Policies for Early Childhood Skills Formation: Accounting for Parental Choices and Non-Cognitive Skills

Iacopo Morchio

Discussion Paper 22/755

18 January 2022

School of Economics

University of Bristol
Priory Road Complex
Bristol
BS8 1TU
United Kingdom

University of Bristol
School of Economics

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

Iacopo Morchio[*][†]

January 2022

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 *et al.* (2010), featuring risky investment in children, time use trade-offs, idiosyncratic income risk and borrowing constraints. Accounting for noncognitive skills implies higher effectiveness of parental investments, and therefore 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 Bristol. E-mail: iacopo.morchio@bristol.ac.uk

[†]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 Árpád Ábrahám, Pedro Gomes, Nezih Guner, Gregory Jolivet, Matthias Kredler, Claudio Michelacci, Christopher Rauh, José Victor Rios-Rull, Loris Rubini, Manuel Santos, Kjetil Storesletten, Ludo Visschers, Yikai Wang, Hakki Yazici 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 would be the impact of introducing a universal child allowance or child tax credit in the US.

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 literature on the technology of skills formation emphasizes the importance of accounting for multiple skills in order to correctly estimate the returns to parental investments.[3] In particular, accounting for noncognitive skills and their feedback to cognitive skills have been shown to be key. This literature argues that policies that target early childhood exhibit large gains.[4]

A related structural literature has instead emphasized the importance of the tradeoffs faced by households who invest time and resources in their offspring.[5] This literature is often more negative on returns to policies, arguing that endogenous responses of parents can limit returns substantially (Bernal & Keane, 2011; Del Boca *et al.*, 2014). One of the most important differences between the two strands of the literature is that the latter tends to focus on unidimensional human capital, or cognitive skills only. 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 noncogni-

[1]See Huggett, Ventura and Yaron (2011); Lee and Seshadri (2019); 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 & Heckman (2007); Cunha *et al.* (2010); Helmers & Patnam (2011). See also Agostinelli & Wiswall (2020) for recent developments.

[4]See Morris *et al.* (2005); Nores *et al.* (2005); Cunha *et al.* (2006); Heckman *et al.* (2013) among others.

[5]See Bernal & Keane (2010, 2011); Del Boca *et al.* (2014); Griffen (2019); Brilli (2012); Youderian (2016); Yum (2020); Daruich (2018); Lee & Seshadri (2019)

tive skills formation proposed by Cunha *et al.* (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. 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 face idiosyncratic wage shocks, and have to choose how much to consume and save, how many hours to 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. Parental investment is endogenously determined in the model by a production function that combines the time and goods inputs of households. 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.

The skills formation technology specifies a relationship such that the future skills of a child are a function of parental skills and investment, and of the child's current skill endowments. 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. Thus, this paper bridges the literature that estimates the technology of skills formation[6] with the literature that uses models to estimate early childhood policy returns.[7] To the best of my knowledge, this is the first paper in which cognitive and noncognitive skills of the child are endogenously determined within a quantitative model of parental investment. I show that accounting for noncognitive skills is crucial, as it implies roughly two times higher policy returns than if they are neglected.

I estimate the model using data from the American Time Use Survey (ATUS), the Children of the National Longitudinal Survey of the Youth (CNLSY/79) and the Inte-

[6]See Cunha & Heckman (2007); Cunha *et al.* (2010); Helmers & Patnam (2011); Todd & Wolpin (2007a); Hanushek & Woessmann (2008).

[7]See Bernal & Keane (2010, 2011); Del Boca *et al.* (2014); Griffen (2019); Brilli (2012); Youderian (2016); Yum (2020); Daruich (2018); Lee & Seshadri (2019); Caucutt & Lochner (2020).

grated Public Use Microdata Series (IPUMS). My estimation results suggest that time is the most important input in parental investment, although goods become relatively more important in late childhood; that time and goods are strong complements; and that households do not dislike time invested in their children as much as working.

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. Another way to understand my contribution is that I provide a theory of endogenous investment that is consistent with the evidence on skills formation, and the abovementioned technology of child development, and that allows the use of the model as a laboratory for policy analysis. Using this technology also allows me to account for both cognitive and noncognitive skills, of both households and offspring. Allowing for endogenous savings and borrowing constraints is important to correctly estimate the returns to child allowance policies: I find that, compared to the average return, policy returns are up to four times higher among parents who were constrained during the investment phase.

Finally, I use the model to simulate the impact of a number of policies. First, I find that the introduction of a universal child allowance policy inspired by the German Kindergeld scheme, worth approximately 5 percent of the average household income in all periods, increases cognitive skills at age 14 by 1.7 percent of a standard deviation and noncognitive skills at age 14 by 2 percent. This increase is higher for low-income, low-skilled and constrained households. The increase in skills is driven by an increase in time invested by households in child care. Even though most 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. I also find that a universal child tax credit with the same cost has a smaller impact on skills formation, increasing cognitive and noncognitive skills by around 0.8 and 1 percent respectively, with returns that are similar across the distribution of income and skills. The reason is that, while increased net wages allows households to invest more goods in child development, they affect the time use tradeoff in the wrong direction: households work slightly more hours and child care time barely increases, resulting in a smaller increase in parental investment compared to the transfer policy. Larger transfers and child tax credits exhibit larger returns, though less than proportionally to the size of the transfer.

As mentioned above, I find that accounting for noncognitive skills increases substantially 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 smaller impact on children's skills, which increase by about one-half of what the two-skills model implies. I show that this result can be explained by the difference in the estimated elasticity of substitution when both skills are accounted for, and by the fact that the cognitive-skills-only model loads all heterogeneity in productivity across households on parental cognitive skills. Thus, parental investment is less effective in the cognitive-skills-only model, and policies have smaller impacts. This result demonstrates the importance of accounting for noncognitive skills to correctly estimate policy returns.[8]

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 literature that estimates the technology of skills formation (Todd & Wolpin, 2003, 2007b; Cunha & Heckman, 2007; Cunha *et al.*, 2010; Helmers & Patnam, 2011; Todd & Wolpin, 2007a; Hanushek & Woessmann, 2008) with the literature that uses models to estimate early childhood policy returns (see Aiyagari *et al.*, 2002; Bernal & Keane, 2010, 2011; Lochner & Monge-Naranjo, 2011; Del Boca *et al.*, 2014; Soytas *et al.*, 2014; Griffen, 2019; Brilli, 2012; Youderian, 2016; Yum, 2020; Daruich, 2018; Lee & Seshadri, 2019; Caucutt & Lochner, 2020). None of the structural papers cited features noncognitive skills; also, most focus on unidimensional unobservable human capital.[9] 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

[8]Cunha *et al.* (2010) show that accounting for two skills matters for how investment should be distributed across developmental stages and initial skills. They perform this experiment in a reduced-form fashion rather than in a model of parental investment, and they do not compare aggregate returns as I do.

[9]Abbott *et al.* (2019) include cognitive and noncognitive skills in the model as exogenous endowments that determine education choices and adult outcomes, whereas this paper focuses on the impact of policy on parental choices and endogenous skill development.

that child allowance policies are relatively ineffective at influencing skills development.

The papers by Daruich (2018), Lee & Seshadri (2019), Youderian (2016), Yum (2020), Abbott (2021) develop macroeconomic models of human capital formation to understand how policies 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 human capital at a disaggregated level (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 the results of the estimation. Section 5 outlines the policy experiments and describes their impact. Section 6 concludes.

2 The model

I build a partial equilibrium model in the spirit of Becker (1964) and Becker & Tomes (1979), featuring five key ingredients: the technology of skills formation, the investment formation technology, time-allocation choices of households, ex-ante heterogeneity in skills for parents and offspring, and borrowing constraints Aiyagari (1994); Huggett (1993), which have been shown to be important during child development (Lochner & Monge-Naranjo, 2012; Carneiro *et al.*, 2015; Caucutt *et al.*, 2020). Before describing the setup of the model, I discuss the technology of skills formation adopted, and set out the notation for the offspring's and parental skills.

2.1 The technology of skills formation

The technology of skills formation is taken from Cunha *et al.* (2010) (CHS from now on). I choose to build the model around this particular technology because of its flexibility and generality: the estimation strategy adopted in CHS allows to account for a number of the features of child development, and of the related data, that have been shown to be key in the literature. It allows to correct for the fact that the inputs of the technology, such as parental investments and skills, and the child's skills, are difficult to measure (Cunha *et al.*, 2010); and for the fact that the technology is likely to have different parametrizations in different developmental phases (see Heckman *et al.*, 2007; Cunha *et al.*, 2006).[10]

CHS estimate the technology of skills formation assuming there exist two different developmental phases, $j = \{1, 2\}$, which correspond to *early childhood* (ages 0-6) and *later childhood* (ages 7-14) respectively. The 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. In what follows, I will denote by $s_{C,t}$ the child's cognitive skills and by $s_{N,t}$ the child's noncognitive skills in period t.

Parents provide three separate 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 differs across phases j and skills $k \in \{C, N\}$, and takes the 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 the next period's skills $s_{k,t+1}$ of each type $k \in \{C, N\}$ are a function of parental investment I_t, the offspring's cognitive and noncognitive skills $\{s_{C,t}, s_{N,t}\}$ at time t and parental cognitive and noncognitive skills $\{s_{C,P}, s_{N,P}\}$. In the work of CHS, periods t are two years long and phases 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 phases $j = \{1, 2\}$ 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.

The technology exhibits four main properties that drive returns to parental investment in the model:

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

[10] In essence, by adopting the estimates of CHS, I assume that their estimation strategy correctly accounts for these difficulties, and that their estimates are correct and unbiased. Some parameters, most notably the elasticities of substitution across phases, exhibit relatively larger standard errors, and therefore I conduct robustness checks with respect to the value of this parameter in Appendix F.

skills, and viceversa.

3. **Efficiency**: in the first phase, investment is more productive than in the second phase, 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 phase of cognitive skills development, the elasticity of substitution between inputs is roughly four times larger than in the second phase; 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, however, inputs become strongly complementary, so that increasing cognitive skills in this phase becomes extremely costly. Noncognitive skills exhibit roughly the same elasticity of substitution across phases.

Appendix A gives graphical examples of how these properties affect returns to parental investment. CHS 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. The specific parametrization adopted in the model is discussed in detail in Section 3.

2.2 Model Environment

Time is discrete and lasts forever. The economy is populated by a measure 1 of households, who live in three stages: the fertile stage, when they consume, work and save but have no children yet; the parenthood stage, when they also make investments in their child; and the final stage, when they consume, work and save and cannot affect their child's skills anymore. I introduce a fertile stage in order to account for asset heterogeneity at birth, which has important consequences for investment: households who have a child earlier may be less prepared in terms of financial assets availability than those who have a child later. Each household can have at most one child.[11]

In all periods of all stages, households enter the period with assets a and with their fixed endowments, to be defined below. They choose consumption c, hours of

[11] The unitary household model is common in the literature (Restuccia & Urrutia, 2004; Daruich, 2018; Caucutt *et al.*, 2020). Extending the model to two parents, or to multiple children, is possible but further increases the computational burden of an already complex model. I leave this to future research.

work n and the next period's assets a'. In all stages, they are endowed with one unit of time so that their total time allocation cannot exceed one. In the parenthood stage, households make additional choices and are characterized by additional state variables, which are specified below.

Finally, households have CRRA preferences over consumption c, so that their utility from consumption can be written as $u(c) = \frac{c^{1-\theta}}{1-\theta}$. They dislike work and time spent with their offspring to differing degrees, so that their disutility from work n and time invested in the offspring x can be written as $g(n, x) = -\zeta(n + \delta x)^{1+\sigma}/1 + \sigma$, where ζ governs the relative disutility of work hours or time invested in children, σ is the inverse labor supply elasticity when $x = 0$ and $\delta > 0$ governs the relative disutility of x versus n.[12] When the offspring is not present, optimization implies $x = 0$.

2.3 The Fertile Stage

Households are born in the fertile stage with no assets ($a = 0$), and draw the vector of fixed endowments $\varOmega = \{s_{C,P}, s_{N,P}, \epsilon_P\}$ from the following distribution:

$$\{\log s_{C,P}, \log s_{N,P}\} \sim N(0, \varSigma_P) \tag{2}$$

$$\epsilon_P = \begin{cases} \underline{\epsilon} < 0 & \text{w.p. } 0.5 \\ 0 & \text{w.p. } 0.5 \end{cases}, \tag{3}$$

where $s_{C,P}, s_{N,P}$ are parental cognitive and noncognitive skills, respectively, and ϵ_P is a permanent wage component which divides households into high-wage and low-wage households, *ceteris paribus*, representing persistent market luck parsimoniously.

During the Fertile stage, households derive utility from consumption and leisure, and take into account the discounted utility of the future. They choose consumption

[12] The typical assumption in the literature is that $\delta = 1$, or that parental utility only depends on leisure, that is, $1-n-x$; see for instance Caucutt *et al.* (2020). However, $\delta = 1$ implies that households with children reduce their labour supply almost exactly by their increase in child care time x when children are present. This is not what I observe in the data: the drop in labour supply after a child is born is a fraction of the increase in child care time. $\delta < 1$ allows the model to be consistent with this time use pattern. More details on time use and hours of work are available in Table XVII in Appendix E.5.

c, work hours n and the next period's assets a', and face the following constraints:

$$c + a' \leq (1+r)a + w(\Omega, \epsilon_t)n\,, \tag{4}$$
$$a' \geq 0\,, \tag{5}$$
$$0 \leq n \leq 1\,. \tag{6}$$

Equation 4 is the budget constraint of the household, stating that consumption and savings cannot exceed the sum of previous assets, income derived from interest on assets ra and labor income. Equation 5 is a borrowing constraint, stating that households cannot borrow. Equation 6 is a time constraint, stating that households can work at most their full time endowment and must supply nonnegative hours of work. The wage at which they work is defined as $w(.)$.[13] In each period t of all stages, the household's wage is given by:

$$\log w(\Omega, \epsilon) = \beta^w_{CP} s_{C,P} + \beta^w_{NP} s_{N,P} + \epsilon_P + \epsilon \tag{7}$$

where β_{CP} is the cognitive skills premium, β_{NP} is the noncognitive skills premium and ϵ is an idiosyncratic wage shock, which follows a finite state Markov transition matrix M_ϵ.

During the fertile stage, the probability that a child is born is given by

$$f(\Omega) = \frac{1}{\exp(\pi_0 + \pi_{CP} s_{C,P} + \pi_{NP} s_{N,P} + \pi_w \epsilon_P)} \tag{8}$$

where the probability of an offspring being born is allowed to depend on all fixed parental characteristics $\{s_{C,P}, s_{N,P}, \epsilon_P\}$. This is to capture the fact that higher-skilled parents, who are often higher-educated, tend to have children later, and therefore have more time to accumulate assets and face childbirth with more financial resources, on average. When an offspring is born, households move to the parenthood stage.

Finally, during the fertile stage households draw an infertility shock that moves

[13] This "household-level" wage is a model simplification, and its data counterpart is a weighted average of the wage earned by each spouse, weighted by the relative labour supply. Therefore, wage shocks within this model capture fluctuations that occur to the wages of both spouses, as well as some intra-household reallocation, and household-level labour supply responses capture the solution to a complex intra-household problem that I do not model here. The question of how shocks to each member of the household affect intra-household allocations, and in turn child development, is very interesting and I leave it to further research.

them directly to the final stage. This happens with probability p^{nf}. This shock is introduced to account for the fact that not all households have children in the data, and that this tends to happen more often to higher-skilled households, who are also associated with delayed fertility. Clearly, this affects the composition of households with children, which may be important for the aggregate returns to different policies.

Denote by V^F the value function in the fertile stage, V_1^C the value function in the first period of parenthood, and V^{T+1} the value function in the final stage. The household's problem is written as follows:

$$\begin{aligned} V^F(\varOmega,\epsilon,a) = \quad & \max_{c,n,a'} \frac{c^{1-\theta}}{1-\theta} - \zeta \frac{n^{1+\sigma}}{1+\sigma} \\ & + (1-p^{nf})\Big[\beta(1-f(\varOmega))\mathbb{E}\left[V^F(\varOmega,\epsilon',a')|\epsilon\right] \\ & + \beta f(\varOmega)\mathbb{E}\left[V_1^C(s_{C,1},s_{N,1},\varOmega,\epsilon',a')|\epsilon\right]\Big] \\ & + p^{nf}\beta\mathbb{E}\left[V^{T+1}(0,0,\varOmega,\epsilon',a')|\epsilon^w\right] \end{aligned}$$

subject to

$$\begin{aligned} c + a' &\leq (1+r)a + w(\varOmega,\epsilon)n \\ a' &\geq 0 \\ 0 \leq n &\leq 1 \end{aligned}$$

2.4 The Parenthood Stage

The parenthood stage lasts $T = 7$ periods, divided in two phases $j = \{1,2\}$ that follow the estimation strategy of CHS: early childhood (phase 1, periods 1–3) and late childhood (phase 2, periods 4–7). At the beginning of the first period of the parenthood stage, a child is born with initial conditions $\{s_{C,1}, s_{N,1}\}$, cognitive and noncognitive skills respectively, which log is drawn from the joint normal distribution $N(0,\varSigma_{CN})$.

