills production function is replaced with the one-skill production function estimated by CHS in the Appendix to their paper.[22]

The one-skill model (column 1 of Table VI) presents two main quantitative issues; first, it is impossible to match the observed intergenerational correlation of cognitive skills at the end of early childhood, which the model grossly overestimates (0.64 against 0.28). As a result, the model predicts an intergenerational correlation at offspring's age 14 of 0.92, which is inconsistent with any study on such correlations. Second, the model cannot generate a large early/late child care time ratio and predicts that college-educated parents spend *less* time than noncollege ones with their children in early childhood: such predictions are inconsistent with the data on time uses.

Finally, policy results are substantially different in the cognitive-skills-only model. The same policies that, in the two-skills model, are relatively effective at influencing children's skills, are utterly ineffective in the cognitive-skills-only model. I will discuss this result in more detail in the policy subsection.

5 Policy Experiments

I explore the effects of applying different transfer and payments schemes to the economy. The first program I simulate is a simple flat transfer to all households, roughly equivalent to the German scheme of child allowances (Kindergeld).

[21] Results for policy simulations in which all elasticities are equal to one are available upon request.

[22] Appendix G features both the estimation results of CHS (Table XVI) and the calibration of my model with cognitive skills only (Table XVII).

The Kindergeld transfer program started in 1936; in 2012, the Kindergeld granted a monthly payment of 184 euros per child to virtually all households who have a child under the age of 18, although it can be extended to age 25 if the child is in school, at university or is doing professional training. The payment is performed for each child in the household, and raises to 190 euros for the third child and 215 for each additional child. The payment extends to citizens of EU countries and of several other countries, provided that they reside in Germany, and is not means-tested.[23]

The 2012 Kindergeld for the first and second child amounted to approximately 5% of the average household income in Germany.[24] For simulation purposes, I perform a flat transfer of 5% of the model-generated average income to all households from birth of the child to age 14 and compare the effect of such policy with respect to the baseline model.

I compare the impact of the Kindergeld with two other means-tested constrained policies. I restrict policies to take the form

$$\tau_{i,t} = \exp(\beta_0 + \beta_t t + \beta_w \text{Wage}_{i,t}) \tag{14}$$

so that the amount of resources $\tau_{i,t}$ given to household i when the offspring is of age t is allowed to vary with the age of the child and with the income of the household.[25] I term these policies "constrained" as they are allowed to depend only on the age of the child and the wage of the household: they will serve as a benchmark with which to compare more flexible policies that will be allowed to depend on all state variables. The exponential form allows some flexibility in the distribution of the policy and ensures that all transfers are positive. Under this restriction, I compute the policies that **maximize average cognitive skills** at the end of childhood such that the total amount of resources devoted to the policy equals 5% of total income produced by the economy from when the offspring is born to when he is 14 years old. I first compute the optimal transfer of this form, and then the optimal direct purchase of investment goods $e_{i,t}$. To be clear, in the first case I manipulate the budget constraint in equation 3 by simply adding the transfer to the right-hand side:

[23] Source: Social Security Throughout the World (http://www.ssa.gov/policy/docs/progdesc/ssptw/).

[24] In 2012, yearly average household income in Germany was 43500 euros. Source: http://www.voxeu.org/article/are-germans-poorer-other-europeans-principal-eurozone-differences-wealth-and-income, data from ECB Household Survey 2013.

[25] I choose to make the policies depend on the wage rather than on income for simplicity of computation. Allowing the transfer to depend on income increases the computational complexity but delivers similar results.

$$c_t + e_t \leq w(s_{C,P}, s_{N,P}, \epsilon_P^w)n_t + \tau_t \quad \text{for } t = 1, ..., T+1\,,$$

so that the transfer can be freely spent by the household. In the second case, I modify the investment equation 2 by adding the direct purchase to the goods invested by the household:

$$I_t = A(\alpha_t x_t^\omega + (1-\alpha_t)(e_t + \tau_t)^\omega)^{1/\omega}$$

I denote the first experiment as "Kindergeld", the second as "Transfer, Age+w" (to denote the dependency of the transfer on these two variables) and the third "Expenditure, Age+w" to indicate that the policy intervenes by investing directly in goods for the offspring's development. Results are summarized in Table VII.

I first comment on the aggregate results from all experiments. All policies are effective at impacting children's skills development. Average final cognitive skills increase on average by 3% of a standard deviation under the Kindergeld, by 4.3% under the optimal constrained transfer and by 6.4% under the optimal constrained direct expenditure policy. Noncognitive skills are also substantially impacted by all policies, rising by 3.9%, 5.6% and 9% of a standard deviation respectively. Also, all policies are effective at reducing the correlation of children's skills with parental income (from -11% in the Kindergeld case to -39% in the direct expenditure case). However, all policies are relatively ineffective at decreasing the intergenerational correlation of cognitive skills. This is compatible with the observation that the intergenerational correlation of skills is very similar in Germany and in the US (Anger and Heineck 2009), although the US does not feature a child allowance policy. Child allowances can reduce the impact of income by itself, that is the fact that higher-income households have more resources for their children, but cannot influence the fact that higher-skilled parents are more productive at raising skillful children.

All policies have a negative impact on labor supply, which drops by almost three hours in early childhood and by two hours in later childhood under the Kindergeld, and by more than six hours in early childhood under the optimal means-tested transfer scheme. In part this is because households have lower incentives to work because the transfer allows to reduce labor supply while increasing consumption: in all cases, more than 96% of the transfer is consumed. This, however, frees up leisure that is partly used to increase time invested in the offspring (+1.5 hours for the Kindergeld, +3.5 hours under the optimal transfer). In the case of the direct expenditure policy, labor supply is less affected: time invested rises

	Kindergeld (1)	Transfer, Age + w (2)	Expenditure, Age + w (3)
Parametrization			
β_t	0.00	-0.27	-0.06
β_w	0.00	-0.51	-0.13
Final Cog Skills			
Mean	+ 3.18%	+ 4.35%	+ 6.40%
Std. Dev.	+ 1.52%	+ 1.01%	+ 2.18%
Final NonCog Skills			
Mean	+ 3.87%	+ 5.59%	+ 9.09%
Std. Dev.	+ 2.23%	+ 2.74%	+ 4.68%
Correlation($s_{C,T+1}, s_{C,P}$)	- 0.81%	- 6.19%	- 5.47%
Correlation($s_{N,T+1}, s_{N,P}$)	- 3.14%	- 14.43%	- 14.21%
Correlation($s_{C,T+1}$, HH Income)	- 5.79%	- 20.66%	- 21.50%
Correlation($s_{N,T+1}$, HH Income)	- 14.92%	- 49.40%	- 57.73%
Time Invested, Avg (1st stg.)	+ 1.48	+ 3.50	+ 1.55
Time Invested, Avg (2nd stg.)	+ 0.58	+ 0.73	+ 0.21
Goods Invested, Avg (1st stg.)	+ 1.08%	+ 1.79%	- 17.40%
Goods Invested, Avg (2nd stg.)	+ 0.71%	+ 0.54%	- 14.88%
Portion of Transfer Consumed (avg)	97.56%	96.99%	
Consumption, Avg.	+ 3.18%	+ 3.19%	+ 1.22%
Hours Worked, Avg. (1st stg.)	- 2.70	- 6.46	- 1.81
Hours Worked, Avg. (2nd stg.)	- 2.06	- 2.20	- 0.82

Table VII: Policy experiments: average changes from baseline. Column (1): flat transfer of 5% of average income. Column (2): transfer worth 5% of average income, optimally allocated according to household wage and age of the child. Column (3): direct increase in expenditure in children worth 5% of average household income, optimally allocated according to household wage and age of the child. Changes in skills are reported as percentage of a standard deviation of baseline skills. Changes in time invested and hours worked are reported in hours.

because of the complementarity between goods invested and parental time. When parents receive a contribution in goods invested, they have an incentive to increase the time they invest in their offspring. Private parental investment in goods is crowded out by the direct expenditure policy, decreasing by more than 14% in all phases of childhood.