During the parenthood stage, households enter each period with their fixed endowments $\varOmega$, the assets a saved from the previous period, and their offspring's skills $\{s_{C,t}, s_{N,t}\}$. Also, households face three different shocks: at the beginning of each period, they receive an idiosyncratic wage shock ϵ. At the end of each period, their offspring's cognitive and noncognitive skills each receive a shock, denoted as $\eta_{C,t}$ and

$\eta_{N,t}$, respectively.

Households decide how much to consume c, how many assets to hold in the next period a', how many hours to work $n \geq 0$, and how much time $x \geq 0$ and goods $e \geq 0$ to invest in their offspring. Their budget constraint reads

$$c + e + a' \leq (1 + r)a + w(\Omega, \epsilon)n\,. \tag{9}$$

In line with Del Boca *et al.* (2014) and others, I assume that households derive utility $h(.)$ from the "quality" of their children in every period of the development stage. I make this assumption to allow households to care, on average, for a multiplicity of different outcomes that depend on the skills of their children (performance in school and in the labour market, behavioural problems, criminal behaviour in adulthood) in a parsimonious fashion. An alternative often used in the literature is to assume that households are altruistic and care for the future value of their children. However, it is not clear if parents only care for the future utility of the child calculated from the child's perspective. Also, in the context of a partial equilibrium single-generation model, this would require imposing an assumption on how this value is determined. Thus, I allow the data to pick appropriate parameters to describe household behaviour instead. I choose the flexible functional form $h(s_{C,t}, s_{N,t}) = \frac{(s_{C,t}^{\psi} s_{N,t}^{1-\psi})^{1-\xi}}{1-\xi}$ with the parameter χ determining the relative importance of children with respect to consumption and leisure, a share parameter ψ determining the relative weight of cognitive vs. noncognitive skills, and the parameter ξ determining parental risk-aversion with respect to investment in children.

Finally, during this stage, investment in children I_t is obtained as a composite of time and goods according to the following Constant Elasticity of Substitution functional form:

$$I_t = A\Big[\alpha_t x^{\omega} + (1 - \alpha_t)e^{\omega}\Big]^{1/\omega}, \tag{10}$$

where A is the scale of the investment function, α_t is a period-specific parameter governing the relative importance of time and goods, and ω governs the elasticity of substitution between them.

Notice that, in this stage, the value function V_t^C is time-dependent, because every period of parenthood is different due to the constraints and the technology changing over time, and to the finite horizon of the child development process. As T is the final period of child development, $V_{t+1}^{C} = V^{T+1}$, the value of the final stage. Thus, the household's problem in all periods of parenthood $\{1, .., 7\}$ can be written as follows:

$$\begin{aligned} V_t^C(s_{C,t}, s_{N,t}, \Omega, \epsilon, a_t) = \quad & \max_{c,e,n,x,a'} \frac{c^{1-\theta}}{1-\theta} - \zeta \frac{(n+\delta x)^{1+\sigma}}{1+\sigma} \\ & + \chi \frac{(s_{C,t}^{\psi} s_{N,t}^{1-\psi})^{1-\xi}}{1-\xi} \\ & + \beta \mathbb{E}\left[V_{t+1}^{C}(s_{C,t+1}, s_{N,t+1}, \Omega, \epsilon', a')\right] \end{aligned}$$

subject to

$$\begin{aligned} c + e + a' &\leq (1+r)a + w(\Omega, \epsilon)n \\ a' &\geq 0 \\ 0 &\leq n + x \leq 1, \quad n, x \geq 0 \\ I_t &= A\left[\alpha_t x^{\omega} + (1-\alpha_t)e^{\omega}\right]^{1/\omega} \\ s_{C,t+1} &= [\gamma_{j,C,1} s_{C,t}^{\phi_{j,C}} + \gamma_{j,C,2} s_{N,t}^{\phi_{j,C}} + \gamma_{j,C,3} I_t^{\phi_{j,C}} + \gamma_{j,C,4} s_{C,P}^{\phi_{j,C}} + \gamma_{j,C,5} s_{N,P}^{\phi_{j,C}}]^{1/\phi_{j,C}} \eta_{C,t+1} \\ s_{N,t+1} &= [\gamma_{j,N,1} s_{C,t}^{\phi_{j,N}} + \gamma_{j,N,2} s_{N,t}^{\phi_{j,N}} + \gamma_{j,N,3} I_t^{\phi_{j,N}} + \gamma_{j,N,4} s_{C,P}^{\phi_{j,N}} + \gamma_{j,N,5} s_{N,P}^{\phi_{j,N}}]^{1/\phi_{j,N}} \eta_{N,t+1} \end{aligned}$$

2.5 The Final Stage

After the parenthood stage, households get utility from their offspring's final skills, but cannot influence them anymore. They enter each period with their fixed endowments Ω, the assets saved from the previous period, and face an idiosyncratic wage shock ϵ. They still consume (c), work (n) and save (a').

The household's problem in the final stage can be written as follows:

$$\begin{aligned} V_{T+1}(s_{C,T+1}, s_{N,T+1}, \Omega, \epsilon, a) = \quad & \max_{c,n,a'} \frac{c^{1-\theta}}{1-\theta} - \zeta \frac{n^{1+\sigma}}{1+\sigma} \\ & + \chi^F \frac{(s_{C,T+1}^{\psi} s_{N,T+1}^{1-\psi})^{1-\xi}}{1-\xi} \\ & + \beta \mathbb{E}\left[V_{T+1}(s_{C,T+1}, s_{N,T+1}, \Omega, \epsilon', a')\right] \end{aligned}$$

subject to

$$
\begin{aligned}
c + a' &\leq (1+r)a + w(\Omega, \epsilon)n \\
a' &\geq 0 \\
0 \leq n &\leq 1
\end{aligned}
$$

Notice that households are allowed to care differently for the quality of their children in the final stage, compared to the parenthood stage, through the parameter χ^F, which may be different from χ. The idea is to allow households to care less (or more) for their children's quality after the development process has ended.

2.6 Implications for parental choices

One useful analytical result (derived in Appendix B) is that, during the child development process, optimal goods invested e_t^* in an interior solution solve the following equation:

$$
e_t^* = \left(\delta w(\Omega, \epsilon) \frac{1-\alpha_t}{\alpha_t}\right)^{\frac{1}{1-\omega}} x_t^*, \tag{11}
$$

so that they are increasing in the input share of goods $1-\alpha_t$, in time invested y_t, in how much households dislike time invested relative to work hours δ and in the wage faced by the household $w(\Omega, \epsilon)$. The first result is trivial; the second is due to input complementarity; the third is because goods are a more attractive investment option when the shadow price of time is higher; and the last is because goods are relatively cheaper, compared to time, for households who are facing a higher wage.

Appendix C details the solution algorithm used to solve the model, which presents some computational challenges due to the large state space, and to the need for the model to be estimated.

3 Data and Estimation

The model is estimated using the Simulated Method of Moments (McFadden, 1989). I use data from the American Time Use Survey 2003-2017 (ATUS from now on), the Children of the National Longitudinal Survey of the Youth 1979 (CNLSY/79 from now on) and the Integrated Public Use Microdata Series (IPUMS) 2000-2017. I use the

ATUS to measure child care time, how child care time relates to parental education, and how child care time and hours of work relate to earnings. I use the CNLSY/79 to measure cognitive and noncognitive test scores, both of mothers and offspring, household income during parenthood, and time to the birth of the first child. I use IPUMS to measure hours of work and fertility by education.

First, I set a number of parameters exogenously and, to adopt the technology of skills formation in Cunha *et al.* (2010), I make a number of choices that allow the model to be consistent with the technology. 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. As in CHS, periods 1,2,3 of parenthood correspond 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 from 8 onwards belong to the final stage. Three ingredients of the model are taken from the paper by Cunha *et al.* (2010): the parametrization of the technology of skills formation, the initial joint distribution of skills at birth Σ_C and the initial distribution of parental skills Σ_P. The parameters of the technology are reported in Table XIII in Appendix E.3, 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.

The discount factor β is set to 0.9216, which is equivalent to 0.96 at the yearly level, a standard value in the literature. Following Osuna & Rios-Rull (2003), I set the time endowment of households to be 200 hours per week, excluding sleep and personal maintenance. Finally, I set the interest rate $r = 0.08$, just shy of $\frac{1}{\beta} - 1$, the point at which there is no steady state in the final stage of the model (Huggett, 1993; Aiyagari, 1994). Before proceeding to describe the moments used, I establish the data equivalents of the theoretical concepts presented in Section 2.

Child care time x_t in the model is matched to primary child care time, measured in the ATUS microdata. I use the data from Aguiar *et al.* (2021), which is the merge of several surveys of time use from 2003 to 2017. I target averages from the ATUS, rather than from the previous survey on time use (the AHTUS), because of the larger sample size and narrower time window, which makes the data easier to compare across years.[14] Finally, I calculate average child care time by education groups and age of the offspring, after cleaning out differences attributable to survey years, age, race and number of children in the household. I discuss in detail the methodology I adopt, the choice of data and the estimation sample in Appendix D.1.

Hours of work n_t are matched to average actual hours worked by households in the IPUMS microdata. I use data from the period 2000-2017 to be consistent with the choice of period in the ATUS data, and calculate average hours worked by education groups and by age of the offspring, again after cleaning out differences attributable to survey years, age and race. Details can be found in Appendix D.2.

The data from the CNLSY/79 is the same as in Cunha *et al.* (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 and socioemotional development, such as temperamental scores. For very early ages (0-2), the best predictors of future tests are measured; for instance, when they estimate cognitive skills at birth, CHS use gestation length, birth weight and motor-social development.

I obtain part of the estimation targets from the estimation of "skill factors" from assessments 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

$$Z_{i,j,t} = \alpha_{i,t} + \beta_{i,t} log(s_{C,j,t}) + \epsilon_{i,j,t} \,, \tag{12}$$

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 $\beta_{i,t}$.[15] In this

[14]See also Ramey & Ramey (2010) for a discussion of the changes in reported child care time between the AHTUS and the ATUS.

[15]See Cunha *et al.* (2010) for a discussion of the application of this methodology in the context of

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.[16] The underlying identifying assumption is that, for two measurements i, j of the same child such that $i \neq j$, $\mathbb{COV}(\epsilon_{i,t}, \epsilon_{j,t}) = 0$. The estimation targets the correlation patterns of the offspring's skills and parental skills over childhood. For consistency in the use of the technology, I estimate the factors following closely the choice of variables described in CHS.[17]

In line with CHS, I consider parental skills to be the mother's skills. Model-simulated skills have a data counterpart in the factors calculated from tests. Therefore, the household concept adopted in the model can be defined as: a household in which the mother has cognitive skills $s_{C,P}$ and noncognitive skills $s_{N,P}$.

This gives rise to two additional challenges to estimating the model using child care time data. First, while the CNLSY/79 data do not include time use information, the ATUS does not include measures of parental skills. This is important because, in order to estimate the model's parameters, it would be useful to target different time use patterns for skilled and unskilled parents. Second, the ATUS is an individual time diary, so it is not possible to match individuals to their partners to calculate household-level child care time. This is important because, if skilled mothers marry skilled men more frequently than unskilled ones, the average time use of a household with a skilled mother may be affected. To address both challenges, I compute education-gender-specific averages of child care time and work hours, and combine these as explained below.

To solve the first challenge, and bridge the data on parental skills in the CNLSY/79 with the data on time use in the ATUS, I use education as a proxy of skills. I estimate an auxiliary Probit model in which completing college is modeled as a function of

skills formation.

[16]CHS 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.

[17]Table X in Appendix D.3 provides basic statistics for the mother's factors and for the child's factors at ages 5-6 and 13-14, showing that they match closely the results by Cunha *et al.* (2010). The conditions I state are enough to estimate skill factors with arbitrary mean and scale, but not enough to estimate the technology of skills formation. I do not have to impose further assumptions in this case because I only rely on correlations between factors as targets for the model, which are mean and scale invariant.

parental skills in the CNLSY/79 data:

$$Pr(\text{College}) = \Phi(\beta_0^{\text{col}} + \beta_C^{\text{col}} \log s_{C,P} + \beta_N^{\text{col}} \log s_{N,P})\,, \tag{13}$$

where Φ is the normal CDF. The estimates can be found in Table XV in Appendix E.4. I then use the estimates of equation (13) in model-generated data to split households into college-educated and noncollege-educated ones. In the model, this has no consequence except that I attribute a specific household's time use to the college-educated group or to the noncollege-educated group. These simulated time-use patterns are then matched to the corresponding time use targets by education groups.

To address the second challenge, I combine child care time averages by education and gender, using the measures of assortative mating in Eika *et al.* (2019), to obtain a measure of household-level time use. The details of the procedure, and a discussion of the assumptions underlying it, are provided in Appendix D.1.

Turning to the shares of time in the investment function α_t, allowing them to differ flexibly in all periods t would add seven further parameters to the estimation. To reduce the dimensionality of the estimation problem, I restrict α_t to follow a shape-preserving cubic polynomial in the time period t, and let only α_1, α_3 and α_7 vary freely.[18] I use information from the Report on Expenditures on Children by Families by the US Department of Agriculture (2012) to estimate these shares. Details are provided in Appendix D.4.

It is worthwhile to discuss the identification of ω, the parameter governing the elasticity of substitution between time and goods in the investment function. In principle, this would require access to data on time use, expenditure on children and parental investment in the same dataset. Equation (11), however, provides another avenue to identify the elasticity of substitution between inputs ω, by adopting an indirect inference approach. Taking logs of both sides, I obtain

$$\log x_t^* = \underbrace{\frac{1}{\omega-1}\log\delta + \frac{1}{\omega-1}\log\left(\frac{1-\alpha_t}{\alpha_t}\right)}_{\text{Constant term}} + \frac{1}{\omega-1}\log w(\Omega,\epsilon) + \log e_t^*\,. \tag{14}$$

Equation (14) suggests that, if it were possible to observe e_t^* and the household-level

[18]In practice, this is done by interpolating between the three points using the function *pchip* in Matlab. Alternative approaches, like estimating a simpler square polynomial, yield identical results while being harder to interpret.

wage $w(\Omega, \epsilon)$, the parameter ω could be recovered by regressing child care time on the household-level wage. However, e_t^* is unobserved and endogenous to the wage $w(\Omega, \epsilon)$ by definition, creating an omitted variable problem that makes ω impossible to identify with a simple regression. The household wage $w(\Omega, \epsilon)$ is also unobserved. However, this equation suggests that an auxiliary model in which child care time is regressed on household-level income instead can help identify ω, as income is determined also by the household-level wage. In this auxiliary model, e_t^* is treated as an unobserved component that is correlated both with the dependent variable and the regressor, and income is a proxy variable for $w(\Omega, \epsilon)$. Thus, I estimate a linear regression between child care time and log household income, both in model-generated data and in the ATUS data, and ask the estimation to bring the model and the data as close as possible in this dimension, to help identify ω. Further details are provided in Appendix D.1.

Finally, I restrict the process of idiosyncratic shocks ϵ to follow a three-states Markov chain M_ϵ, where I normalize the middle shock $\epsilon_2 = 0$ without loss of generality and estimate the values of the low shock ϵ^1 and of the high shock ϵ^3.

Summarizing, there are 29 parameters to be estimated: eight household preference parameters $(\theta, \sigma, \zeta, \delta, \chi, \psi, \xi, \chi^F)$; two parameters pertaining to the cognitive and noncognitive skill wage premium (β_{CP}, β_{NP}); one parameter governing persistent income inequality $\underline{\epsilon}$; five parameters governing fertility $(\pi_0, \pi_{CP}, \pi_{NP}, \pi_w, p^{nf})$; eight parameters pertaining to the size of idiosyncratic shocks, and governing their transition dynamics (ϵ^1, ϵ^3, 6 transition probabilities for the 3x3 Markov transition matrix); and five parameters for the investment equation (A, three coefficients for $\{\alpha_t\}_{t=1}^7$ and ω). The estimation targets 207 moments, which construction is described in detail in Appendix D.