While the Kindergeld is flat by construction, both optimal transfers are decreasing with the age of the child and with the amount of resources available to the household. I break down the changes in parental investment that occur under the Kindergeld experiment in Table VIII to understand why optimal transfers take this form and which mechanisms are determining the policy returns I simulate.

In early childhood, a flat transfer has the strongest impact on low-income households, and on households with low parental or offspring's skills. The larger impact on low-income and low-skilled households is due to the fact that the same transfer has a stronger impact on labor supply for households who earn a low wage on the labor market, which increases time available for leisure and for investment in children. The gradient in parental skills is weaker because the elasticity of the wage to skills is lower than 1, and because higher-skilled households are more productive at raising skillful children, thus they have an incentive to invest relatively more. Nevertheless, the fact that higher-skilled parents have higher income dominates, thus high-skilled households increase investment by less. The larger impact on households who have low-skilled offspring is instead due to the elasticity of substitution between inputs in early childhood. As additional resources become available, remediating adverse initial conditions is easier than improving an already skilled offspring.

In late childhood the income gradient in the transfer's impact remains the same, but the magnitude of the impact on time invested reduces to less than half. This is because investment in late childhood has smaller returns, thus parents have a smaller incentive to change their behavior when the transfer is fixed. The gradient on initial offspring's skills disappears, as shocks and the pattern of investment have made the initial conditions of the offspring less relevant.

Policy returns are highly heterogeneous over the income distribution: increases in final cognitive skills are larger among low-income households and high-skilled households.

Overall, this implies that the optimal transfer should be decreasing in the age of the offspring, in order to reap the benefits from the higher productivity of early childhood, and decreasing in the wage faced by the household, in order to focus on impacting low-income households who are most affected by the additional resources. In fact, I find that both the optimal constrained transfer and the optimal constrained increase in investment expenditure follow this intuition.

Finally, I compare the policy results of the constrained policies (the Kindergeld, the age-and-income-dependent transfer and the age-and-income-dependent expenditure) under the baseline model with policy returns implied by the cognitive-skills only model. Results are

	D1	Q1	Q2	Q3	Q4
Time Invested (1st stg.)					
by Initial Cog Skill	+ 1.51	+ 1.50	+ 1.49	+ 1.47	+ 1.45
by Initial NonCog Skill	+ 1.48	+ 1.48	+ 1.48	+ 1.48	+ 1.48
by Parental Cog Skill	+ 1.90	+ 1.80	+ 1.57	+ 1.39	+ 1.15
by Parental NonCog Skill	+ 1.65	+ 1.61	+ 1.52	+ 1.43	+ 1.35
by Parental Income	+ 3.00	+ 2.46	+ 1.56	+ 1.14	+ 0.75
by Residual Wage	+ 2.90	+ 2.40	+ 1.56	+ 1.16	+ 0.78
Time Invested (2nd stg.)					
by Initial Cog Skill	+ 0.58	+ 0.58	+ 0.58	+ 0.58	+ 0.58
by Initial NonCog Skill	+ 0.58	+ 0.58	+ 0.58	+ 0.58	+ 0.58
by Parental Cog Skill	+ 0.68	+ 0.66	+ 0.61	+ 0.56	+ 0.48
by Parental NonCog Skill	+ 0.63	+ 0.62	+ 0.59	+ 0.57	+ 0.54
by Parental Income	+ 1.16	+ 0.95	+ 0.61	+ 0.45	+ 0.30
by Residual Wage	+ 1.17	+ 0.96	+ 0.61	+ 0.45	+ 0.30
Investment (1st stg.)					
by Initial Cog Skill	+ 5.39%	+ 5.36%	+ 5.32%	+ 5.26%	+ 5.20%
by Initial NonCog Skill	+ 5.30%	+ 5.28%	+ 5.28%	+ 5.29%	+ 5.29%
by Parental Cog Skill	+ 6.56%	+ 6.27%	+ 5.60%	+ 5.02%	+ 4.25%
by Parental NonCog Skill	+ 5.83%	+ 5.71%	+ 5.41%	+ 5.14%	+ 4.89%
by Parental Income	+ 10.12%	+ 8.45%	+ 5.61%	+ 4.22%	+ 2.85%
by Residual Wage	+ 9.87%	+ 8.31%	+ 5.60%	+ 4.27%	+ 2.96%
Investment (2nd stg.)					
by Initial Cog Skill	+ 3.84%	+ 3.81%	+ 3.82%	+ 3.81%	+ 3.79%
by Initial NonCog Skill	+ 3.78%	+ 3.78%	+ 3.79%	+ 3.82%	+ 3.84%
by Parental Cog Skill	+ 4.24%	+ 4.18%	+ 4.00%	+ 3.74%	+ 3.31%
by Parental NonCog Skill	+ 4.08%	+ 4.01%	+ 3.86%	+ 3.74%	+ 3.63%
by Parental Income	+ 6.93%	+ 5.88%	+ 4.06%	+ 3.12%	+ 2.16%
by Residual Wage	+ 7.10%	+ 5.97%	+ 4.04%	+ 3.09%	+ 2.14%
Final Cog Skills					
by Initial Cog Skill	+ 3.22%	+ 3.19%	+ 3.21%	+ 3.18%	+ 3.14%
by Initial NonCog Skill	+ 3.12%	+ 3.14%	+ 3.16%	+ 3.19%	+ 3.23%
by Parental Cog Skill	+ 2.32%	+ 2.62%	+ 3.22%	+ 3.46%	+ 3.42%
by Parental NonCog Skill	+ 3.27%	+ 3.24%	+ 3.19%	+ 3.16%	+ 3.13%
by Parental Income	+ 5.12%	+ 4.57%	+ 3.41%	+ 2.76%	+ 1.98%
by Residual Wage	+ 6.24%	+ 5.16%	+ 3.37%	+ 2.51%	+ 1.69%
Final NCog Skills					
by Initial Cog Skill	+ 3.92%	+ 3.89%	+ 3.88%	+ 3.89%	+ 3.82%
by Initial NonCog Skill	+ 3.69%	+ 3.72%	+ 3.83%	+ 3.92%	+ 4.01%
by Parental Cog Skill	+ 5.05%	+ 4.75%	+ 4.09%	+ 3.62%	+ 3.02%
by Parental NonCog Skill	+ 4.04%	+ 3.97%	+ 3.92%	+ 3.82%	+ 3.77%
by Parental Income	+ 7.77%	+ 6.39%	+ 4.10%	+ 3.01%	+ 1.98%
by Residual Wage	+ 7.53%	+ 6.24%	+ 4.11%	+ 3.07%	+ 2.07%

Table VIII: Kindergeld experiment: changes from baseline across the skills and income distribution. Changes in skills are reported as percentage of a standard deviation of baseline skills. Changes in time invested and hours worked are reported in hours.

summarized in Table IX.

When noncognitive skills are not accounted for, all policies considered are substantially

	Kindergeld (1)	Transfer, Age + w (2)	Expenditure, Age + w (3)
Parametrization			
β_t	0.00	-0.27	-0.06
β_w	0.00	-0.51	-0.13
Final Cog Skills			
Mean	+ 0.84%	+ 0.52%	+ 2.17%
Std. Dev.	+ 0.54%	+ 0.04%	+ 0.75%
Correlation($s_{C,T+1}, s_{C,P}$)	+ 0.03%	- 0.04%	- 0.10%
Correlation($s_{C,T+1}$, HH Income)	- 0.79%	- 1.14%	- 3.92%
Time Invested, Avg (1st stg.)	+ 0.79	+ 2.03	+ 2.44
Time Invested, Avg (2nd stg.)	+ 0.44	+ 0.39	+ 1.34
Goods Invested, Avg (1st stg.)	+ 1.08%	+ 2.62%	- 26.35%
Goods Invested, Avg (2nd stg.)	+ 0.72%	+ 0.62%	- 21.53%
Portion of Transfer Consumed (avg)	97.84%	96.34%	
Consumption, Avg.	+ 3.73%	+ 3.75%	+ 1.38%
Hours Worked, Avg. (1st stg.)	- 2.24	- 5.48	- 2.90
Hours Worked, Avg. (2nd stg.)	- 1.96	- 1.95	- 1.90

Table IX: Policy experiments: average changes from base case, **Cognitive-Skills-Only** model. Column (1): flat transfer of 5% of average income. Column (2): transfer worth 5% of average income, optimally allocated according to household wage and age of the child. Column (3): direct increase in expenditure in children worth 5% of average household income, optimally allocated according to household wage and age of the child. Changes in skills are reported as percentage of a standard deviation of baseline skills. Changes in time invested and hours worked are reported in hours.

less effective at raising children's skills, while still decreasing labor supply. The cognitive-skills-only model implies returns to policies that are about one-fourth of those of the two-skills model. This result highlights the importance of accounting for both cognitive and noncognitive skills when simulating the impact of policies aimed at fostering early skills accumulation.

5.1 More Flexible Policies

I now explore how much is lost by not allowing policies to vary with all state variables. I allow transfers (and expenditure increases) to take the form

$$\tau_{i,t} = \exp(\beta_0 + \beta_t t + \beta_w \text{Wage}_{i,t} + \beta_C s_{C,t} + \beta_N s_{N,t} + \beta_{CP} s_{C,P} + \beta_{NP} s_{N,P}) \,, \tag{15}$$

so that policies are allowed to vary with all state variables and with the age of the child. Results are summarized in Table X.

Unfortunately, the approximation algorithm limits the degree of flexibility of these policies, since parameter values that are too large produce heavy nonlinearities that the approximation does not handle well. The approximation stops being reliable once transfers for the lowest income families (bottom 1%) become more than 500 times larger than the average transfer in the economy. Thus, I restrict these policies to have a less steep gradient.

The more flexible policies improve over the best of the constrained policies. Final cognitive skills increase by more than 8.5% in the case of the optimal transfer, and by more than 8% in the case of the optimal expenditure increase. However, this means that the constrained policy gets quite close to the best optimal policy, achieving more than three-quarters of the increase in skills obtainable through a more flexible scheme. This result is interesting because, while observing a child's skills in all periods can be very costly for a policy-maker, the age of a child and household income are variables that are commonly collected. Thus, my results suggest that reasonable efficiency of child allowance policies can be attained without excessive costs associated to additional collection of information.

Interestingly, when allowing policies to be more flexible, the transfer becomes preferable to the direct increase in expenditure. While allowing the transfer to be more flexible allows to target unskilled households with highly skilled children, which benefit the most from the extra time available to parents, targeting the increase in expenditure seems to be limited by the extent to which such increase crowds out private expenditure.

6 Conclusions

I develop a model of skills formation and household choices, grounded in the literature on Cognitive and NonCognitive Skills, and I show that it can help explaining several stylized facts on child care time and cognitive/noncognitive achievement. I use the model to simu-

	Expenditure, Age + w (1)	Transfer All States (2)	Expenditure, All States (3)
Parametrization			
β_t	-0.06	0.06	-0.12
β_w	-0.13	0.01	-0.07
β_C	0.00	0.50	0.42
β_N	0.00	0.50	0.32
β_{CP}	0.00	-0.44	0.20
β_{NP}	0.00	-0.02	-0.13
Final Cog Skills			
Mean	+ 6.40%	+ 8.63%	+ 8.15%
Std. Dev.	+ 2.18%	+ 5.16%	+ 5.12%
Final NonCog Skills			
Mean	+ 9.09%	+ 11.81%	+ 10.39%
Std. Dev.	+ 4.68%	+ 11.06%	+ 7.68%
Correlation($s_{C,T+1}, s_{C,P}$)	- 5.47%	- 5.51%	- 0.48%
Correlation($s_{N,T+1}, s_{N,P}$)	- 14.21%	- 2.59%	- 9.56%
Correlation($s_{C,T+1}$, HH Income)	- 38.83%	- 27.00%	- 34.86%
Correlation($s_{N,T+1}$, HH Income)	- 57.73%	- 43.43%	- 49.02%
Time Invested, Avg (1st stg.)	+ 1.55	+ 4.56	+ 2.43
Time Invested, Avg (2nd stg.)	+ 0.21	+ 1.68	+ 0.47
Goods Invested, Avg (1st stg.)	- 17.40%	+ 3.36%	- 19.66%
Goods Invested, Avg (2nd stg.)	- 14.88%	+ 2.04%	- 17.53%
Portion of Transfer Consumed (avg)		92.69%	
Consumption, Avg.	+ 1.22%	+ 1.39%	+ 0.92%
Hours Worked, Avg. (1st stg.)	- 1.81	- 4.07	- 2.41
Hours Worked, Avg. (2nd stg.)	- 0.82	- 3.08	- 1.05

Table X: Changes from baseline to optimal policies depending on all states. All policies sum to 5% of total household income. Column (1): Best constrained expenditure policy, depending on wage and age of the child only. Column (2): transfer worth 5% of average income, optimally allocated according to household income, age of the child, offspring's skills and parental skills. Column (3): direct increase in expenditure in children, optimally allocated according to income, age of the child, offspring's skills and parental skills. Changes in skills are reported as percentage of a standard deviation of baseline skills. Changes in time invested and hours worked are reported in hours.

late the effect of applying different policies to the US economy and find that even the less sophisticated policy available (a flat transfer) is effective at increasing children's skills by a significant amount (+3%). When policies are allowed to vary with the age of the child and with the income of the household, returns become substantially larger (+6%) and even larger if further dependency on parental and children's skills is allowed. Also, such policies are effective at reducing intergenerational persistence in several dimensions. Finally, my results suggest that returns to transfer policies are substantially heterogeneous across households and higher among low-income and high-skilled households.