Table I summarizes the estimation strategy: while most parameters affect all moments at the same time, some parameters are mainly identified by a subset of moments. For instance, the household's preferences are mainly identified by the time use patterns by education, including child care time and work hours. The skill premia of cognitive and noncognitive skill of the mother are identified by an auxiliary Mincer equation, described in Appendix D.3. The income process, together with the elasticity of labor supply, is mainly identified by matching a large number of income patterns found in the data. The fertility equation is identified by an equivalent auxiliary model, estimated on CNLSY/79 data, and by matching the fraction of households who ever had children before the age of 45, by education. Finally, as mentioned above, the shares

Table I. Summary of estimation strategy

Parameters	Symbols	Moment	Source	#
Preferences				
for child	$\chi, \chi^F, \psi, \xi, \delta$	Average child care time	ATUS	14
		Correlations between skill factors	CNLSY/79	76
leisure, consumption	ζ, θ, σ	Average work hours	IPUMS	14
		Relationship between n and y	ATUS	14
Investment function				
Scale	A	Normalization	-	2
Shares of time	$\alpha_1, \alpha_3, \alpha_7$	Shares of income spent on children	USDA (2015)	7
Elasticity	ω	Relationship between x and y	ATUS	14
Wage premia	γ_C^w, γ_N^w	Income premia of mother's skills	CNSLY/79	2
Permanent shock	ϵ_P	Lifetime income inequality	CNLSY/79	1
Income process	3×3 M_ϵ, ϵ^1, ϵ^3	Income persistence, variance, asymmetry, curtosis, Markov transition by quintiles	CNLSY/79	60
Fertility	$\pi_0, \pi_{CP}, \pi_{NF}, \pi_w$	Time before child is born	CNSLY/79	4
	p^{nf}	Probability of parenthood by age 45	IPUMS	2

Note: n stands for hours of work, x for child care time, y for household income. See main text for the description of abbreviations.

of time $\{\alpha_t\}_{t=1}^7$ are identified by targeting the shares of income spent on children by households in USDA data.

4 Estimation Results and Goodness of Fit

The model manages to produce moments that are relatively close to most of the overidentifying restrictions, and all parameters are estimated relatively precisely. As moments exhibit relatively low standard errors, this leads to statistical rejection of the null hypothesis of equality between the data and the model-generated statistics. However, at the local level, I cannot reject equality of the model-simulated moments and the data equivalents at the 95 percent confidence level in more than 57 percent of the cases.

Most of the parameters exhibit relatively low standard errors. However, it must be stressed that all standard errors assume that the model is correct, and that there is no estimation uncertainty regarding the technology of skills formation, which parametrization I assume as fixed. Therefore, I only infer that the moments used to estimate the model have sufficient variation to identify the model's parameters. I discuss identification and diagnose the estimation more in detail in Appendix E.1.

I now turn to the interpretation of the estimated coefficients. I start from the household's preference parameters estimates, presented in Table II. I find that households are relatively risk averse with regard to investment in children, as the concavity parameter associated to the child's skills (1.397) is higher than that of consumption (0.835). The importance of the child's skills for household utility is substantially lower when households cannot affect child development anymore (0.081 vs. 0.684). δ, the relative disutility of child care time relative to work hours, is substantially lower than 1, which is necessary to rationalize the fact that, when households spend time with their children, they do not drop labor supply by as much as their increase in child care time. I also find that the estimated labor supply elasticity at the household level is around 0.72, on the high end of the scale of empirical estimates based on individual-level data (Keane, 2011). However, this value does not seem unreasonable given that households have more margins of adjustment of their labor supply than individuals, and that women exhibit relatively higher elasticities (see Blundell & MaCurdy, 1999; Blau & Kahn, 2007).

Turning to the estimates of the parameters of the investment function (see Table III), I find that the coefficients α_t associated with parental time in the investment function are always substantially higher than 0.5, suggesting that time invested in children remains the main input throughout the child's development. However, its importance is highest at birth, which is consistent with the empirical evidence in Schoellman (2016), and goods become more important as the child ages, which is consistent with Del Boca *et al.* (2014); Brilli (2015). The estimated value of ω (-1.033) implies that time and goods are relatively strong complements in parental investment.

The estimated income process, skill premia and persistent heterogeneity level (see Table IV) fit the patterns of income dynamics, observed Mincer coefficients and lifetime income inequality quite well. The performance of the model in these dimensions is presented in Tables XXIII and XXIV in Appendix E.5.

The performance of the model against all targets is shown in detail in Tables XVII –XXIV in Appendix E.5.

Table II. Estimation results: household's preferences

Parameter	Interpretation	Value
ζ	Disutility of work	9.037
		(0.16262)
δ	Relative disutility from child care time vs. work	0.719
		(0.01390)
χ	Relative utility of children	0.684
		(0.01018)
χ^F	Final relative utility of children	0.081
		(0.00210)
ψ	Relative weight of cognitive skills	0.057
		(0.02490)
ξ	Concavity of utility from children	1.397
		(0.03273)
θ	Risk aversion in consumption	0.835
		(0.01024)
σ	Inverse labor supply elasticity	1.389
		(0.02475)

Note: Author's calculations. Asymptotic standard errors in parentheses.

Table III. Estimation results: investment function

Parameter	Interpretation	Value
A	Scale of investment	14.861
		(0.15042)
α_1	Share of goods at time 1	0.980
		(0.00503)
α_3	Share of goods at time 3	0.921
		(0.00906)
α_7	Share of goods at time 7	0.711
		(0.02995)
ω	Complementarity parameter in investment	-1.033
		(0.03042)

Note: Author's calculations. Asymptotic standard errors in parentheses.

5 Policy analysis

In this Section, I simulate a number of different policies, evaluate their performance from the point of view of increasing children's cognitive and noncognitive skills, and

Table IV. Estimation results: wage equation, heterogeneity and shocks

Parameter	Interpretation	Value
ϵ^P	Permanent income variation	-0.785
		(0.00004)
ϵ^1	Income shock 1	-1.044
		(0.01688)
ϵ^3	Income shock 3	0.368
		(0.00650)
β_{CP}	Cognitive skills wage premium	0.334
		(0.00729)
β_{NP}	Noncognitive skills wage premium	0.260
		(0.01738)
Markov transition probabilities		
p_{11}	Markov transition, shock 1 to 1	0.054
		(0.00010)
p_{12}	Markov transition, shock 1 to 2	0.181
		(0.00010)
p_{21}	Markov transition, shock 2 to 1	0.043
		(0.00007)
p_{22}	Markov transition, shock 2 to 2	0.465
		(0.00007)
p_{31}	Markov transition, shock 3 to 1	0.049
		(0.00007)
p_{32}	Markov transition, shock 3 to 2	0.446
		(0.00008)

Note: Author's calculations. Asymptotic standard errors in parentheses.

study the channels through which these policies operate.[19] To be precise, I simulate the impact of a universal childcare allowance and that of a child tax credit. The manner in which I introduce the technology of skills formation in the model somewhat limits the type of policies that can be directly evaluated using the model, in the sense

[19] I choose to focus on returns to skill accumulation rather than other outcomes that are a function of skills for three reasons. First, economic returns to skills, such as wage premia, have changed substantially in the past and are likely to change further in the future, making quantifications less reliable. Second, higher cognitive and noncognitive skills are associated to a large number of outcomes, like higher wages, higher education, better health, lower incarceration probability (Heckman *et al.*, 2006, 2013), which relative importance for welfare is difficult to pin down. Third and final, the partial equilibrium model I set up does not take into account that these returns are endogenous to aggregate skill accumulation.

Table V. Estimation results: fertility-related parameters

Parameter	Interpretation	Value
π_0	Constant of fertility equation	1.501
		(0.00005)
π_{CP}	Cognitive skills coefficient on fertility	0.466
		(0.00004)
π_{NP}	Noncognitive skills coefficient on fertility	0.111
		(0.00004)
π_w	Permanent income coefficient on fertility	0.124
		(0.00005)
p^{nf}	Probability of no more children	0.046
		(0.00005)

Note: Author's calculations. Asymptotic standard errors in parentheses.

that introducing more complex policies such as schooling changes require more serious departures from the assumptions of CHS. However, I also perform some thought experiments to evaluate the potential impact of the availability of child care and of policies aimed at delaying fertility. As these require a more substantial departure from the model's assumptions and more important changes to its structure, I dedicate separate subsections to these below.

To simulate the abovementioned policies, I manipulate the budget constraint of households during childhood: the budget constraint becomes

$$c_t + e_t + a_{t+1} \leq (1+r)a_t + (1+\tau)w(\Omega, \epsilon_t)n_t + z_t\,, \tag{15}$$

where τ is the child tax credit rate, and z_t is the cash transfer received by households with children in period t. The unconditional cash transfer I simulate is a simple flat transfer to all households, roughly equivalent to the German universal child allowance scheme, called Kindergeld.[20]

The Kindergeld transfer program started in 1936; in 2012, the Kindergeld granted

[20] In principle it would be interesting to study the impact of an in-kind transfer that manipulates e_t directly. However, Del Boca *et al.* (2014) have already shown that in-kind transfers are more effective than flat transfers in quantitative models of parental choices, and I find that this is true in this model too. Also, within this model it is difficult to account for the possibility that, if in-kind transfers are provided in the form of vouchers, households may decide to sell them rather than use them for child development. In this sense, a cash transfer can also serve as a lower bound for the impact of in-kind transfers.

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.[21]

The 2012 Kindergeld for the first and second child amounted to approximately five percent of the average household income in Germany.[22] For simulation purposes, I introduce a flat transfer of 5 percent of the model-generated average income to all households from the birth of the child to age 14, and compare the effect of such policy with respect to the baseline model. For comparison, while having the advantage of replicating a real-world policy, this transfer is much smaller than those simulated in Del Boca *et al.* (2014), equating roughly one-fourth of the 250 dollars per week used in that paper.

The child tax credit rate τ_t is chosen so that the cost of the policy is equivalent to that of the Kindergeld, totaling five percent of the average household income, to facilitate comparison across policies. Finally, to test whether larger transfers and tax credits have larger effects, I also perform the same experiments with a budget of 25 percent of average household income.

I perform all these experiments as "pure" counterfactuals: that is, I apply the policies while keeping constant the initial endowments, skills and histories of shocks of the simulated population. In this way, I can compute the returns of the policies on different categories of households, focusing on the heterogeneous impacts of the policies across subgroups. Thus, I also condition policy outcomes on whether the borrowing constraint of households was binding in at least one period of the baseline simulation. I refer to these households as "constrained".

The results of my main experiments are summarized in Table VI. All policies are quite effective at increasing average skills: in particular, the unconditional cash transfer increases cognitive skills by 1.7 percent of a standard deviation, and noncognitive skills by 1.9 percent.

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

[22]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.

Intuitively, policy returns in terms of increases in cognitive skills are particularly large among households who were constrained in their choices during early (+1.9 percent cognitive, +2.3 percent noncognitive) or late childhood (+2.2 and +2.6 percent respectively), and even larger if households were constrained in both (+2.8 and +3.2 percent), similarly to Caucutt & Lochner (2020).[23] Further, policy returns are particularly large among low-permanent income, low-skilled households, who were constrained in both the early and late phases of development: for these households, the offspring's cognitive (noncognitive) skills increase by almost 4.2 (4.9) percent of a standard deviation, which represents a very large percentage increase within the group, as their offspring exhibit skills that are around 29 percent lower than the average at age 14. Finally, returns to cash transfers are substantially smaller (but not negligible) among unconstrained high-income, high-skilled households (+0.66 percent for cognitive skills), consistently with evidence in Heckman & Mosso (2014). It is worthwhile to stress that, for the sake of comparison, Del Boca *et al.* (2014) find similar aggregate returns to a transfer that is approximately four times as large (more than 1000 dollars a month).

What accounts for these results? The unconditional cash transfer operates through both an increased availability of financial resources, especially for poorer households, and a change in the opportunity cost of time. When households receive the transfer, they decrease their work hours by 1.34 hours per week in early childhood and 0.87 hours in late childhood. This frees up time that can be invested in the child: while overall leisure increases, time invested in the child increases by 0.64 hours per week in early childhood and 0.27 hours per week in late childhood, on average. Complementarity of inputs implies that goods invested increase as well. The result is increased investment in children, which translates in higher skills at the end of the child's development.[24]

[23]Being constrained in early childhood is defined as choosing $a_{t+1} = 0$ in any $t \in \{1, .., 3\}$. Being constrained in late childhood is defined as choosing $a_{t+1} = 0$ is any $t \in \{4, .., 7\}$. I find that around one-quarter of households are constrained for at least one period of child development; this is approximately twice as many households than in Caucutt & Lochner (2020), but it is also a consequence of the shorter time period: in this paper there are seven periods of childhood, whereas there are only two in theirs. Thus, being constrained for one period in this model is not the same as being constrained for one period in their paper.

[24]I simulate all policies as "money from the sky" to isolate the effect of a single change on the economy, but it is interesting to understand how results would change if one were to simulate a balanced-budget version of the same policy. I also simulate a version of the Kindergeld policy where a proportional tax on labor finances the transfer. I find that even the balanced-budget version of the flat transfer policy increases both cognitive and noncognitive skills by around 0.7 percent of a standard deviation.

Intuitively, flat transfers somewhat depress savings: when they anticipate that they will receive a transfer during the child's development period, households reduce their savings by about 2.8 percent of their permanent income, and they are also more likely to be constrained during parenthood. However, this small decrease in savings is not enough to offset the positive impacts of the policy on final skills.

Child tax credits operate through a different channel: by increasing the net wage that households receive for every hour worked during parenthood, they change the trade-off of households in favor of increasing work hours, and reducing child care time. However, the increased net wage yields an increase in goods invested in the offspring for fixed time investments (recall equation (11)), and due to input complementarity this reverses the negative impact of the child tax credit on child care time. As a result, time investments increase slightly, goods investments increase and overall investment in children increases. Thus, both final cognitive and noncognitive skills increase in response to introducing the child tax credit, by 0.83 and 1 percent respectively.

Larger transfers and tax credits have larger effects on child development: a 25 percent tax credit raises cognitive (noncognitive) skills by 4.17 (5.04) percent; a transfer worth 25 percent of average household income raises them by 7.46 (9.35) percent. Therefore, I calculate that policy returns are approximately three times as large as previously estimated in Del Boca *et al.* (2014) when noncognitive skills are accounted for.

My results on the impact of child tax credits are also compatible with evidence from Dahl & Lochner (2012), who find that increases in household income caused by the Earned Income Tax Credit (EITC) are associated with large short-term changes in skills, of around 6 percent of a standard deviation. Consistently with their findings, when I focus on low-income households, who are the main beneficiaries of the EITC in the United States, I find that large tax credits, similar to the size of the EITC, increase cognitive skills by approximately 5 percent.

Interestingly I find that, while being effective at increasing average skills, transfers are relatively ineffective at reducing inequality and intergenerational persistence. The standard deviation of skills slightly increases under all policies considered, and the intergenerational correlation of skills drops by less than 0.5 percent. In short, the policy tends to have little effect on the fact that, by the properties of the technology of skills formation, higher-skilled parents are more productive at raising skilled children. This result can help explain why countries like Germany and the US, while exhibiting very different levels of child support from the government, exhibit very similar levels of intergenerational persistence of skills (Anger & Heineck, 2009).

5.1 What happens if noncognitive skills are neglected?

In this subsection, I show that accounting for noncognitive skills makes a substantial difference when simulating the returns to different policies. To this end, I first substitute the technology of skills formation in the model with the technology of *cognitive-only* skills formation estimated by Cunha *et al.* (2010). The associated parameters are shown in Table XIV in Appendix E.3. I then perform the same policy experiments as in the previous subsection.

Results are summarized in Table VII. I find that, when noncognitive skills are neglected, all policies have substantially smaller impacts on skills development, by at least 35 percent. Policies have a smaller impact on households' choices, which in turn translate to smaller increases in aggregate skills. Returns are particularly underestimated among constrained and low-income, low-skilled households.

5.1.1 The crucial mechanisms behind policy returns

What accounts for the large difference in policy returns when noncognitive skills are not accounted for? I run a number of quantitative experiments in which I change the features of the technology one by one, in order to investigate the reasons behind this result. In the first, I shut down cross-productivity, setting $\gamma_{j,C,2} = 0$, and redistribute this coefficient proportionally to all others.[25] In the second, I set the parameters governing the elasticity of substitution between inputs of the technology of cognitive skills formation equal to those of the cognitive-skills-only technology. That is, I set $\phi_{j,C} = \phi_{j,CO}$. In the third and last, I keep the elasticities constant but set all the coefficients of the inputs equal to those of the cognitive-skills-only technology: that is, I set the coefficients multiplying noncognitive skills $\gamma_{j,C,2} = \gamma_{j,C,5} = 0$, the coefficient multiplying cognitive skills $\gamma_{j,C,1} = \gamma_{j,CO,1}$, etcetera, while keeping the elasticities of substitution unchanged. I term this the "cognitive-only loadings" experiment.