I find that accounting for noncognitive skills is essential to get the correct returns to policies. When noncognitive skills are neglected, simulated policies are ineffective at increasing children's skills and reducing inequality.

While the framework developed in this paper accounts for several channels of persistency and inequality in skills accumulation, it abstracts from other sources of dispersion in parental investment such as income shocks, borrowing constraints and imperfect information about the child's skills. Future avenues of research might include incorporating these ingredients in the study of policy impacts on cognitive and noncognitive skills accumulation.

References

Agostinelli, Francesco, & Wiswall, Matthew. 2017. Estimating the Technology of Skills Formation. *working paper.*

Aguiar, Mark, & Hurst, Erik. 2007. Measuring Trends in Leisure: The Allocation of Time over Five Decades. *Quarterly Journal of Economics*, **122**(3).

Allard, Mary Dorinda, Bianchi, Suzanne M., Stewart, Jay, & Wight, Vanessa R. 2004. Inequality in Parental Investment in Child-Rearing: Expenditures, Time, and Health. *In:* Neckerman, Kathryn M. (ed), *Social Inequality*. New York: Russel Sage Foundation.

Allard, Mary Dorinda, Bianchi, Suzanne M., Stewart, Jay, & Wight, Vanessa R. 2007. Comparing Childcare measures in the ATUS and in Earlier Time-Diary Studies. *Monthly Labor Review*, **130**(5).

Anger, Silke, & Heineck, Guido. 2009. Do Smart Parents Raise Smart Children? The Intergenerational Transmission of Cognitive Abilities. *Journal of Population Economics*, **23**(3).

Becker, Gary S., & Tomes, Nigel. 1976. Child Endowments and the Quantity and Quality of Children. *Journal of Political Economy*, **84**(4).

Becker, Gary S., & Tomes, Nigel. 1979. An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility. *Journal of Political Economy*, **87**(6).

Bernal, Raquel, & Keane, Michael P. 2010. Quasi-Structural Estimation of a Model of Child Care Choices and Child Cognitive Ability Production. *Journal of Econometrics*, **156**.

Bernal, Raquel, & Keane, Michael P. 2011. Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers. *Journal of Labor Economics*, **29**(3).

Bianchi, Suzanne M. 2000. Maternal Employment and Time with Children: Dramatic Change or Surprising Continuity? *Demography*, **37**(4).

Boca, Daniela Del, Flinn, Christopher, & Wiswall, Matthew. 2014. Household Choices and Child Development. *Review of Economics Studies*, **81**.

Brilli, Ylenia. 2012. Mother or Market Care? A Structural Estimation of Child Care Impacts on Child Development. *Working Paper*.

Carlson, Marcia J., & Corcoran, Mary E. 2001. Family Structure and Children's Behavioral and Cognitive Outcomes. *Journal of Marriage and Family*.

Caucutt, Elizabeth M., & Lochner, Lance. 2012. Early and Late Human Capital Investments, Borrowing Constraints, and the Family. *working paper*.

Coleman, J. S. 1966. Equality of Educational Opportunity. *Washington DC, US dept. of Health, Education and Welfare, Office for Education*.

Cunha, Flavio, & Heckman, James J. 2007. The Technology of Skill Formation. *American Economic Review*.

Cunha, Flavio, Heckman, James J., & Schennach, Susanne M. 2010. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*, **78**(3).

Daruich, Diego. 2017. From Childhood to Adult Inequality: Parental Investments and Early Childhood Development. *Working Paper*.

Eika, Lasse, Mogstad, Magne, & Zafar, Basit. 2017. Educational Assortative Mating and Household Income Inequality. *Working Paper*.

Erosa, Andrès, Fuster, Luisa, & Kambourov, Guergoui. 2012. Towards a micro-founded theory of aggregate labor supply. *Working Paper*.

Flynn, James R. 2009. *What is Intelligence? Beyond the Flynn Effect*. Cambridge University Press.

Griffen, Andrew. accepted. Evaluating the Effects of Child Care Policies on Children's Cognitive Development and Maternal Labor Supply. *Journal of Human Resources*.

Guryan, Jonathan, Hurst, Erik, & Kearney, Melissa. 2008. Parental Education and Parental Time with Children. *Journal of Economic Perspectives*, **22**(3).

Guvenen, Fatih, & Kuruscu, Burhanettin. 2009. A Quantitative Analysis of the Evolution of the U.S. Wage Distribution: 1970-2000. *NBER Macroeconomics Annual*, **24**.

Hanushek, Eric A., & Woessmann, Ludger. 2008. The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*.

Heckman, James J., Moon, Seong Hyeok, Pinto, Rodrigo, Savelyev, Peter A., & Yavitz, Adam. 2010a. The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, **94**(1-2).

Heckman, James J., Malofeeva, Lena, Pinto, Rodrigo, & Savelyev, Peter A. 2010b. Understanding the Mechanisms through which an Influential Early Childhood Program Boosted Adult Outcomes. *Unpublished Manuscript*.

Helmers, Christian, & Patnam, Manasa. 2011. The formation and evolution of childhood skill acquisition: Evidence from India. *Journal of Development Economics*.

Huggett, Mark, Ventura, Gustavo, & Yaron, Amir. 2011. Sources of Lifetime Inequality. *American Economic Review*, **101**(7).

Keane, Michael P. 2011. Labor Supply and Taxes: A Survey. *Journal of Economic Literature*, **49**(4).

Keane, Michael P., & Wolpin, Kenneth I. 1996. The Career Decisions of Young Men. *Journal of Political Economy*, **105**(3).

Lee, Sang Yoon, & Seshadri, Ananth. forthcoming. On the Intergenerational Transmission of Economic Status. *Journal of Political Economy*.

McGrattan, Ellen R., & Rogerson, Richard. 2004. Changes in Hours Worked, 1950-2000. *Federal Reserve Bank of Minneapolis Quarterly Review*, **28**(1).

Osuna, Victoria, & Rios-Rull, Josè Victor. 2003. Implementing the 35 hours workweek by Means of Overtime Taxation. *Review of Economic Dynamics*, **6**(1).

Pistaferri, Luigi. 2003. Anticipated and Unanticipated Wage Changes, Wage Risk, and Intertemporal Labor Supply. *Journal of Labor Economics*, **21**(3).

Ramey, Garey, & Ramey, Valerie A. 2010. The Rug Rat Race - Comments and Discussion. *Brookings Papers on Economic Activity*.

Restuccia, Diego, & Urrutia, Carlos. 2004. Intergenerational Persistence of Earnings; The Role of Early and College Education. *The American Economic Review*, **94**(5).

Robinson, John P., & Godbey, Geoffrey. 1999. *Time for Life: The Surprising Ways Americans Use Their Time, 2nd Edition*. Pennsylvania State University Press.

Sandberg, John F., & Hofferth, Sandra L. 2001. Changes in children's time with parents: United States, 1981-1997. *Demography*, **38**(3).

Todd, Petra E., & Wolpin, Kenneth I. 2007. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*.

Youderian, Xiaoyan. 2016. Human Capital Production with Parental Time Investment in Early Childhood. *Working Paper*.