Results are summarized in Table VIII. I find that most of the difference is accounted

[25]I do this to maintain constant returns to scale and to keep constant the relative contributions of all other coefficients. Formally, I divide each coefficient within the technology by $(1-\gamma_{j,C,2})$ and then set $\gamma_{j,C,2} = 0$.

for by the difference in the coefficients governing the relative importance of inputs within the technology, $\gamma_{j,C,i}$. The higher elasticity of substitution between inputs of the technology that accounts for noncognitive skills can explain around 20 percent of the difference in the returns to flat transfers, and around 50 percent of the difference in the returns of tax credits. The difference in the importance of inputs accounts for 100 percent or more of the difference for both policies. Cross-productivity has a negligible impact on policy returns.

The reason is that the effectiveness of parental investment is determined by two factors: the first is the productivity of investment in each period (the coefficient $\gamma_{j,k,3}$, for all skills k and phases j), and the second is the self-productivity of skills (the coefficient $\gamma_{j,k,1}$), which determines how persistent the effect of current investments is through dynamic complementarity. Overall, the cognitive-skills-only technology assigns relatively more weight to parental heterogeneity in cognitive skills ($\gamma_{j,k,4}$) than the two-skills technology, particularly during the second phase. This limits the effectiveness of investment, particularly among low-skilled households, who account for most of the returns of policies when noncognitive skills are accounted for, due to their relatively lower opportunity cost of time.

5.2 Policies that delay fertility

The model also provides a laboratory to understand the impact of one of the channels through which delaying fertility can improve children's outcomes: parents who have children later have more time to accumulate assets, and therefore be less frequently constrained in their choices. I simulate the impact of policies that delay fertility by modifying the constant π_0 so that households have children two years later, on average.

I find that the asset accumulation channel has a non-negligible impact on children's final skills: household assets at the birth of the child increase by 3.1 percent, and final cognitive (noncognitive) skills increase by 0.4 (0.2) percent respectively. The result is driven by an increase in time spent with children in the baseline: households that are more asset-rich spend less time working and more time with their children, particularly during early childhood.

The results from this policy experiment should be taken as a lower bound of the true potential returns of policies that delay fertility: the reason is that this model does not feature human capital accumulation, which may be a crucial driver of the returns to this type of policies. If households accumulate human capital on-the-job, delaying children implies higher average wages at birth, therefore larger investment in children-related goods through Equation (11) and, because of input complementarity, potentially larger time investments. In addition, households who anticipate that their income is likely to grow in the future may save less when young find themselves more often constrained. I leave the question of how important this channel is to future research.[26]

5.3 The impact of child care

Another interesting thought experiment that the model can be used for is: what would be the impact of changing the investment provider of a child for one year? This experiment can be viewed as a particular form of child care provision, in which the child is cared for by an investment provider with specific skills. This experiment is rather outside the model's scope, as the model and the technology of skills formation have been estimated using households as reference investment providers, and it is not clear whether we can think of external child care providers as perfect substitutes for parents. In addition, professional childcare providers can share their time between multiple children, reducing their time cost per child, whereas this experiment assumes that time is still provided by a household. Nevertheless, it is still interesting to understand the role that alternative child care providers may play.

To answer this question, I simulate the effect of substituting a year of parental investment in early childhood (from age 3-4 to 5-6, so during period 3), with an amount of investment equivalent to the average parental investment in those years (both in terms of time and goods), while the mother's skills that enter the technology are substituted with average skills. In a sense, another way to think about this policy is that the child is moved for one period to a different household, which is going to perform investment. I apply this policy to all households and study the change in skills throughout the distribution. I also perform this policy in two ways: in the *full response* simulation, I make parents aware that they are not going to invest in their child in that period, and that the child is going to be cared for by someone else. In the *partial response* simulation, I maintain the same policy functions as in the baseline solution,

[26]Extending the model to include stochastic human capital accumulation on-the-job is theoretically easy, but increases the computational complexity further by adding another dynamic state variable.

but apply the different investment in that period only in the simulation stage. The idea is to understand how much of the return to this policy is due to the response of households, who take into account that their offspring will receive external investment in period 3, and how much is accounted for by the mechanical difference in skills and investment.

In the full response simulation, aggregate cognitive (noncognitive) skills increase by about 0.5 (1.8) percent. While the average return is only slightly positive, returns are very heterogeneous across households. The increase is larger than 6 percent, for both skills, for households who were constrained and have low permanent income and low skills. Perhaps unsurprisingly, the children's skills decrease for unconstrained households with high income and high skills, who deliver higher investment than the average. More interestingly, households respond to the childcare policies by increasing hours of work substantially during early childhood, but also by increasing their time and goods investments during late childhood, sustaining the investments their children obtained in the childcare period.

The households' endogenous response turns out to be crucial for the aggregate policy returns: when I shut down this response, the "childcare" policy still redistributes from richer and high-skilled to poorer and low-skilled households, but has small negative effects on aggregate skills. The crucial response that is missing in this simulation is the increase in work hours in early childhood induced by the childcare policy, which allows households to accumulate assets and spend more time with their children in late childhood.

To get a rough estimate of the cost of providing this policy, I multiply the amount of child care time needed by the average wage earned by households in the economy with the same level of skills. I obtain that the "childcare" policy costs roughly one-third of the yearly household income of parents with children, making it the most expensive and least cost-effective policy simulated in this environment. However, such a policy could be made substantially cheaper (by about 63 percent) by setting wages such that only providers with low permanent income ($\epsilon^P < 0$) would find it convenient to provide childcare services. It could also be made more cost-effective by focusing the childcare efforts on low-income, low-skilled households, which exhibit large positive returns to delegating childcare.

Once again, I find that neglecting noncognitive skills implies substantially smaller policy returns even for these thought experiments (see Table XXV in Appendix F).

In particular, the childcare experiment yields negative returns, underestimating the impact of these policies among low-income, low-skilled households by a factor of three.

5.4 Robustness

All results are clearly conditional on the specific parametrization of the model, and on the specific parametrization of the technology of skills formation. While I do everything I can to obtain precise estimates, and use a flexible technology of skills formation, there remain at least two important potential sources of uncertainty surrounding the parametrization of the model. The first is that, while most of the parameters of the skills formation technology estimated by Cunha *et al.* (2010) are relatively precisely estimated, the elasticities of substitution exhibit relatively larger standard errors. The second is that the parameters ψ, governing the relative importance of cognitive skills for parental utility, is hard to pin down in a model where investment is a single variable that is shared in the production of cognitive and noncognitive skills: see also the relatively larger standard error associated to this parameter in Table II.

To address the first concern, I perform a robustness check of my main results in which I compute again the returns to policies, after changing all the technologies of skills formation to Cobb-Douglas, therefore adopting a substitution elasticity equal to one for all phases and skills. In this way, I check that my results are not overly dependent on the relatively less precise estimates of the elasticities. Results are presented in Table XXVI in Appendix F. I find that policy returns are substantially unchanged when the elasticity of substitution is set to one, suggesting that these parameters are not crucial for my results.

To address the second concern, I perform three robustness checks in which I repeat my main simulations giving different values to the parameter ψ, up to four standard deviations away from its baseline estimate (see Table XXVII in Appendix F. Again, I find that my results are substantially unchanged for different values of ψ: while returns to policies are slightly higher when ψ has a higher value, returns to flat transfers and tax credits are almost identical when ψ ranges from 0.01 to 0.2.

6 Conclusions

I have estimated policy returns using a model of household choices and child development, which embeds the technology of skills formation of Cunha *et al.* (2010) and allows for both cognitive and noncognitive skills, endogenous time allocation choices, savings and borrowing constraints. I show that even policies that can be thought to be relatively ineffective *a priori*, such as flat unconditional cash transfers, affect skills accumulation substantially by freeing time that households can invest in their children. Further, I show that accounting for noncognitive skills is key, as policy returns are approximately halved when the technology of skills formation neglects them.

The model I set up in this paper is a single-generation partial equilibrium model; as such, there are at least two interesting questions for future research. The first is how much the gains in cognitive and noncognitive skills of the offspring translate in further increases in skill accumulation in subsequent generations: as Daruich (2018) argues, increased human capital today has an impact on the human capital accumulation of future generations, which is a further return from these policies. In this context, accounting for both skills is likely to prove important as I show that policy returns in terms of noncognitive skills are even higher than for cognitive skills. The second is how policies that influence child development affect the general equilibrium of the economy: by affecting the supply of skills, these policies affect wages and skill premia in the long run. In turn, these feedback effects are likely to affect the incentives of households to work and invest in their children. While both questions were impossible to answer with the model developed in this paper, relatively straightforward model extensions, accompanied by increased available computational resources, may allow answers that will help increase our understanding of the interactions between child development and the macroeconomy.

Table VI. Returns to child tax credit and universal child allowance.

Variable	Baseline	Tax Credit		Kindergeld	
		5%	25%	5%	25%
Cognitive skills	-	+0.83	+4.17	+1.68	+7.46
Never constrained	+2.11	+0.81	+4.08	+1.50	+6.77
& high income, high skills	+57.25	+0.67	+3.35	+0.66	+3.22
Constrained late	-6.04	+0.88	+4.44	+2.19	+9.40
Constrained early	-4.95	+0.82	+4.11	+2.03	+8.57
Constrained early & late	-8.72	+0.89	+4.49	+2.83	+11.14
& low income, low skills	-29.10	+1.04	+5.27	+4.17	+15.89
Standard deviation	100.00	+0.21	+1.06	+1.08	+4.06
Correlation with $s_{C,P}$	42.79	42.81	42.92	42.61	42.26
Correlation with $s_{N,P}$	24.43	24.42	24.36	24.24	23.74
Correlation with income	25.61	25.52	25.17	25.02	23.09
Noncognitive skills	-	+0.99	+5.04	+1.92	+8.35
Never constrained	+1.26	+0.97	+4.94	+1.68	+7.46
& high income, high skills	+27.09	+0.68	+3.44	+0.61	+2.93
Constrained late	-3.58	+1.05	+5.36	+2.63	+11.04
Constrained early	-2.68	+0.99	+5.03	+2.27	+9.38
Constrained early & late	-5.93	+1.07	+5.43	+3.21	+12.42
& low income, low skills	-19.48	+1.29	+6.57	+4.88	+18.15
Standard deviation	100.00	+0.23	+1.20	+1.37	+4.99
Correlation with $s_{C,P}$	15.43	15.32	14.86	14.80	13.09
Correlation with $s_{N,P}$	20.32	20.28	20.12	20.03	19.27
Correlation with income	14.03	13.87	13.21	13.23	10.79
Work hours, early	56.43	+0.16	+0.89	-1.34	-7.56
Work hours, late	64.31	+0.18	+1.00	-0.87	-4.30
Child care time, early	28.55	+0.03	+0.12	+0.64	+3.16
Child care time, late	12.87	+0.02	+0.12	+0.27	+1.24
Goods invested, early	-	+2.42	+12.78	+2.21	+10.15
Goods invested, late	-	+2.51	+13.28	+1.94	+8.48
Assets at birth	26.92	+0.20	+1.09	-2.77	-10.49
Ever constrained	25.63	+0.13	+0.78	+2.62	+16.14
early	6.26	-0.00	-0.02	+0.43	+16.66
late	21.81	+0.14	+0.86	+2.32	+5.02

Note: Skills and goods are in log units $\times$ 100. Changes in skills are expressed in log points deviations, normalized by the standard deviation of skills, $\times$ 100. Other changes in variables are expressed in the unit of measure of the variable. Time is in hours per week. Assets are in permanent income units, that is, $a/w(\Omega, 0)$. Low- (high-) permanent income households are defined as those with $\epsilon_P < 0$ $(= 0)$. Low- (high-) skilled households are defined as those with cognitive skills below (equal to or above) the median.

Table VII. Policy returns when noncognitive skills are neglected.

Variable	Baseline	Tax Credit		Kindergeld	
		5%	25%	5%	25%
Cognitive skills	-	+0.54	+2.57	+0.79	+3.37
Never constrained	+1.34	+0.54	+2.56	+0.73	+3.19
& high income, high skills	+126.62	+0.57	+2.69	+0.41	+1.94
Constrained late	-6.03	+0.55	+2.63	+1.17	+4.83
Constrained early	-8.99	+0.53	+2.52	+0.90	+3.65
Constrained early & late	-10.13	+0.56	+2.65	+1.35	+5.24
& low income, low skills	-48.05	+0.60	+2.87	+1.91	+7.01
Standard deviation	100.00	+0.16	+0.75	+0.51	+1.85
Correlation with $s_{C,P}$	91.05	91.08	91.17	91.07	91.14
Correlation with $s_{N,P}$	35.28	35.27	35.25	35.23	35.09
Correlation with income	47.58	47.56	47.50	47.24	45.91
Work hours, early	68.55	+0.22	+1.10	-1.07	-5.21
Work hours, late	66.47	+0.22	+1.12	-0.79	-3.54
Child care time, early	2.88	+0.00	+0.02	+0.08	+0.38
Child care time, late	5.76	+0.00	+0.02	+0.10	+0.43
Goods invested, early	-	+2.61	+12.95	+2.60	+11.56
Goods invested, late	-	+2.53	+12.61	+1.55	+6.77
Assets at birth	18.93	+0.22	+1.14	-2.02	-6.73
Ever constrained	15.16	+0.21	+0.73	+8.76	+17.94
early	7.01	-0.02	-0.11	+10.67	+23.18
late	9.15	+0.23	+0.88	-0.73	-4.07

Note: Skills and goods are in log units $\times$ 100. Changes in skills are expressed in log points deviations, normalized by the standard deviation of skills, $\times$ 100. Other changes in variables are expressed in the unit of measure of the variable. Time is in hours per week. Assets are in permanent income units, that is, $a/w(\Omega, 0)$. Low- (high-) permanent income households are defined as those with $\epsilon_P < 0$ $(= 0)$. Low- (high-) skilled households are defined as those with cognitive skills below (equal to or above) the median.

Table VIII. Accounting for lower policy impacts in the cognitive-only model

Policy	Baseline	Cognitive-only	No cross-productivity	Lower elasticity	Cognitive-only loadings
Kindergeld	+1.68	+0.79 (-53%)	+1.71	+1.48	+0.70
Tax credit	+0.83	+0.54 (-35%)	+0.84	+0.68	+0.55

Note: Changes in skills are expressed in log points deviations, normalized by the standard deviation of skills, $\times$ 100.

Table IX. Returns to delaying fertility and childcare policies.

Variable	Baseline	Fertility Delay	Childcare Full	Childcare Partial
Cognitive skills	-	+0.37	+0.48	-0.31
Never constrained	+2.11	-0.12	+0.28	-0.51
& high income, high skills	+57.25	+0.38	-7.65	-8.45
Constrained late	-6.04	+1.39	+0.63	-0.27
Constrained early	-4.95	+1.97	+2.22	+1.78
Constrained early & late	-8.72	+4.29	+2.86	+2.44
& low income, low skills	-29.10	+5.15	+6.29	+5.74
Correlation with $s_{C,P}$	42.79	42.78	39.46	39.39
Correlation with $s_{N,P}$	24.43	24.32	20.44	20.41
Correlation with income	25.61	25.64	22.42	22.48
Noncognitive skills	-	+0.24	+1.80	+0.78
Never constrained	+1.26	-0.18	+1.57	+0.60
& high income, high skills	+27.09	+0.27	-3.94	-4.74
Constrained late	-3.58	+1.14	+2.20	+0.97
Constrained early	-2.68	+2.30	+3.14	+2.15
Constrained early & late	-5.93	+2.39	+3.70	+2.70
& low income, low skills	-19.48	+1.87	+6.22	+5.01
Correlation with $s_{C,P}$	15.43	15.25	11.98	12.07
Correlation with $s_{N,P}$	20.32	20.41	15.37	15.39
Correlation with income	14.03	14.12	11.24	11.37
Work hours, early	56.43	-0.55	+4.42	0.00
Work hours, late	64.31	-0.16	-0.64	-0.00
Child care time, early	28.55	+0.23	+0.30	0.00
Child care time, late	12.87	+0.05	+0.20	+0.00
Goods invested, early	-	+0.70	+3.31	+0.00
Goods invested, late	-	+0.32	+1.29	+0.03
Assets at birth	26.92	+3.14	-2.24	0.00
Ever constrained	25.63	-1.73	+2.14	+0.01
early	6.26	-0.54	+10.74	0.00
late	21.81	-1.39	-7.45	+0.01

Note: Skills and goods are in log units $\times$ 100. Changes are expressed in log points deviation normalized by the standard deviation of skills. Time is in hours per week. Assets are in permanent income units, that is, $a/w(\Omega, 0)$. Low- (High-) permanent income households are defined as those with $\epsilon_P < 0$ $(= 0)$. Low- (High-) skilled households are defined as those with cognitive skills below (equal to or above) the median.

Appendix A The Technology of Skills Formation

To shed some light on the inner workings of the technology of skills formation central to the model, this Appendix gives examples of the main implications of the technology of skills formation estimated by Cunha *et al.* (2010). The features of the technology, along with the specific parametrization used in the model (available in Table XIII in Appendix E.3), produce a number of derived results that give insights into how parental investment should behave if households knew the technology of skills formation.

It is useful to summarize again the main properties of the technology here:

1. **Self-Productivity**: skills exhibit self-productivity in the sense that $\gamma_{j,C,1} > 0, \gamma_{j,N,2} > 0$ for $j = \{1, 2\}$.

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

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

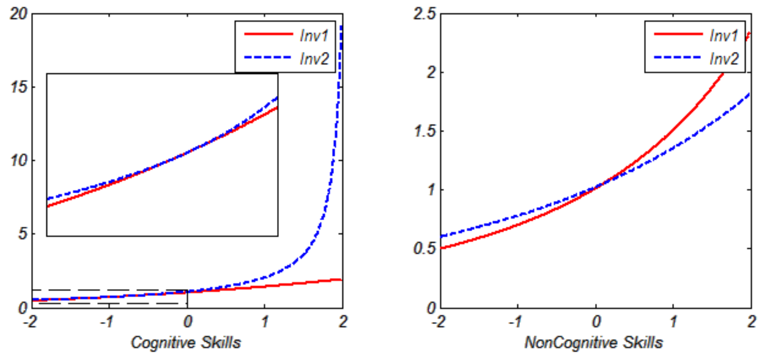
4. **Complementarity**: in the first phase of cognitive skills development, the elasticity of substitution between inputs is substantially larger than in the second phase. Noncognitive skills exhibit similar elasticities of substitution across phases.

First of all, in the first phase it is easier to increase cognitive skills; the amount of investment required to increase skills by 1 percent of a standard deviation is lower in the first phase with respect to the second. Figure 1 exemplifies this pattern by showing how much investment is required in each phase in order to increase skills by 1 percent of a standard deviation. The graph for cognitive skills clearly shows that the required amount of investment increases exponentially with starting skills in the second phase. Noncognitive skills do not exhibit such a clear pattern for the productivity of investment.

Given that returns to investment are larger in the first phase for cognitive skills, and similar for noncognitive skills, we should expect investment to be higher in early childhood rather than in later childhood, if households understand the properties of the technology.

Another feature of the technology is that investments in the second and first phase are strongly complementary: this happens because first phase investment enters second-

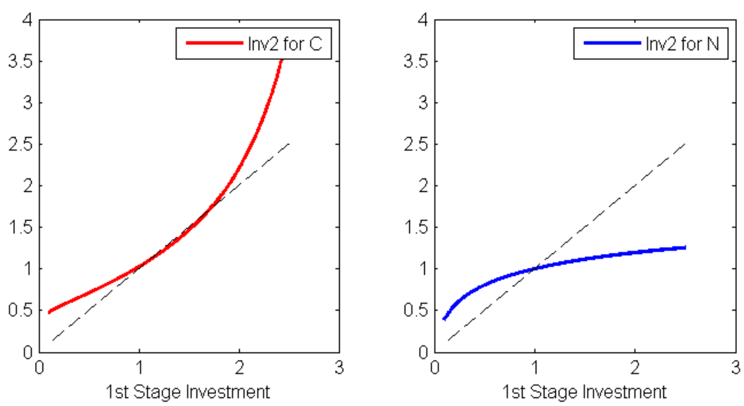
Figure 1. Investment required to increase skills



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

phase skills production through the self-productivity of future periods' skills. Hence, the more investment is performed today, the more it is required tomorrow, even just to keep skills constant.

Figure 2. Investment needed to maintain skills



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

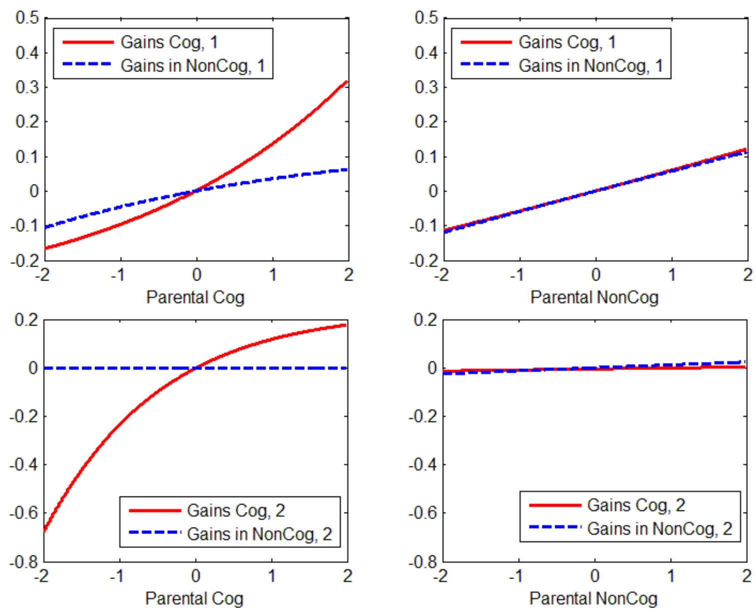
Figure 2 shows how much investment is required in order to keep skills constant in the second phase, after investing x units in the first phase, for a median household.

It is easy to see that the required amount of second-phase investment is increasing in first-phase investment for both skills.

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 phase will, on average, invest more also in the second phase.

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.

Figure 3. Skills gains by parental skills, for fixed investment



Note: Gain in skills (as fraction of a standard deviation) for a fixed quantity of investment, by log standardized parental skills and by developmental phase; initial child's skills fixed at the median.

For instance, when a mother's cognitive skills are one standard deviation above the median, the first-phase 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.

Appendix B Analytical Results

I rewrite the problem of the household and make explicit the Lagrange multipliers associated to the constraints. In every period $\tau \in \{1, .., 7\}$ of the parenthood stage, the problem of the household can be written as:

$$\max_{\{c_t, e_t, n_t, x_t, a_{t+1}\}_{t=\tau}^{\infty}} \mathbb{E}_\tau \Bigg[\sum_{t=\tau}^{\infty} \beta^{t-\tau} \frac{c_t^{1-\theta}}{1-\theta} - \sum_{t=\tau}^{7} \beta^{t-\tau} \zeta \frac{(n_t + \delta x_t)^{1+\sigma}}{1+\sigma} + \sum_{t=8}^{\infty} \beta^{t-\tau} \zeta \frac{(n_t)^{1+\sigma}}{1+\sigma}$$
$$+ \sum_{t=\tau}^{7} \beta^{t-\tau} \chi \left[\frac{(s_{C,t}^{\psi} s_{N,t}^{1-\psi})^{1-\xi}}{1-\xi} \right] + \sum_{t=8}^{\infty} \beta^{t-\tau} [\chi \frac{(s_{C,8}^{\psi} s_{N,8}^{1-\psi})^{1-\xi}}{1-\xi}] \Bigg]$$

subject to

$$\begin{array}{lll}
(\lambda_t) & c_t + e_t + a_{t+1} \leq (1+r)a_t + w(\Omega, \epsilon_t) & \forall t \in \{\tau, .., 7\} \\
(\lambda_t) & c_t + a_{t+1} \leq (1+r)a_t + w(\Omega, \epsilon_t)n_t & \forall t \in \{8, .., \infty\} \\
(\iota_t) & a_{t+1} \geq 0 & \forall t \\
(\mu_t) & I_t = A\left[\alpha_t x_t^{\omega} + (1-\alpha_t)e_t^{\omega}\right]^{1/\omega} & \forall t \in \{\tau, .., 7\}
\end{array}$$

and subject also to time constraints and to the skills formation technology in periods $t \in \{\tau, .., 7\}$, associating the multipliers $\kappa_{C,t}$ to the cognitive skills formation technology and $\kappa_{N,t}$ to the noncognitive skills formation technology, omitted from the above formulas for readability.

Assuming an interior solution for time allocation, such that $0 < x_t + n_t < 1$ and $n_t, x_t > 0$, the first-order conditions of the problem are as follows:

$$c_t: \quad c_t^{-\theta} = \lambda_t\,, \tag{16}$$
$$n_t: \quad \zeta(n_t + \delta x_t)^{\sigma} = \lambda_t w(\Omega, \epsilon_t)\,, \tag{17}$$
$$x_t: \quad \zeta\delta(n_t + \delta x_t)^{\sigma} = \mu_t A \alpha_t \frac{x_t^{\omega-1}}{\omega} \left[\alpha_t x_t^{\omega} + (1-\alpha_t)e_t^{\omega}\right]^{\frac{1-\omega}{\omega}}\,, \tag{18}$$
$$e_t: \quad \lambda_t = \mu_t A (1-\alpha_t) \frac{e_t^{\omega-1}}{\omega} \left[\alpha_t x_t^{\omega} + (1-\alpha_t)e_t^{\omega}\right]^{\frac{1-\omega}{\omega}} \tag{19}$$
$$a_{t+1}: \quad \lambda_t - \iota_t = \beta(1+r)\lambda_{t+1}\,. \tag{20}$$

In some cases, for some parametrizations of the CES technology of skills formation, it is possible that the derivative of the technology at investment $= 0$ is finite, and therefore in this case optimal investment may be equal to zero, in which case $x_t = e_t = 0$ and the above first-order conditions for x_t and e_t may not apply. In what follows, I show that this case is covered by the equation for optimal goods obtained below.

Substituting (17) into (18) yields

$$\delta \lambda_t w(\Omega, \epsilon_t) = \mu_t A \alpha_t \frac{x_t^{\omega-1}}{\omega} \left[\alpha_t x_t^{\omega} + (1-\alpha_t) e_t^{\omega}\right]^{\frac{1-\omega}{\omega}} . \tag{21}$$

Now, substituting (19) into (21) gives

$$\delta w(\Omega, \epsilon_t)(1-\alpha_t) \frac{e_t^{\omega-1}}{\omega} \left[\mu_t A \left[\alpha_t x_t^{\omega} + (1-\alpha_t) e_t^{\omega}\right]^{\frac{1-\omega}{\omega}}\right] = \alpha_t \frac{x_t^{\omega-1}}{\omega} \left[\mu_t A \left[\alpha_t x_t^{\omega} + (1-\alpha_t) e_t^{\omega}\right]^{\frac{1-\omega}{\omega}}\right]$$

and eliminating identical terms yields

$$\delta w(\Omega, \epsilon_t)(1-\alpha_t) e_t^{\omega-1} = \alpha_t x_t^{\omega-1} .$$

Finally, trivial manipulation yields

$$e_t^* = \left[\delta w(\Omega, \epsilon_t) \frac{1-\alpha_t}{\alpha_t}\right]^{\frac{1}{1-\omega}} x_t ,$$

obtaining equation (11) for optimal investment in goods in the main text. Notice that, even when households decide not to invest in the offspring, $x_t^* = 0$ and $e_t^* = 0$, so equation (11) obtains the correct solution even in corner cases.

There is one more case to be considered, and that is when the optimal hours of work choice $n_t^* = 0$. This is possible with these particular preferences because, when child care time $x > 0$, the marginal disutility of work at $n = 0$ is not equal to zero. If households choose to work zero hours, equation (17) does not hold and it is not possible to obtain the explicit equation for optimal goods (11). Thus, using only equations (18) and (19), I can write

$$\frac{c_t^{-\sigma}}{\zeta \delta (\delta x_t)^{\sigma}} = \frac{1-\alpha_t}{\alpha_t} \left[\frac{e_t^{\omega-1}}{x_t^{\omega-1}}\right] ,$$

and by rearranging I find that optimal invested goods when $n_t^* = 0$ solve the implicit equation:

$$e_t^{1-\omega} = \frac{1-\alpha_t}{\alpha_t} \left(\zeta \delta^{1+\sigma} x_t^{1-\omega+\sigma}\right) \left((1+r)a_t - a_{t+1} - e_t\right)^{\theta} \ . \tag{22}$$

In equation (22), I have substituted c_t with the solution from the budget constraint when $n_t = 0$. It is easy to see that, once again, the solution holds also when $x_t = 0$, and that optimal invested goods when $n_t^* = 0$ are increasing in time invested in the offspring x_t and in assets a_t for constant savings a_{t+1}.

Appendix C Solution Algorithm

The model features a relatively complex problem: a large state space, comprising three continuous state variables a, s_C, s_N, one discrete state variable ϵ and three dimensions of fixed heterogeneity, $s_{C,P}$, $s_{N,P}$ and ϵ_P, each affecting the problem nonlinearly. Since the problem of the household changes over time, I resort to Value Function iteration to solve the problem.

In order to simplify the problem, first-order conditions are used whenever possible to solve for the optimal solution given other controls. In practice, this means that the optimization algorithm only solves the value function for a' and x, while the other optimal choices are calculated by solving the First Order Conditions of the problem.

To find the solution, I use iterated maximization of the value function, with two nested Golden Search algorithms working to find the maximum of the value function first for x, then for a' given x. Golden Search is chosen as the maximization algorithm because it is the fastest and most reliable derivative-free algorithm, since derivative-based optimization can fail when the objective function is not necessarily differentiable at all points, which can be the case with a discretized state space and interpolation of future values.

The value function is maximized as follows: given x, optimal goods e are obtained by solving equation (11). Given a, a', e and the state variables, the budget constraint can be substituted into (17) to solve for optimal work hours n. n is found by finding the solution to the first-order condition 17 using the bisection method. If optimal work hours n are equal to zero, e is substituted with its solution from equation 22, given a, a' and x. This allows computing the value function for given values of x and a'. By the iterated maximum theorem, maximizing by one variable at a time is equivalent to maximizing for both at the same time, provided that the objective function is well-behaved. All objective functions are continuous and concave, and when fixing either x or a', the value function is continuous and concave in the other argument.

The solution algorithm works as follows:

1. The state space is discretized, placing continuous state variables on a grid. a is placed on a 25-points log-spaced grid, to allow for higher precision closer to the borrowing constraint, while the logs of s_C and s_N are placed on 11-points linearly spaced grids, which lower and upper bounds are allowed to go one standard deviation below the .1% lowest and above the 99.9% highest percentiles of skills

generated in simulation runs.

2. The algorithm starts from the final stage: here one only needs to solve for a', as there is no investment in the offspring and no time choice except for n, which is solved using the simpler first-order condition

$$\zeta n_t^{\sigma} = c_t^{-\theta} w(\Omega, \epsilon_t) \,,$$

where c_t is obtained from the budget constraint, as explained above. Solving this stage requires convergence of the value function, because time goes on to infinity. I iterate the solution until the supremum norm $\max |V_{k+1} - V_k|$ is lower than a tolerance parameter, set at 1e-6.

3. Next, the parenthood stage is solved backwards starting from $t = 7$ and ending at $t = 1$, using the solution method described above.

4. Finally, the fertile stage is solved, again until convergence of the value function.

Appendix D Details on Data and Moment Construction

D.1 ATUS data on child care time

I use the same data as Aguiar *et al.* (2021), which consists in the 2003-2017 waves of the American Time Use Survey (ATUS). The ATUS surveys a sample of Current Population Survey respondents, around three months after they exited the CPS. ATUS data consist of a 24-hour time diary of the previous day, split into 15-minute intervals, which records the activities in which respondents spent their time. Aguiar *et al.* (2021) group activites into six broad categories: market work, leisure, home production, job search, education and, most importantly for this study, childcare. I adopt their definition of child care time, which follows closely previous work by Guryan *et al.* (2008) and Aguiar *et al.* (2013).

To bring the data closer to the model equivalent of the unitary household, I restrict my attention to married individuals, between the ages of 25 and 54, who are not students, whose youngest child is younger than 15 years old and was born when they were between 20 and 44 years of age.