Yum, Minchul. 2018. Parental Time Investment and Intergenerational Mobility. *Working Paper*.

Appendix A Analytical Results

First order conditions for consumption c_t give

$$c_t^{-\theta} = \lambda_t$$

where λ_t is the multiplier associated to the budget constraint of the household. First order conditions with respect to labor time n_t gives

$$(1 - n_t - x_t) = \left(\frac{\zeta}{w\lambda_t}\right)^{1/\sigma}$$

substituting the first equation inside the second yields that households trade off leisure and consumption according to the equation

$$(1 - n_t - x_t) = \left(\frac{\zeta c_t^{\theta}}{w}\right)^{1/\sigma}$$

Taking first order conditions with respect to e_t yields

$$\lambda_t = \mu_t A(1-\alpha)e_t^{\omega-1}\left(\alpha x_t^{\omega} + (1-\alpha)e_t^{\omega}\right)^{\frac{1-\omega}{\omega}}$$

substituting the multiplier λ_t and dividing the last expression by equation 7 yields

$$e_t = \left[(\frac{1-\alpha}{\alpha})w\right]^{\frac{1}{1-\omega}} x_t \tag{16}$$

A.1 Relations between investment and parameters

Consider the multipliers μ_t associated to the constraint $I_t = Ax_t^{\alpha}e_t^{1-\alpha}$ as the function

$$\mu_t = \mu_t(K_t, S_t, x_t)$$

where the function expresses the marginal productivity of investment as a function of the stage t, of the parameters K_t encompassing the efficiency of investment[26], of the state

[26]In practice, many coefficients may enter K_t; for instance, in a Cobb-Douglas specification of the technology of skills formation, K_t includes the scale of the function and the exponent of the investment variable. In a CES specification, K_t includes the coefficient that multiplies the investment variable inside the CES

S_t and of the amount of time spent with the child x_t. By definition $\partial\mu_t/\partial K_t > 0$; moreover, since the technology has decreasing returns to every single input, $\partial\mu_t/\partial x_t < 0$.

Proposition 1 (1)**.** *Suppose that*

- ***A1****:* $\alpha > 1/2$*;*

then we have that $\frac{\partial x_t}{\partial K_t} > 0$*, that is, households respond to increased productivity in investment by increasing time invested in the offspring.*

Furthermore, if we have that

- ***A2****: Preferences satisfy Balanced Growth Path, that is* $\theta = 1$*;*

- ***A3****:* $\sigma \in [0,1], \zeta > \frac{1-\alpha}{\alpha\sigma}$*;*

then $\frac{\partial n_t}{\partial K_t} < 0$*, that is households respond to higher productivity in investment by decreasing hours of work.*

Proof. First of all, by implicit function theorem we have

$$\frac{\partial F}{\partial n_t}\frac{\partial n_t}{\partial K_t} + \frac{\partial F}{\partial x_t}\frac{\partial x_t}{\partial K_t} = -\frac{\partial F}{\partial K_t}$$

Call equation 5 F_1 and equation 6 F_2; substituting expression 8 for e_t and the budget constraint in equation 5 and applying the result above yields

$$\left[-\sigma\zeta(1-n_t-x_t)^{\sigma-1} - \theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}\right]\frac{\partial n_t}{\partial K_t} +$$
$$\left[-\sigma\zeta(1-n_t-x_t)^{\sigma-1} + \frac{1-\alpha}{\alpha}\theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}\right]\frac{\partial x_t}{\partial K_t} = 0$$
$$\left[\sigma\zeta(1-n_t-x_t)^{-\sigma-1}\right]\frac{\partial n_t}{\partial K_t} + \left[\sigma\zeta(1-n_t-x_t)^{-\sigma-1}\right]\frac{\partial x_t}{\partial K_t} = (\frac{\partial\mu_t}{\partial K_t} - \frac{\partial\mu_t}{\partial x_t})A\alpha^\alpha w^{1-\alpha}(1-\alpha)^{1-\alpha}$$

Trivial manipulation of the two equations yields

aggregator.

$$\frac{\partial n_t}{\partial K_t} = -\frac{\partial x_t}{\partial K_t}\frac{\sigma\zeta(1-n_t-x_t)^{\sigma-1} - \frac{1-\alpha}{\alpha}\theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}}{\sigma\zeta(1-n_t-x_t)^{\sigma-1} + \theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}}$$
$$\frac{\partial n_t}{\partial K_t} = -\frac{\partial x_t}{\partial K_t} + (\frac{\partial \mu_t}{\partial K_t} - \frac{\partial \mu_t}{\partial x_t})\frac{A\alpha^\alpha w^{1-\alpha}(1-\alpha)^{1-\alpha}(1-n_t-x_t)^{\sigma+1}}{\sigma\zeta}$$

Solving the system for $\frac{\partial x_t}{\partial K_t}$ we have

$$\frac{\partial x_t}{\partial K_t}\left[1 - \frac{\sigma\zeta(1-n_t-x_t)^{\sigma-1} - \frac{1-\alpha}{\alpha}\theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}}{\sigma\zeta(1-n_t-x_t)^{\sigma-1} + \theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}}\right] =$$
$$(\frac{\partial \mu_t}{\partial K_t} - \frac{\partial \mu_t}{\partial x_t})\frac{A\alpha^\alpha w^{1-\alpha}(1-\alpha)^{1-\alpha}(1-n_t-x_t)^{\sigma+1}}{\sigma\zeta}$$

It is clear that the right hand side is always positive, being the product of only positive terms, because $\frac{\partial \mu_t}{\partial K_t} > 0$ by construction and $\frac{\partial \mu_t}{\partial x_t} < 0$ because of the CES form of the skills formation technology and of the utility of skills. Hence, we have that $\frac{\partial x_t}{\partial K_t} > 0$ if and only if

$$\frac{\sigma\zeta(1-n_t-x_t)^{\sigma-1} - \frac{1-\alpha}{\alpha}\theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}}{\sigma\zeta(1-n_t-x_t)^{\sigma-1} + \theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}} < 1$$

and since leisure $(1-n_t-x_t)$ is always positive and consumption $wn_t - w(\frac{1-\alpha}{\alpha})x_t$ is always bigger than zero (by the fact that marginal utility approaching zero is infinity), this can be trivially shown to be equivalent to

$$\alpha > 1/2$$

which proves the first statement. Suppose now that $\alpha > 1/2$, that is $\frac{\partial x_t}{\partial K_t} > 0$; by the previous equation,

$$\frac{\partial n_t}{\partial K_t} = -\frac{\partial x_t}{\partial K_t}\left[\frac{\sigma\zeta(1-n_t-x_t)^{\sigma-1} - \frac{1-\alpha}{\alpha}\theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}}{\sigma\zeta(1-n_t-x_t)^{\sigma-1} + \theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1}}\right]$$

Since the denominator is positive, $\frac{\partial n_t}{\partial \mu_t} < 0$ if and only if

$$\sigma\zeta(1-n_t-x_t)^{\sigma-1} > \frac{1-\alpha}{\alpha}\theta(wn_t - w(\frac{1-\alpha}{\alpha})x_t)^{\theta-1} \tag{17}$$

and in the case of Balanced Growth path ($\theta = 1$), a sufficient condition for 17 to hold is that $\sigma \in [0,1], \zeta > \frac{1-\alpha}{\alpha\sigma}$, completing the proof. $\square$

It is not possible to state the proposition generally for any value of the concavity of leisure and any risk aversion parameter; however, condition 17 suggests that if the share of goods $1-\alpha$ in the investment function is low and ζ is higher than 1, the solution should satisfy the condition in the majority of cases for reasonable values of labor supply and time spent with children, but may possibly be violated in extreme cases.

Appendix B The Technology of Skills Formation

Covariance Matrix				
	Child's Cog. Skills at birth	Child's NonCog. Skills at Birth	Mother's Cog. Skills	Mother's Noncog. Skills
Child Cog. Skills	0.1777			
Child NonCog. Skills	-0.0204	0.2002		
Mother's Cognitive	0.0182	0.0592	0.5781	
Mother's Noncognitive	0.0050	0.0261	0.0862	0.0667

Correlation Matrix				
	Child's Cog. Skills at birth	Child's NonCog. Skills at Birth	Mother's Cog. Skills	Mother's Noncog. Skills
Child Cog. Skills	1.0000			
Child NonCog. Skills	-0.1081	1.0000		
Mother's Cognitive	0.0569	0.1741	1.0000	
Mother's Noncognitive	0.0463	0.2260	0.4390	1.0000

Table XI: Variance/Covariance Matrix and Correlation Matrix for initial conditions of parental and offspring's skills [source: Cunha, Heckman, Schennach (2010), Appendix].

Technology of **Cognitive** Skills Formation				
		1st Stage		2nd Stage
Cog Skills	$\gamma_{1,C,1}$	0.485	$\gamma_{2,C,1}$	0.884
NonCog Skills	$\gamma_{1,C,2}$	0.062	$\gamma_{2,C,2}$	0.011
Investment	$\gamma_{1,C,3}$	0.261	$\gamma_{2,C,3}$	0.044
Parental Cog	$\gamma_{1,C,4}$	0.035	$\gamma_{2,C,4}$	0.051
Parental NonCog	$\gamma_{1,C,5}$	0.157	$\gamma_{2,C,5}$	0.011
Complementarity	$\phi_{1,C}$	0.585	$\phi_{2,C}$	-1.220
Elasticity of Substitution $1/(1-\phi)$		2.409		0.450
Variance of Shocks	$\eta_{1,C}$	0.165	$\eta_{2,C}$	0.098
Technology of **NonCognitive** Skills Formation				
		1st Stage		2nd Stage
Cog Skills	$\gamma_{1,N,1}$	0.000	$\gamma_{2,N,1}$	0.002
NonCog Skills	$\gamma_{1,N,2}$	0.602	$\gamma_{2,N,2}$	0.857
Investment	$\gamma_{1,N,3}$	0.209	$\gamma_{2,N,3}$	0.104
Parental Cog	$\gamma_{1,N,4}$	0.014	$\gamma_{2,N,4}$	0.000
Parental NonCog	$\gamma_{1,N,5}$	0.175	$\gamma_{2,N,5}$	0.037
Complementarity	$\phi_{1,N}$	-0.464	$\phi_{2,N}$	-0.522
Elasticity of Substitution $1/(1-\phi)$		0.683		0.657
Variance of Shocks	$\eta_{1,N}$	0.203	$\eta_{2,N}$	0.102

Table XII: The Technology for Cognitive and Noncognitive Skill Formation; estimated by Cunha, Heckman, Schennach (2010) [pag. 919] taking into account Investment endogeneity; skills linearly anchored to educational attainment; factors normally distributed, standard errors in parentheses.

Appendix C Summary Statistics - Skill Factors

	Mean	Std. Dev.	Skewness	N	R^2 of factor
Child's Cognitive Factor, Age 5-6					
Peabody Picture Vocabulary	0.475	0.906	-0.103	809	31.4 %
PIAT Math	0.271	1.039	0.886	1101	37.9 %
PIAT Reading Recognition	0.246	1.015	1.466	1074	96.5 %
PIAT Reading Comprehension	0.240	0.978	1.294	1025	95.1 %
Child's Noncognitive Factor, Age 5-6					
Behavior Problem Index/ Antisocial Raw Score	0.092	0.937	-1.107	1453	55.9 %
Behavior Problem Index/ Anxiety Raw Score	-0.066	1.067	-0.820	1461	49.9 %
Behavior Problem Index/ Headstrong Raw Score	-0.098	0.996	-0.039	1462	72.3 %
Behavior Problem Index/ Hyperactive Raw Score	0.010	0.972	-0.417	1461	58.1 %
Behavior Problem Index/ Conflict Raw Score	0.064	0.905	-1.882	1463	41.1 %
Child's Cognitive Factor, Age 13-14					
PIAT Math	0.424	0.921	-0.220	1063	64.5 %
PIAT Reading Recognition	0.336	0.876	-0.639	1064	78.8 %
PIAT Reading Comprehension	0.427	0.937	-0.270	1056	72.4 %
Child's Noncognitive Factor, Age 13-14					
Behavior Problem Index/ Antisocial Raw Score	0.117	0.971	-1.148	1125	63.5 %
Behavior Problem Index/ Anxiety Raw Score	-0.088	1.053	-0.595	1138	64.8 %
Behavior Problem Index/ Headstrong Raw Score	-0.07	0.998	0.002	1143	68.3 %
Behavior Problem Index/ Hyperactive Raw Score	0.044	0.974	-0.715	1138	59.3 %
Behavior Problem Index/ Conflict Raw Score	-0.024	1.033	-1.577	1142	52.4 %
MOTHER's Cognitive Factor					
Mom's Arithmetic Reasoning Test Score	0.172	0.933	0.168	2207	83.7 %
Mom's Word Knowledge Test Score	0.302	0.822	-0.836	2207	70.9 %
Mom's Paragraph Composition Test Score	0.377	0.827	-1.121	2207	66.0 %
Mom's Numerical Operations Test Score	0.343	0.875	-0.469	2207	54.7 %
Mom's Coding Speed Test Score	0.468	0.879	-0.445	2207	41.1 %
Mom's Mathematical Knowledge Test Score	0.185	0.972	0.269	2207	77.4 %
MOTHER's NonCognitive Factor					
Mom's Self-Esteem: "I am a person of worth"	3.534	0.516	-0.343	2207	43.1 %
Mom's Self-Esteem: "I have good qualities"	3.382	0.531	0.025	2207	48.5 %
Mom's Self-Esteem: "I am a failure"	3.477	0.580	-0.649	2207	52.9 %
Mom's Self-Esteem: "I am as capable as others"	3.326	0.549	-0.217	2207	36.7 %
Mom's Self-Esteem: "I have nothing to be proud of"	3.480	0.625	-1.082	2207	46.0 %
Mom's Self-Esteem: "I have a positive attitude"	3.200	0.576	-0.250	2207	51.6 %
Mom's Self-Esteem: "I wish I had more self-respect"	2.876	0.787	-.206	2207	38.2 %
Mom's Self-Esteem: "I feel useless at times"	2.650	0.774	0.300	2207	32.5 %
Mom's Self-Esteem: "I sometimes think I am no good"	3.005	0.808	-0.298	2207	41.9 %
Mom's Rotter Score: "I have no control"	2.897	1.156	-0.600	2207	5.5 %
Mom's Rotter Score: "I make no plans for the future"	2.543	1.159	-0.002	2207	8.1 %
Mom's Rotter Score: "Luck is big factor in life"	3.154	0.974	-1.107	2207	4.5 %
Mom's Rotter Score: "Luck plays big role in my life"	2.426	1.144	-0.025	2207	2.5 %

Table XIII: Summary Statistics of Variables used to identify latent Cognitive and Noncognitive Factors. Source: data extract from NLSY79/CNSLY79 by Cunha, Heckman and Schennach (2010).