I target average child care time spent by individuals by college education and by age of the youngest child. In line with the technology of skills formation and the model period, I aggregate the youngest child's age in two-year intervals, following all definitions from Cunha *et al.* (2010), where the first interval corresponds to the first period of parenthood and includes ages 0-1 (the only period overlapping with the next in the data on assessments), the second period includes ages 1-2, the third includes ages 3-4 and so forth. To compute the targets, I first run the following regression, separately for men and women:

$$
\begin{aligned}
x_{i,g,t} = \quad & \sum_{k=1}^{7} \beta_{g,NC,k} \mathbb{I}(k = t \text{ \& Education}_i < \text{College})_{i,g,t} \\
&+ \sum_{k=1}^{7} \beta_{g,C,k} \mathbb{I}(k = t \text{ \& Education}_i = \text{College})_{i,g,t} + \Delta_g X_{i,g,t} + \epsilon_{i,g,t}\,,
\end{aligned}
\tag{23}
$$

where i indexes individuals, g indexes gender, t indexes the model period (or the age group of the child, equivalently), $\beta_{g,NC,k}$ are the coefficients associated to each gender $g \in \{M, F\}$, each period k and non-college-educated individuals, $\beta_{g,C,k}$ are the equivalent coefficients for college-educated individuals, $X_{i,g,t}$ are control variables, and $\mathbb{I}(k = t \text{ \& Education...})$ are dummy variables equal to one if the period is equal to t and if the education of the individual is as requested in the condition, and zero otherwise. Thus, after estimating the coefficients, I predict residualized average child care time, $\hat{x}_{g,e,t}$ by gender g, education e of the parent and age t of the child, by setting all controls to specific values, so that $X_{i,g,t} = X$, and $\epsilon_{i,g,t} = 0$. Specifically, I include in X age, race dummies, survey year dummies, the number of children under 18 in the household and state dummies. When I calculate $\hat{x}_{g,e,t}$, I set age to 35, race to white, the survey year to 2003, the number of children under 18 to one and the state to Alabama (dummy=1). None of these specific choices affects results in any other way than shifting all averages upwards or downwards, with no impact on the differences across education groups and ages. Observations are weighted with the ATUS recommended weight.

The procedure above produces as outputs $7 \times 2 \times 2 = 28$ numbers, representing average child care time spent in each period by parents of either gender, with college and noncollege education. It is useful to remember that, in the CNLSY/79 data, I cannot observe fathers, so the cognitive skills of the household in the model are mapped to the mother's skills, and the mother's skills are associated to the mother's education in the model through equation (13). In order to aggregate gender-specific averages and produce a household-level average, that is only time- and education-specific, I combine the gender-specific averages using a weight a that represents assortative matching. Thus, average child care time of households with a college-educated mother $\hat{x}_{e,t}$ is calculated as:

$$\hat{x}_{e,t} = \hat{x}_{F,C,t} + \left[a\hat{x}_{M,C,t} + (1-a)\hat{x}_{M,NC,t}\right] . \tag{24}$$

The assortative matching parameter $a = 0.6890 = 11.3/(24 - 7.6)$ is taken from the 2014 working paper version of Eika *et al.* (2019) [see (Eika *et al.*, 2014), available at `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.693.8503&rep=rep1&type=pdf`, Table C.1 in the Appendix], using the parameters estimated for 2007. It is calculated as the ratio between the percentage of women who are college-educated and who are married to a college-educated husband (11.3) and the total percentage of non-single college-educated women (24.0, the percentage of college-educated women, minus 7.6, those who are single).

One important underlying assumption of the above procedure is that child care time provided by the father and the mother are perfect substitutes and have equal weight in the household-level investment function. While this assumption is unlikely to hold in this form, estimating the elasticity of substitution between the father's time and the mother's time is challenging without better data. One might think that some activities can exclusively be performed by the mother, i.e. breastfeeding. However, I argue that a substantial degree of substitutability between the mother's and the father's time is necessary to rationalize the empirical evidence showing that, although children of single parents are at a disadvantage, such disadvantage is too small to be reconciled with strong complementarity between fathers' and mothers' time. For instance Carlson & Corcoran (2001) show that the difference in cognitive test scores between children raised in single-parent households, compared with those raised in intact households, is statistically insignificant after controlling for income and Army Force Qualification Test score of the mother.

Another important underlying assumption is that assortative matching, after accounting for education, is orthogonal to other observable and nonobservable characteristics. The above methodology is a compromise between using precise time use data and making do with available information. The methodology I adopt can in principle be amended to account for more complex assortative matching patterns, or data on time use of households, instead of individuals, can be used to better estimate the above averages.

After the above procedure, I am left with $7 \times 2 = 14$ targets $\hat{x}_{e,t}$, representing average child care time by the household's education. One final issue is that the standard error of $\hat{x}_{e,t}$ is not obvious, because it is a combination of the standard errors of the two gender-specific moments, and depends on the correlation between the specific measures $\hat{x}_{g,e,t}$. The higher the correlation, the higher the standard error of $\hat{x}_{e,t}$. Specifically:

$$\begin{aligned} VAR(\hat{x}_{e,t}) = \quad & VAR(\hat{x}_{F,C,t}) + a^2 VAR(\hat{x}_{M,C,t}) + (1-a^2) VAR(\hat{x}_{M,NC,t}) \\ & + a\rho_{FC,MC}\sqrt{VAR(\hat{x}_{F,C,t}) VAR(\hat{x}_{M,C,t})} \\ & + (1-a)\rho_{FC,MNC}\sqrt{VAR(\hat{x}_{F,C,t}) VAR(\hat{x}_{M,NC,t})} \\ & + a(1-a)\rho_{MC,MNC}\sqrt{VAR(\hat{x}_{M,C,t} VAR(\hat{x}_{M,NC,t})} \end{aligned}$$

where $\rho_{ge,g'e'}$ is the correlation between $\hat{x}_{g,e,t}$ and $\hat{x}_{g',e',t}$. I choose to err on the side of caution and assume a correlation coefficient of 1 across all measures, and use that to calculate the standard error of $\hat{x}_{e,t}$.

Further, I use the ATUS data to estimate the relationship between child care time and log earnings in all periods, motivated by Equation (14) as explained in the main text. I estimate the period- and education-specific auxiliary models

$$\begin{aligned} x_{i,g,e,t} = &\sum_{k=1}^{7} \beta_{g,e,k} \mathbb{I}(k=t) \log \text{Earnings}_{i,g,e,t} \\ &+ \Delta_{e,t} X_{i,e,t} + \epsilon_{i,e,t} \,, \end{aligned} \tag{25}$$

where I regress child care time on log earnings and the same controls X as in the regression in Equation (23). I then perform the same regression in model-generated data, and ask the model to get as close as possible to the age-education-specific coefficients estimated according to equation (25). This adds 14 more targets to the estimation, one for each period and education group.[27]

Finally, I also estimate the relationship between work hours and log earnings with an identical methodology. This adds a further 14 targets, again one for each period and education group. I perform the same regression between work hours and log earnings in model-generated data, and ask the model to get as close as possible to the estimated age-education-specific coefficients. Summarizing, I calculate a total of 42 targets using ATUS data.

D.2 IPUMS data on work hours and fertility

The Integrated Public Use Microdata Series is a publicly available dataset which includes variables from the Current Population Survey, a widely used cross-sectional survey that is representative of the US population. To be consistent with the ATUS sample described in the previous subsection, I use the 2000-2017 waves to construct series of average work hours by education, gender and age of the youngest child, and to obtain a measure of their relationship with household earnings. I use a procedure identical to that described in Subsection D.1, selecting the sample according to the same

[27] In principle, Equation (14) has the log of child care time as the left-hand variable. However, when using time use surveys that are obtained as 24-hour diaries, there may be days in which households report zero child care time. Estimating the equation in logs would require to throw these households out of the sample; to avoid this, I perform a regression in levels rather than in logs both in the data and in the model.

criteria, performing separate regressions by gender, and regressing actual individual hours of work on dummy variables that interact dummies for the age of the youngest child with education dummies. Controls include the same variables as in Subsection D.1. After obtaining the estimates, I aggregate these at the household level following the same procedure described above, with the same assortative matching parameter $a = 0.6890$. This gives a further 14 targets.

I also use the IPUMS data to obtain two further moments to be targeted in the estimation, that summarize the differential fertility of college-vs-noncollege educated individuals. I use the variable FREVER to estimate the fraction of college- and noncollege-educated individuals who ever had children before the age of 45. Summarizing, I calculate a total of 16 targets using IPUMS data.

D.3 CNLSY/79 data on skills, income and fertility

I use data from the Children of the National Longitudinal Survey of the Youth 1979 (CNLSY/79). The data from the CNLSY/79 is the same as in Cunha *et al.* (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 construct time-invariant log skill factors $\{\tilde{s}_{C,P}, \tilde{s}_{N,P}\}$ for the mother and time-varying factors $\{\tilde{s}_{C,t}, \tilde{s}_{N,t}\}$ for the child, using exactly the same variables as Cunha *et al.* (2010). Notice that these are not the exact data counterparts of the log of skills in the model $\{s_{C,P}, s_{N,P}, s_{C,t}, s_{N,t}\}$ because, as mentioned in the main text, factors have arbitrary mean and scale. For this reason, I only use correlations between each type of skill and over time as model targets, as correlations are mean- and scale-invariant. Table X provides basic statistics on the variables in the dataset for ages 5-6 and 13-14, showing that they match closely the results by CHS.

Table X. Summary statistics of variables used to identify latent skill factors.

	Mean	Std. Dev.	Skewness	N	R^2 of factor
Child's Cognitive Factor, Ages 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, Ages 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, Ages 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, Ages 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					
Arithmetic Reasoning Test Score	0.172	0.933	0.168	2207	83.7 %
Word Knowledge Test Score	0.302	0.822	-0.836	2207	70.9 %
Paragraph Composition Test Score	0.377	0.827	-1.121	2207	66.0 %
Numerical Operations Test Score	0.343	0.875	-0.469	2207	54.7 %
Coding Speed Test Score	0.468	0.879	-0.445	2207	41.1 %
Mathematical Knowledge Test Score	0.185	0.972	0.269	2207	77.4 %
Mother's NonCognitive Factor					
Self-Esteem: "I am a person of worth"	3.534	0.516	-0.343	2207	43.1 %
Self-Esteem: "I have good qualities"	3.382	0.531	0.025	2207	48.5 %
Self-Esteem: "I am a failure"	3.477	0.580	-0.649	2207	52.9 %
Self-Esteem: "I am as capable as others"	3.326	0.549	-0.217	2207	36.7 %
Self-Esteem: "I have nothing to be proud of"	3.480	0.625	-1.082	2207	46.0 %
Self-Esteem: "I have a positive attitude"	3.200	0.576	-0.250	2207	51.6 %
Self-Esteem: "I wish I had more self-respect"	2.876	0.787	-.206	2207	38.2 %
Self-Esteem: "I feel useless at times"	2.650	0.774	0.300	2207	32.5 %
Self-Esteem: "I sometimes think I am no good"	3.005	0.808	-0.298	2207	41.9 %
Rotter Score: "I have no control"	2.897	1.156	-0.600	2207	5.5 %
Rotter Score: "I make no plans for the future"	2.543	1.159	-0.002	2207	8.1 %
Rotter Score: "Luck is big factor in life"	3.154	0.974	-1.107	2207	4.5 %
Rotter Score: "Luck plays big role in my life"	2.426	1.144	-0.025	2207	2.5 %

Source: Author's calculations on CNSLY/79 data from Cunha *et al.* (2010)

After obtaining the variables above, I calculate the following moments, for both cognitive and noncognitive skills of the child: for each period $t \in \{1, ..7\}$, the correlation between a factor in period t and that factor in $t+1$ (that is, $\rho(s_{C,t}, s_{C,t+1}$

and $\rho(s_{N,t}, s_{N,t+1})$, so 14 moments); for each period $t \in \{1,..7\}$, the correlation between a factor in period t and the other factor in period $t+1$ (that is, $\rho(s_{C,t}, s_{N,t+1})$ and $\rho(s_{N,t}, s_{C,t+1})$, so again 14 moments); for each period $t \in \{1,..8\}$, the correlation of each factor with the mother's factors (that is, $\rho(s_{k,t}, s_{j,P})$, $\forall k \in \{C, N\}$ and $\forall j \in \{C, N\}$; so 32 moments); and finally, correlation of each factor with family income $\tilde{y}_t$, defined below, giving another 16 moments. Therefore, the estimation targets a total of 76 moments related to the offspring's skills development.

Regarding income, I use family income as the data equivalent of household income in the model. For all the auxiliary models that follow, I rescale the mother's skill factors so that they have the same variance as parental skills in the model (following the covariance matrix in Table XII in Appendix E.3). I first estimate the returns to cognitive and noncognitive skills of the mother by estimating the auxiliary Mincer equation:

$$\log \text{Income}_{i,t} = \beta_0 + \beta_C \bar{s}_{C,P} + \beta_N \bar{s}_{N,P} + \Gamma X_{i,t} + \epsilon_{i,t} \, , \tag{26}$$

where $\bar{s}_{C,P}$ is the mother's cognitive skill factor estimated from the data, $\bar{s}_{N,P}$ is the mother's noncognitive skill factor estimated from the data, and $X_{i,t}$ is a vector of controls that includes a quadratic polynomial in age and yearly dummies. Since household income combines asset choices and labor supply choices, which are endogenous with respect to unobserved fixed heterogeneity and to skills, both in the data and in the model, I again adopt an indirect inference approach. I perform the same regression in model-generated data and ask the model to replicate the same skill premia (two moments) that I estimate in the CNLSY/79 data.

I then construct residualized log income from log family income y_t: I regress log family income on a quadratic polynomial in the mother's age and year dummies, and use the regression to predict log income $\hat{y}_t$ using only these variables. I then use the residual $\tilde{y}_t = y_t - \hat{y}_t$ to construct the income-related moments targeted by the estimation. Notice that, after cleaning out differences across years and by the mother's age, household income still depends on the household's skills and on fixed heterogeneity across households, both in the data and in the model-generated data. For each period $t \in \{1,..,7\}$, I construct variance, skewness and curtosis of residualized log income $\tilde{y}_t$ (21 moments); for each period $t \in \{1,..,6\}$ I construct the covariance of residualized log income between t and $t+1$ (6 moments); for each period $t \in \{1,..,5\}$, I construct the covariance of residualized log income between t and $t+2$ (5 moments); finally, I divide residualized log income $\tilde{y}_t$ in quintiles and estimate the full Markov transition matrix across quintiles from one period to the next, aggregating across time periods (25 moments).[28] Finally, I calculate the household-level average over time $\tilde{y} = \frac{\sum_1^7 \tilde{y}_t}{7}$ to obtain a measure of lifetime log income, and calculate its variance across households, to obtain a target that is informative of persistent income inequality across households over lifetimes. This gives a total of 60 income-related moments; I estimate the associated standard errors via bootstrap with 100 repetitions.

Turning to the estimation of the fertility equation (8), I estimate the following auxiliary model on the CNLSY/79 sample of mothers older than 20:

$$\log d_i = \gamma_0 + \gamma_1 \tilde{s}_{C,P} + \gamma_2 \tilde{s}_{N,P} + \gamma_3 \bar{y}_i + \epsilon_i \tag{27}$$

where d_i is the number of years between when the mother was aged 20 and the birth of the first child; $\tilde{s}_{C,P}$ and $\tilde{s}_{N,P}$ are the mother's cognitive and noncognitive factors, respectively; and $\bar{y}_i$ is the previously estimated measure of lifetime log income. I cannot simply introduce the estimates of these coefficients in the fertility equation in the model for two reasons. First, the above regression can be biased, because the sample does not include mothers who never had children, who may be over-represented among higher-skilled and higher-income mothers. Second, family income is endogenous with respect to expected fertility and to unobserved characteristics, both in the data and in the model. To overcome these issues, I treat the parameter estimates of γ_0, γ_1, γ_2 and γ_3 as four additional moments targeted by the estimation, and I run an identical regression on model-simulated data, replicating the same empirical issues in the simulated moments. The fertility moments calculated in IPUMS data, as explained above, provide additional restrictions that are useful to estimate the model coefficients and the probability that a household stops being fertile p^{nf}.