Appendix D Solution Algorithm

The algorithm uses first order conditions when possible and solves the household's problem backwards, using polynomial approximation of next period's value function. Shocks are approximated with a three-state symmetric shock, that has mean zero and variance as in Table XII. The reason is that shocks are independent, hence for n-state shocks the next period's value function must be computed n^2 times, greatly increasing the computational burden. A brief description of the algorithm follows:

1. Start from final period $t = T + 1$; extract 10^5 points in the continuous state space of $S_{T+1} = \{s_{C,T+1}, s_{N,T+1}, s_{C,P}, s_{N,P}, \epsilon_P^w\}$, distributed as uniform Sobol numbers from -10 to +10 standard deviations of the covariance matrix in Table XI. The algorithm oversamples the tails on purpose (the true distribution is jointly normal at the initial period, and roughly normal in later periods), because it is more difficult to approximate the value function far from the median. Also, simulations show that the variance of offspring's skills increases by approximately 70 % from period 1 to period $T + 1$, so it is important to consider a more dispersed final distribution. I use Sobol numbers because they span the state space more efficiently than uniform random numbers.

2. Given the state, the solution at period t is found as follows. The objective function at time t, before maximization, is

$$Q_t(S_t, c_t, n_t, e_t, x_t) = \frac{c_t^{1-\theta}}{1-\theta} + \zeta\frac{(1-n_t-x_t)^{1-\sigma}}{1-\sigma} + W(s_{C,t}, s_{N,t}) + \beta\mathbb{E}\Big[V_{t+1}(S_{t+1})\Big]$$

   which is a function of the state $S_t = \{s_{C,t}, s_{N,t}, s_{C,P}, s_{N,P}, \epsilon_P^w\}$ and of the controls c_t, n_t, x_t, e_t. Clearly, $V_t = \max_{\{c_t,n_t,x_t,e_t\}} Q_t(S_t, c_t, n_t, x_t, e_t)$, subject to all the constraints of the dynamic problem in section 3. In all periods, the choices x_t and e_t map offspring's skills $s_{C,t}, s_{N,t}$ into next period's skills $s_{C,t+1}, s_{N,t+1}$, hence in the state S_{t+1} and in the value function $V_{t+1}(S_{t+1})$. Then it is necessary to predict next period's value function.

At period $T+1$ such prediction is unnecessary, because $V_{T+2}=0$. Assume that the map from the state S_{t+1} to the value function $V_{t+1}(S_{t+1})$ is known.

3. Start with a guess x_t. Given x_t and the state S_t, one of the controls e_t is the explicit solution to equation 7 in the optimum; call it $e_t^*(x_t)$. Given $e_t^*(x_t)$, consumption c_t can be backed out from the budget constraint given the labor choice n_t. Then n_t is the solution to the equation 5, which is solved by bisection method; denote these last two choices as $c_t^{**}(x_t)$,$n_t^{**}(x_t)$. As a result, given the state S_t, now Q_t is only a function of the state S_t and of the choice of x_t, because all other controls are functions of x_t. Denote this objective function as $Q_t(S_t, c_t^{**}(x_t), n_t^{**}(x_t), e_t^*(x_t), x_t)$. Notice that $Q_t(S_t, c_t^{**}(x_t), n_t^{**}(x_t), e_t^*(x_t), x_t)$ is not equal to $\max_{c_t,n_t,e_t} Q_t(S_t, c_t, n_t, e_t, x_t)$ for all x_t, except in the optimum x_t^*, because $c_t^{**}, n_t^{**}, e_t^*$ are not the optimal controls given a suboptimal x_t since they exploit the first order conditions of the problem, which are jointly true only in the optimum. However,

$$Q_t(S_t, c_t^{**}(x_t), n_t^{**}(x_t), e_t^*(x_t), x_t) \leq \max_{c_t,n_t,e_t} Q_t(S_t, c_t, n_t, e_t, x_t)$$

because, for given x_t, all other controls are suboptimal. Finally,

$$\max_{x_t} Q_t(S_t, c_t^{**}(x_t), n_t^{**}(x_t), e_t^*(x_t), x_t) = \max_{c_t,n_t,e_t,x_t} Q_t(S_t, c_t, n_t, e_t, x_t)$$

which ensures that the solution is the same. Then, $\tilde{V}_t(x_t, S_t)$ is maximized on x_t using golden search. The algorithm gives the solution at arbitrary precision for every state point; the chosen precision is 10^{-7}. One possible concern is that, given the transformation described above, the function might lose single-peakedness, which is a strict requirement of golden search. Using another algorithm, i.e. grid-based maximization methods, delivers substantially the same results but is much slower and produces greater approximation errors. Finally,

$$V_t(S_t) = \max_{c_t,n_t,e_t,x_t} Q_t(S_t, c_t, n_t, e_t, x_t)$$

so the maximizers for Q_t given S_t deliver the desired policy functions.

4. Now we have 10^5 values of the value function $V(s_{C,t}, s_{N,t}, s_{C,P}, s_{N,P}, \epsilon_P^w)$ associated to state points S_t. Name X_t the $10^5 \times K$ matrix which stores an n-th degree polynomial[27] in the values of the extracted state points at time t. To explain further, X_t includes columns of 10^5 realizations of $1, s_{C,t}, s_{C,t}^2, s_{C,P}, s_{C,P}^2, s_{C,t}s_{C,P}, ...$ and so on. A polynomial regression is performed:

$$V_t = X_t \beta_{V_t} + \epsilon_t \tag{18}$$

which coefficients are stored in the $K \times 1$ vector β_{V_t}.

5. Now go back one period and back to point 2, predicting next-period value function V_{t+1} by using $\beta_{V_{t+1}}$. The algorithm stops at the solution of period 2 because predicting V_1 is unnecessary.

After the household's problem is solved, I extract random jointly lognormal initial conditions for every household using the variance/covariance matrix XI and I solve again each household's problem, starting in period 1, to obtain policy functions. The reason is that the value function is easier to approximated by polynomial approximation than the policy functions for n_t, x_t, e_t.

The most nonstandard part of the algorithm is the polynomial approximation, which has to be reliable in order not to produce large errors in the solution. In this case the approximation is quite precise; define the relative ex-post approximation error ϵ^{approx} as

$$\epsilon^{\text{approx}} = \frac{\hat{V}_t(S_t) - V_t(S_t)}{\text{std}(V_t(S_t))}$$

that is, the difference between the ex-ante prediction of the value function $\hat{V}_t(S_t)$ and the ex-post solution to the problem $V_t(S_t)$, found with maximization, normalized by the standard deviation of $V_t(S_t)$. The average of such relative error is always under 10^{-4} for every t and the maximum relative error is under 10^{-2}.