D.4 USDA data on expenditure on children

I leverage the US Department of Agriculture data on household expenditure on children to estimate the shares of income that go in child-related expenses, and use these as

[28] In principle, 5 of these moments are redundant as a Markov transition matrix will always sum up to 1 on the rows. Using this restriction allows to use fewer moments and obtain tighter error bounds for estimates, but does not alter results substantially.

targets to estimate the shares of time α_t in the investment function (and conversely the share of goods $1 - \alpha_t$). I use data from the USDA report of 2011 (Lino, 2012), which include estimated annual expenditures on a child by middle-income husband-wife families, on average for the United States. I use the estimates for an average before-tax household income of 79,940, shown in Table 1 at page 26 of the report, and obtain the total costs that I consider the data equivalent of the model concept of "invested goods" e_t by summing the expenses in food, healthcare and childcare/education. These are further multiplied by 1.25, using the equivalence scale suggested in the report for households with a single child. Estimation results do not vary significantly if other costs are included, except that a higher share is obviously attributed to goods in that case. There are two difficulties with using this data in the estimation: First, the data does not align precisely with the model periods, because the USDA report uses 3-year intervals. To solve this problem, I interpolate values for intermediate periods: for instance, for the period corresponding to ages 3-4 in the model, expenses are calculated as the average between the average USDA expenses at age 3 and the average USDA expenses at age 4. The second difficulty is that standard errors of the estimates are not provided in the report; as the data used by the authors comprise 11,800 husband-wife households with children, and around one-third of them were classified in the middle-income group, standard errors are likely to be relatively small. I make the assumption that the USDA estimates are relatively precise and set their standard error to 3% of the calculated moments. For a point of comparison, consider that $\mathbb{VAR}(\bar{x}) = \mathbb{VAR}(x)/N$ where x is the variable of interest and N is the number of observations. Therefore, this standard error implies that, for a share of expenses of 8% of household income (which is approximately the average of the shares I calculate), the standard deviation of the share of household expenditures in children is:

$$\mathbb{STD}(x) = \mathbb{STD}(\bar{x}) \times \sqrt{N} = \underbrace{3\% \times 8\%}_{\mathbb{STD}(\bar{x})} \times \sqrt{11,800/3} \simeq 0.03 \times 0.08 \times 62.71 \simeq 0.15\,,$$

meaning that in the data it would not be uncommon to observe households spending nothing at all for their children in those categories, or more than three times the average share. Summarizing, this provides 7 additional targets to the estimation.

Appendix E Details of Estimation

E.1 Further discussion of identification

Let Θ be a parametrization of the model, $\hat{m}$ the $K \times 1$ vector of targets estimated in the data, $m_S(\Theta)$ the corresponding $K \times 1$ moments generated by S simulations of the model, and W the $K \times K$ weighting matrix. I follow Altonji & Segal (1996) and set W to be a diagonal matrix which entries are equal to the inverse of the variance of each corresponding moment.[29] Whenever the variance of each moment does not have an analytical solution, I calculate it via bootstrap with 100 repetitions. In practice, the estimation algorithm solves

$$\Theta^* = \arg\min_{\Theta}(m_S(\Theta) - \hat{m})^T W (m_S(\Theta) - \hat{m}),$$

where ^{T} denotes the transpose operator and Θ^* is the optimal solution. The Method of Simulated Moments estimator I use is consistent and asymptotically normally distributed under the regularity conditions in Pakes & Pollard (1989) and Duffie & Singleton (1990).

The model has been estimated with Differential Evolution (Storn & Price, 1997; Price, 2013), a greedy multistart evolutionary algorithm in which every node was obtaining its function value by further performing inner simplex optimization (Nelder & Mead, 1965) around the test point. The final estimate has further been refined with derivative-based methods.

I diagnose the final estimation results by verifying that the loss function is responsive to changes in parameters. If the loss function is flat around the optimum, this is an obvious sign of poor identification, as this means that the chosen moments are unresponsive when parameters change. To address this concern, I follow Adda *et al.* (2017) and I compute the change in the loss function for values of the parameters that are 1 percent away from the optimum (see Figure 4 in Appendix E.6). I find that the loss function is responsive to all changes, implying that the predicted moments are indeed sensitive to all the parameters.

[29] This choice is similar to other papers in the literature that use data from multiple sources, for instance Adda *et al.* (2017).

E.2 Normalizing final skills

To identify the scale of the investment function A, I add two restrictions, asking the model to generate final cognitive skills $s_{C,8}$ and noncognitive skills $s_{N,8}$ such that $\mathbb{E}[s_{C,8}] = \mathbb{E}[s_{N,8}] = 0$. Any choice for the standard error of these moments is arbitrary, and I associate a relatively small variance of $1/1000$.

The estimation strategy uses a total of 207 moments: 205 described in Appendix D, and the 2 normalizations described above.

E.3 Exogenously set parameters of the Technology of Skills Formation

Table XI. List of exogenous model parameters

Parameter	Value
Technology of skills formation	Cunha *et al.* (2010)
Covariance matrix, initial conditions	Cunha *et al.* (2010)
Duration of one period	2 years
Time endowment	200 hours/week
β	0.92
r	0.08

Table XII. Parametrization of Covariance/Correlation Matrices of offspring's and parental skills.

Covariance Matrix				
	$s_{C,1}$	$s_{N,1}$	$s_{C,P}$	$s_{N,P}$
$s_{C,1}$	0.1777			
$s_{N,1}$	-0.0204	0.2002		
$s_{C,P}$	0.0182	0.0592	0.5781	
$s_{N,P}$	0.0050	0.0261	0.0862	0.0667

Correlation Matrix				
	$s_{C,1}$	$s_{N,1}$	$s_{C,P}$	$s_{N,P}$
$s_{C,1}$	1.0000			
$s_{N,1}$	-0.1081	1.0000		
$s_{C,P}$	0.0569	0.1741	1.0000	
$s_{N,P}$	0.0463	0.2260	0.4390	1.0000

Source: Cunha *et al.* (2010), Appendix. *Note:* Variance/Covariance Matrix and Correlation Matrix for initial conditions of parental and offspring's skills. The offspring's skills are denoted $s_{C,1}$ (cognitive) and $s_{N,1}$ (noncognitive). The mother's skills are denoted $s_{C,P}$ (cognitive) and $s_{N,P}$ (noncognitive).

Table XIII. Parametrization of the technology for cognitive and noncognitive skills formation.

Technology of cognitive skills formation				
	1st Stage		2nd Stage	
Cognitive Skills	$\gamma_{1,C,1}$	0.485	$\gamma_{2,C,1}$	0.884
NonCognitive 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 Cognitive	$\gamma_{1,C,4}$	0.035	$\gamma_{2,C,4}$	0.051
Parental NonCognitive	$\gamma_{1,C,5}$	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				
Cognitive Skills	$\gamma_{1,N,1}$	0.000	$\gamma_{2,N,1}$	0.002
NonCognitive 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 Cognitive	$\gamma_{1,N,4}$	0.014	$\gamma_{2,N,4}$	0.000
Parental NonCognitive	$\gamma_{1,N,5}$	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

Source: Cunha *et al.* (2010) [pag. 919]. *Note:* parameters estimated by CHS taking into account investment endogeneity; skills linearly anchored to educational attainment; factors normally distributed.

Table XIV. Parametrization of the technology for **cognitive-only** skills formation.

Technology of cognitive skills formation				
	1st Stage		2nd Stage	
Cognitive Skills	$\gamma_{1,CO,1}$	0.303	$\gamma_{2,CO,1}$	0.448
Investment	$\gamma_{1,CO,3}$	0.319	$\gamma_{2,CO,3}$	0.098
Parental Cognitive	$\gamma_{1,CO,4}$	0.378	$\gamma_{2,CO,4}$	0.454
Complementarity	$\phi_{1,CO}$	-0.180	$\phi_{2,CO}$	-0.781
Elasticity of Substitution $1/(1-\phi)$		0.847		0.562
Variance of Shocks	$\eta_{1,CO}$	0.193	$\eta_{2,CO}$	0.050

Source: Cunha *et al.* (2010) [pag. 45 of Appendix A14]. *Note:* parameters estimated by CHS taking into account investment endogeneity and unobserved heterogeneity; skills linearly anchored to educational attainment; factors normally distributed.

E.4 Auxiliary models used in estimation

Table XV. Probit regression of being college-educated.

	College
Mother's Cognitive Factor	1.394***
	(0.078)
Mother's Noncognitive Factor	0.181
	(0.170)
Constant	-1.160***
	(0.055)
Observations	1581
Pseudo-R2	0.278

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: CNLSY/79. *Note:* Author's calculations. The mother's skill factors are scaled as in the model.

Table XVI. Mincer regression: income premia to cognitive and noncognitive skills.

	(1)
	Log Family Income
Mother's cognitive factor	0.306***
	(0.011)
Mother's noncognitive factor	0.231***
	(0.030)
Constant	-0.889***
	(0.256)
Observations	13467
R2	0.262

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: CNLSY/79. *Note:* Author's calculations. The mother's skill factors are scaled as in the model. Controls include a quadratic polynomial in the mother's age and year dummies.

E.5 Goodness of fit

Table XVII. Goodness of fit: average time allocation choices

	College		Non-college	
Moment	Data	Simulated	Data	Simulated
Hours of work, child aged 0-1	0.295 (0.00321)	0.289	0.266 (0.00305)	0.280
Hours of work, child aged 2-3	0.314 (0.00343)	0.307	0.283 (0.00316)	0.295
Hours of work, child aged 4-5	0.323 (0.00371)	0.300	0.294 (0.00336)	0.288
Hours of work, child aged 6-7	0.326 (0.00389)	0.328	0.304 (0.00351)	0.318
Hours of work, child aged 8-9	0.339 (0.00403)	0.330	0.313 (0.00364)	0.320
Hours of work, child aged 10-11	0.346 (0.00417)	0.334	0.320 (0.00377)	0.324
Hours of work, child aged 12-13	0.354 (0.00428)	0.338	0.325 (0.00388)	0.329
Child care time, child aged 0-1	0.200 (0.01053)	0.142	0.169 (0.01054)	0.141
Child care time, child aged 2-3	0.149 (0.01069)	0.138	0.126 (0.01061)	0.137
Child care time, child aged 4-5	0.123 (0.01102)	0.152	0.108 (0.01082)	0.150
Child care time, child aged 6-7	0.101 (0.01138)	0.083	0.095 (0.01097)	0.081
Child care time, child aged 8-9	0.085 (0.01155)	0.073	0.073 (0.01124)	0.071
Child care time, child aged 10-11	0.071 (0.01184)	0.061	0.061 (0.01144)	0.059
Child care time, child aged 12-13	0.054 (0.01210)	0.046	0.044 (0.01161)	0.045

Source: ATUS 2003-2017 and IPUMS 2001-2017. *Note:* Author's calculations. Standard errors of moments in parentheses.

Table XVIII. Goodness of fit: shares of income spent on children

Moment	Data	Simulated
Share of income spent on goods invested, age 0-1	0.080 (0.00895)	0.067
Share of income spent on goods invested, age 2-3	0.079 (0.00891)	0.083
Share of income spent on goods invested, age 4-5	0.079 (0.00887)	0.124
Share of income spent on goods invested, age 6-7	0.074 (0.00859)	0.077
Share of income spent on goods invested, age 8-9	0.080 (0.00895)	0.081
Share of income spent on goods invested, age 10-11	0.086 (0.00928)	0.080
Share of income spent on goods invested, age 12-13	0.092 (0.00961)	0.070

Source: USDA 2013 report. *Note:* Author's calculations. Standard errors of moments in parentheses.

Table XIX. Goodness of fit: correlation between child's skills and mother's skills

	Mother's cognitive		Mother's noncognitive	
Moment	Data	Simulated	Data	Simulated
ρ(child's cognitive,mother's skills)				
at age 0-1	-0.062 (0.07203)	0.051	0.059 (0.07204)	0.043
at age 2-3	0.092 (0.06128)	0.145	0.152 (0.06083)	0.150
at age 4-5	0.356 (0.04678)	0.187	0.242 (0.04857)	0.195
at age 6-7	0.284 (0.04386)	0.213	0.153 (0.04520)	0.221
at age 8-9	0.268 (0.02612)	0.326	0.126 (0.02690)	0.240
at age 10-11	0.357 (0.02544)	0.381	0.183 (0.02676)	0.245
at age 12-13	0.420 (0.02610)	0.412	0.213 (0.02810)	0.246
at age 14	0.430 (0.02787)	0.428	0.184 (0.03035)	0.244
ρ(child's noncognitive,mother's skills)				
at age 0-1	0.068 (0.06968)	0.162	0.130 (0.06925)	0.211
at age 2-3	0.197 (0.03497)	0.153	0.174 (0.03513)	0.217
at age 4-5	0.192 (0.05733)	0.159	0.223 (0.05695)	0.231
at age 6-7	0.210 (0.02584)	0.168	0.202 (0.02588)	0.244
at age 8-9	0.209 (0.02555)	0.160	0.203 (0.02558)	0.230
at age 10-11	0.183 (0.02624)	0.154	0.205 (0.02612)	0.219
at age 12-13	0.140 (0.02769)	0.153	0.142 (0.02768)	0.210
at age 14	0.134 (0.02976)	0.154	0.144 (0.02971)	0.203

Source: CNLSY/79. *Note:* Author's calculations. Standard errors of moments in parentheses.

Table XX. Goodness of fit: correlation of child's skills over time

	Cognitive in $t+1$		Noncognitive in $t+1$	
Moment	Data	Simulated	Data	Simulated
ρ(child's cognitive,next period skill)				
at age 0-1	0.053 (0.22909)	0.407	0.016 (0.27732)	-0.052
at age 2-3	0.338 (0.13310)	0.477	0.136 (0.09320)	0.042
at age 4-5	-0.018 (0.28863)	0.501	0.255 (0.05206)	0.080
at age 6-7	0.676 (0.03696)	0.773	0.243 (0.04797)	0.109
at age 8-9	0.788 (0.01859)	0.811	0.145 (0.02979)	0.122
at age 10-11	0.826 (0.01721)	0.831	0.187 (0.02959)	0.127
at age 12-13	0.839 (0.01786)	0.839	0.224 (0.03167)	0.132
ρ(child's noncognitive,next period skill)				
at age 0-1	0.068 (0.23516)	0.047	-0.308 (0.28686)	0.546
at age 2-3	0.192 (0.06063)	0.121	0.228 (0.13250)	0.609
at age 4-5	0.330 (0.13217)	0.149	0.688 (0.04707)	0.631
at age 6-7	0.192 (0.02924)	0.141	0.688 (0.02101)	0.843
at age 8-9	0.199 (0.02887)	0.144	0.732 (0.01985)	0.845
at age 10-11	0.212 (0.02963)	0.146	0.735 (0.02019)	0.845
at age 12-13	0.241 (0.03129)	0.148	0.703 (0.02246)	0.843

Source: CNLSY/79. *Note:* Author's calculations. Standard errors of moments in parentheses.

Table XXI. Goodness of fit: correlation of child's skills and household income

	Cognitive skills		Noncognitive skills	
Moment	Data	Simulated	Data	Simulated
ρ(skills, log income)				
child aged 0-1	-0.035 (0.07733)	0.036	0.092 (0.07381)	0.113
child aged 2-3	0.082 (0.06266)	0.088	0.093 (0.03678)	0.087
child aged 4-5	0.260 (0.04967)	0.122	0.240 (0.05908)	0.097
child aged 6-7	0.150 (0.04735)	0.150	0.168 (0.02736)	0.111
child aged 8-9	0.168 (0.02809)	0.194	0.185 (0.02711)	0.113
child aged 10-11	0.173 (0.02834)	0.219	0.198 (0.02765)	0.116
child aged 12-13	0.235 (0.03014)	0.234	0.177 (0.02969)	0.123
child aged 14	0.236 (0.03243)	0.252	0.182 (0.03200)	0.137

Source: CNLSY/79. *Note:* Author's calculations. Standard errors of moments in parentheses.

Table XXII. Goodness of fit: relationship between time use and household income

	College		Noncollege	
Beta of regression on log income of...	Data	Simulated	Data	Simulated
Child care time				
child aged 0-1	0.008 (0.00413)	-0.006	0.004 (0.00222)	-0.003
child aged 2-3	0.003 (0.00413)	-0.013	-0.003 (0.00223)	-0.013
child aged 4-5	-0.001 (0.00417)	-0.013	-0.006 (0.00223)	-0.013
child aged 6-7	-0.008 (0.00414)	-0.011	-0.008 (0.00224)	-0.009
child aged 8-9	-0.011 (0.00425)	-0.010	-0.010 (0.00224)	-0.008
child aged 10-11	-0.012 (0.00428)	-0.008	-0.013 (0.00225)	-0.006
child aged 12-13	-0.016 (0.00434)	-0.006	-0.016 (0.00227)	-0.004
Hours of work				
child aged 0-1	0.038 (0.00413)	0.024	0.043 (0.00454)	0.020
child aged 2-3	0.042 (0.00412)	0.062	0.046 (0.00452)	0.055
child aged 4-5	0.041 (0.00422)	0.064	0.043 (0.00458)	0.057
child aged 6-7	0.043 (0.00422)	0.054	0.048 (0.00452)	0.048
child aged 8-9	0.044 (0.00432)	0.054	0.048 (0.00456)	0.048
child aged 10-11	0.046 (0.00438)	0.053	0.047 (0.00460)	0.048
child aged 12-13	0.046 (0.00457)	0.052	0.046 (0.00467)	0.047

Source: ATUS 2003-2017. *Note:* Author's calculations. Standard errors of moments in parentheses.