[27] It follows that the number of terms K is given by $\binom{n+5-1}{n}$, where n is the degree of the polynomial and 5 is the number of variables.

Appendix E Correlation Matrices of Data Factors

	Cog. Skills at age 6	NonCog. Skills at age 6	Mom's Cog. Skills	Mom's NonCog. Skills
Cog. Skills at age 6	1.0000			
NonCog. Skills at age 6	**0.1566*** 0.0007	1.0000		
Mom's Cog. Skills	**0.2836*** 0.0000	0.2097* 0.0000	1.0000	
Mom's NonCog. Skills	0.1528* 0.0000	0.2021* 0.0000	0.4167* 0.0000	1.0000

Table XIV: Pairwise Correlation Matrix of Skills Factors, with significance level (* = 0.001 significance); all families (minimum N = 460). Calibration targets are displayed in bold. Source: data extract from NLSY79/CNSLY79 by Cunha, Heckman and Schennach (2010).

	Cog. Skills at age 14	NonCog. Skills at age 14	Mom's Cog. Skills	Mom's NonCog. Skills
Cog. Skills at age 14	1.0000			
NonCog. Skills at age 14	0.2494* 0.0000	1.0000		
Mom's Cog. Skills	0.4303* 0.0000	0.1336* 0.0000	1.0000	
Mom's NonCog. Skills	0.1844* 0.0000	0.1444* 0.0000	0.4167* 0.0000	1.0000

Table XV: Pairwise Correlation Matrix of Skills Factors, with significance level (* = 0.001 significance); all families (minimum N = 1009). Source: data extract from NLSY79/CNSLY79 by Cunha, Heckman and Schennach (2010).

Appendix F Data on Child Care Time

The data of Ramey and Ramey (2010) consist in 13 time diary surveys for US, for years 1965, 1975, 1985, 1992-1994, 1995, 1998, 2000, and all years from 2003 to 2008. I use the time diaries from 2003 to 2008 to compute the averages with which I compare the model. The online Appendix of Ramey and Ramey provides details on how child care time is computed, i.e. which activity codes have been included in total child care time.

I do not use time diaries data prior to 2003 for two reasons. First, as Ramey and Ramey notice, several researchers have doubted the comparability between the 1993 survey with other surveys (see for instance Allard et al. (2007), Robinson and Godbey (1999), Bianchi et al. (2004)). Second, as Ramey and Ramey (2010) and Aguiar and Hurst (2007) show, child care time from 1995 to 2003 increased substantially for all groups.

I include into the sample only married parents between the ages of 25 and 54, who had a child after age 20 and prior to age 44; hence I use a total of 20592 observations. Of these, 9594 (46.5%) are males. College-educated individuals are 8582 (41.7%). Of all parents, 52.6% have a child under the age of 5. I use the ATUS recommended weights for computing averages. Results for averages and ratios are robust to the inclusion of older individuals.

I follow the definition of child care time in Ramey and Ramey (2010), considering it as the sum of primary, educational and recreational child care time. My results are comparable to those of Hurst's comment to Ramey and Ramey; more details on how the averages have been calculated can be found in the Appendix.

Appendix G The Cognitive-Skills-Only Model

Technology of **Cognitive** Skills Formation				
		1st Stage		2nd Stage
Cog Skills	$\gamma_{1,C,1}$	0.303	$\gamma_{2,C,1}$	0.448
NonCog Skills	$\gamma_{1,C,2}$	0	$\gamma_{2,C,2}$	0
Investment	$\gamma_{1,C,3}$	0.319	$\gamma_{2,C,3}$	0.098
Parental Cog	$\gamma_{1,C,4}$	0.378	$\gamma_{2,C,4}$	0.454
Parental NonCog	$\gamma_{1,C,5}$	0	$\gamma_{2,C,5}$	0
Complementarity	$\phi_{1,C}$	-0.180	$\phi_{2,C}$	-0.781
Elasticity of Substitution $1/(1-\phi)$		0.847		0.561
Variance of Shocks	$\eta_{1,C}$	0.193	$\eta_{2,C}$	0.050

Table XVI: The Technology for Only Cognitive Skill Formation (used for counterfactual experiment); estimated by Cunha, Heckman, Schennach (2010) [Online Appendix] taking into account Investment endogeneity; skills linearly anchored to educational attainment; factors normally distributed, standard errors in parentheses.

Parameter	Value	Target	Data	Model
Preferences				
ζ	0.31	Hours worked	0.330[a]	0.363
χ	0.82	Avg. Time in Child Care when child < 6	0.134[b]	0.109
ψ	1.00	Ratio Early/Late Time	1.767[c]	1.407
ξ	1.57	Correlation($s_{C,4}, s_{C,P}$)	0.284[c]	0.611
Income Equation				
$\beta_{C,\text{model}}$	0.42	Mincer returns to $s_{C,P}$	0.288	0.277
$\beta_{N,\text{model}}$	0.51	Mincer returns to $s_{N,P}$	0.074	0.072
Variance(ϵ_P^w)	0.48	Var. of Mincer Residuals	0.218	0.223
Investment				
A	9.021	Mean($s_{C,3}$) $= 0$		
α_1	0.229	Share of income spent in child, ages 0-2	0.046[c]	0.041
ω	-8.642	Indirect Inference, Time on Wage	0.135	0.103

[a] See McGrattan and Rogerson (2004).
[b] Source: author's calculations on 2003-2008 ATUS.
[c] Source: author's calculations on CNLSY/79.

Table XVII: Calibration of parameters endogenously determined; targets, data moments and simulated moments.

	Model	Data
Early (0-6) Time in Child Care		
College	10.6	14.9
Noncollege	11.2	13.0
Δ, College/NonCollege	-5.9%	14.1%
Late (7-14) Time in Child Care		
College	8.6	7.5
Noncollege	7.8	6.8
Δ, College/NonCollege	11.3%	10.6%
Hours of Work		
Early (0-6)	35.8	30.8
Final (14)	35.6	33.0
% Change during Early C.	0.6%	-6.8%
Intergenerational Correlations		
$\rho(\theta_{C,4}, \theta_{C,P})$	0.61 (calibr.)	0.28
$\rho(\theta_{C,T+1}, \theta_{C,P})$	0.92	0.43
$\rho(\theta_{C,T+1}$,HH Income)	0.50	0.29

Table XVIII: External Validation, Data vs **Cognitive-Skills-Only** Model. Summary statistics for Time invested in children and Intergenerational Correlations. Data on time use are the author's calculations on the 2003-2008 American Time Use Survey, on married parents aged 25-44 of children aged 0-6 for early time and 7-14 for late time. Numbers are obtained by summing average primary child care time of mothers and average time of fathers, and dividing by the assumed time endowment of 200 hours. The degree of assortative mating is as in Eika, Mogstad and Zafar (2017) for the US. All observations are weighted as recommended by the ATUS. Work hours data are taken from McGrattan and Rogerson (2004) and are calculated as the sum of average working hours for married males plus average working hours for married females. Intergenerational Correlations are the author's calculations on CNLSY/79 data.