Table XXIII. Goodness of fit: income- and fertility-related moments

Moment	Data	Simulated
Mincer return to cognitive skills	0.306 (0.01087)	0.354
Mincer return to noncognitive skills	0.231 (0.03040)	0.278
Variance of lifetime log household income	0.317 (0.00942)	0.297
Asymmetry of income (child aged 0-1)	-0.851 (0.06082)	0.020
Asymmetry of income (child aged 2-3)	-0.988 (0.06329)	-0.858
Asymmetry of income (child aged 4-5)	-0.874 (0.06687)	-0.893
Asymmetry of income (child aged 6-7)	-0.746 (0.07090)	-0.649
Asymmetry of income (child aged 8-9)	-0.638 (0.06505)	-0.626
Asymmetry of income (child aged 10-11)	-0.626 (0.07754)	-0.591
Asymmetry of income (child aged 12-13)	-0.616 (0.08440)	-0.549
Curtosis of income (child aged 0-1)	3.717 (0.21196)	2.178
Curtosis of income (child aged 2-3)	4.455 (0.17935)	4.829
Curtosis of income (child aged 4-5)	4.167 (0.20659)	4.888
Curtosis of income (child aged 6-7)	3.880 (0.17433)	3.983
Curtosis of income (child aged 8-9)	3.769 (0.19536)	3.903
Curtosis of income (child aged 10-11)	3.588 (0.17695)	3.794
Curtosis of income (child aged 12-13)	3.661 (0.22222)	3.667
Covariance of income, one period ahead (child aged 0-1)	0.226 (0.01626)	0.280
Covariance of income, one period ahead (child aged 2-3)	0.276 (0.01888)	0.261
Covariance of income, one period ahead (child aged 4-5)	0.253 (0.01518)	0.265
Covariance of income, one period ahead (child aged 6-7)	0.272 (0.01799)	0.267
Covariance of income, one period ahead (child aged 8-9)	0.289 (0.02104)	0.268
Covariance of income, one period ahead (child aged 10-11)	0.328 (0.02618)	0.268
Covariance of income, two periods ahead (child aged 0-1)	0.210 (0.01883)	0.277
Covariance of income, two periods ahead (child aged 2-3)	0.219 (0.01725)	0.278
Covariance of income, two periods ahead (child aged 4-5)	0.206 (0.01394)	0.275
Covariance of income, two periods ahead (child aged 6-7)	0.244 (0.02023)	0.277
Covariance of income, two periods ahead (child aged 8-9)	0.226 (0.02224)	0.278
Coefficients of regression of log fertility duration		
constant	1.670 (0.01979)	1.663
mother's cognitive skills	0.316 (0.03049)	0.289
mother's noncognitive skills	0.086 (0.08161)	0.061
lifetime log income	0.089 (0.03747)	0.034
Fraction of college-educated parents with child at age 45	0.822 (0.01427)	0.787
Fraction of noncollege-educated parents with child at age 45	0.874 (0.00785)	0.878

Source: CNLSY/79 and IPUMS 2001-2017. *Note:* Author's calculations. Standard errors of moments in parentheses.

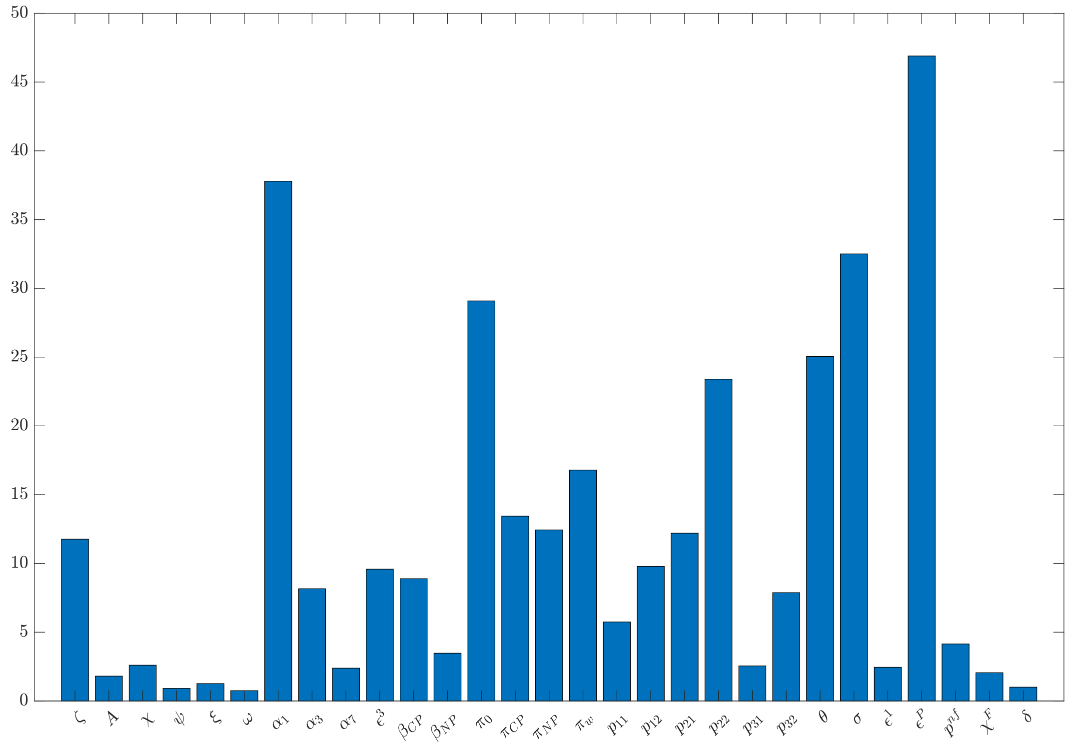
Table XXIV. Goodness of fit: Markov transition matrix of income

	Q1		Q2		Q3		Q4		Q5	
	Data	Sim	Data	Sim	Data	Sim	Data	Sim	Data	Sim
Q1	0.580	0.585	0.228	0.288	0.095	0.067	0.057	0.030	0.040	0.030
	(0.015)		(0.014)		(0.009)		(0.007)		(0.005)	
Q2	0.195	0.235	0.410	0.428	0.240	0.236	0.097	0.100	0.058	0.002
	(0.011)		(0.014)		(0.014)		(0.009)		(0.007)	
Q3	0.078	0.079	0.204	0.149	0.394	0.410	0.229	0.250	0.096	0.112
	(0.009)		(0.013)		(0.014)		(0.013)		(0.008)	
Q4	0.065	0.046	0.095	0.076	0.200	0.164	0.416	0.431	0.225	0.283
	(0.008)		(0.008)		(0.014)		(0.016)		(0.013)	
Q5	0.059	0.046	0.059	0.003	0.090	0.099	0.211	0.227	0.581	0.625
	(0.008)		(0.008)		(0.009)		(0.014)		(0.015)	

Source: CNLSY/79. *Note:* Author's calculations. Standard errors of moments in parentheses.

E.6 Estimation diagnostics

Figure 4. Loss function response to changes in parameters



Note: Each column corresponds to the percentage change in the loss function when the corresponding parameter is increased by 1 percent of its estimated value.

Appendix F Additional policy results

Table XXV. Returns to delaying fertility and childcare policies, cognitive-skills only model.

Variable	Baseline	Fertility Delay	Childcare	
			Full	Partial
Cognitive skills	0.00	+0.13	-0.38	-0.45
Never constrained	+1.34	-0.06	-0.42	-0.49
& high income, high skills	+126.62	+0.08	-5.93	-6.10
Constrained late	-6.03	+1.29	-0.48	-0.55
Constrained early	-8.99	+0.86	+0.20	+0.12
Constrained early & late	-10.13	+2.32	+0.25	+0.17
& low income, low skills	-48.05	+2.81	+1.93	+1.87
Correlation with $s_{C,P}$	91.05	91.02	90.39	90.37
Correlation with $s_{N,P}$	35.28	35.05	35.00	34.99
Correlation with income	47.58	47.37	46.98	46.99
Work hours, early	68.55	-0.33	+0.63	0.00
Work hours, late	66.47	-0.12	-0.07	+0.00
Child care time, early	2.88	+0.02	+0.00	+0.00
Child care time, late	5.76	+0.01	-0.00	-0.01
Goods invested, early	-	+0.66	+3.34	+3.18
Goods invested, late	-	+0.23	+0.05	-0.08
Assets at birth	18.93	+2.78	-0.21	0.00
Ever constrained	15.16	-1.06	-0.31	-0.00
early	7.01	-0.59	+0.08	0.00
late	9.15	-0.56	-0.38	-0.00

Note: Skills and goods are in log units $\times$ 100. Changes are expressed in log points deviation normalized by the standard deviation of skills. Time is in hours per week. Assets are in permanent income units, that is, $a/w(\Omega, 0)$. Low- (High-) permanent income households are defined as those with $\epsilon_P < 0$ $(= 0)$. Low- (High-) skilled households are defined as those with cognitive skills below (equal to or above) the median.

Table XXVI. Robustness: results when all technologies of skills formation are Cobb-Douglas.

Variable	Baseline	Tax Credit 5%	Tax Credit 25%	Kindergeld 5%	Kindergeld 25%
Cognitive skills	-	+0.77	+3.68	+1.72	+7.76
Never constrained	+2.87	+0.75	+3.56	+1.48	+6.84
& high income, high skills	+51.19	+0.51	+2.39	+0.56	+2.69
Constrained late	-6.16	+0.83	+3.96	+2.14	+9.60
Constrained early	-3.01	+0.77	+3.69	+2.14	+8.96
Constrained early & late	-10.12	+0.87	+4.14	+3.09	+12.22
& low income, low skills	-28.72	+1.04	+4.97	+4.57	+17.56
Standard deviation	100.00	+0.17	+0.84	+1.15	+4.51
Correlation with $s_{C,P}$	33.59	33.51	33.19	33.09	31.51
Correlation with $s_{N,P}$	20.63	20.58	20.39	20.35	19.44
Correlation with income	21.20	21.09	20.68	20.53	18.22
Noncognitive skills	-	+1.22	+5.86	+2.29	+10.06
Never constrained	+3.85	+1.19	+5.72	+1.93	+8.72
& high income, high skills	+28.31	+0.84	+4.02	+0.71	+3.47
Constrained late	-9.18	+1.29	+6.19	+3.01	+12.99
Constrained early	+1.06	+1.21	+5.80	+2.64	+11.02
Constrained early & late	-11.66	+1.31	+6.30	+3.81	+15.08
& low income, low skills	-26.12	+1.55	+7.46	+5.63	+21.53
Standard deviation	100.00	+0.24	+1.18	+1.56	+5.89
Correlation with $s_{C,P}$	14.35	14.22	13.70	13.59	11.43
Correlation with $s_{N,P}$	18.06	17.98	17.69	17.64	16.44
Correlation with income	14.15	13.98	13.35	13.25	10.51
Work hours, early	54.42	+0.16	+0.80	-1.42	-8.16
Work hours, late	65.35	+0.19	+0.95	-0.86	-4.25
Child care time, early	32.66	+0.06	+0.29	+0.83	+4.20
Child care time, late	10.84	+0.05	+0.26	+0.28	+1.29
Goods invested, early	-	+2.67	+13.23	+2.47	+11.56
Goods invested, late	-	+2.97	+14.73	+2.36	+10.07
Assets at birth	27.88	+0.22	+1.15	-2.75	-10.67
Ever constrained	31.77	+0.12	+0.51	+3.61	+14.39
early	6.27	-0.01	-0.09	+0.37	+16.98
late	28.26	+0.13	+0.61	+3.44	+4.77

Note: Skills and goods are in log units × 100. Changes are expressed in log points deviation normalized by the standard deviation of skills. Time is in hours per week. Assets are in permanent income units, that is, $a/w(\Omega, 0)$. Low- (High-) permanent income households are defined as those with $\epsilon_P < 0$ $(= 0)$. Low- (High-) skilled households are defined as those with cognitive skills below (equal to or above) the median.

Table XXVII. Robustness: results for different values of the relative importance of cognitive skills for the utility of households.

	Baseline	$\psi = 0.01$	$\psi = 0.1$	$\psi = 0.2$
Kindergeld	1.68	1.67	1.70	1.74
Tax credit	0.88	0.87	0.89	0.90

Note: Changes are expressed in 100 $\times$ log points deviation normalized by the standard deviation of skills.

References

ABBOTT, BRANT. 2021. Incomplete markets and parental investments in children. *Review of Economic Dynamics*.

ABBOTT, BRANT, GALLIPOLI, GIOVANNI, MEGHIR, COSTAS, & VIOLANTE, GIOVANNI L. 2019. Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, **127**(6), 2569–2624.

ADDA, JÉRÔME, DUSTMANN, CHRISTIAN, & STEVENS, KATRIEN. 2017. The career costs of children. *Journal of Political Economy*, **125**(2), 293–337.

AGOSTINELLI, FRANCESCO, & WISWALL, MATTHEW. 2020. Estimating the Technology of Children's Skill Formation. *working paper*.

AGUIAR, MARK, HURST, ERIK, & KARABARBOUNIS, LOUKAS. 2013. Time use during the great recession. *American Economic Review*, **103**(5), 1664–96.

AGUIAR, MARK, BILS, MARK, CHARLES, KERWIN KOFI, & HURST, ERIK. 2021. Leisure Luxuries and the Labor Supply of Young Men. *Journal of Political Economy*, **129**(2), 337–382.

AIYAGARI, S. RAO. 1994. Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics*, **109**(3).

AIYAGARI, S RAO, GREENWOOD, JEREMY, & SESHADRI, ANANTH. 2002. Efficient investment in children. *Journal of economic theory*, **102**(2), 290–321.

ALTONJI, JOSEPH G, & SEGAL, LEWIS M. 1996. Small-sample bias in GMM estimation of covariance structures. *Journal of Business & Economic Statistics*, **14**(3), 353–366.

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. 1964. Human capital theory. *Columbia, New York*, **1964**.

Becker, Gary S, & Tomes, Nigel. 1979. An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of political Economy*, **87**(6), 1153–1189.

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).

Blau, Francine D, & Kahn, Lawrence M. 2007. Changes in the labor supply behavior of married women: 1980–2000. *Journal of Labor economics*, **25**(3), 393–438.

Blundell, Richard, & MaCurdy, Thomas. 1999. Labor supply: A review of alternative approaches. *Handbook of labor economics*, **3**, 1559–1695.

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

Brilli, Ylenia. 2015. Mother's time allocation, child care and child cognitive development.

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

Carneiro, Pedro Manuel, Lopez Garcia, Italo, Salvanes, Kjell G, & Tominey, Emma. 2015. Intergenerational mobility and the timing of parental income. *NHH Dept. of Economics Discussion Paper*.

Caucutt, Elizabeth M, & Lochner, Lance. 2020. Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy*, **128**(3), 1065–1147.

Caucutt, Elizabeth M, Lochner, Lance, Mullins, Joseph, & Park, Youngmin. 2020. *Child skill production: Accounting for parental and market-based time and goods investments*. Tech. rept. National Bureau of Economic Research.

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., Lochner, Lance, & Masterov, Dimitriy V. 2006. Interpreting the Evidence on Life Cycle Skill Formation. *Chap. 12 of:* North-Holland, Amsterdam: (ed), *Handbook of the Economics of Education.* E. A. Hanushek and F. Welch.

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

Dahl, Gordon B, & Lochner, Lance. 2012. The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*, **102**(5), 1927–56.

Daruich, Diego. 2018. The macroeconomic consequences of early childhood development policies. *FRB St. Louis Working Paper.*

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

Duffie, Darrell, & Singleton, Kenneth J. 1990. *Simulated moments estimation of Markov models of asset prices.*

Eika, Lasse, Mogstad, Magne, & Zafar, Basit. 2014. Educational assortative mating and household income inequality. *Working Paper.*

Eika, Lasse, Mogstad, Magne, & Zafar, Basit. 2019. Educational assortative mating and household income inequality. *Journal of Political Economy*, **127**(6), 2795–2835.

Griffen, Andrew S. 2019. Evaluating the Effects of Childcare Policies on Children's Cognitive Development and Maternal Labor Supply. *Journal of Human Resources*, **54**(3), 604–655.

Guryan, Jonathan, Hurst, Erik, & Kearney, Melissa. 2008. Parental education and parental time with children. *Journal of Economic perspectives*, **22**(3), 23–46.

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, Pinto, Rodrigo, & Savelyev, Peter. 2013. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, **103**(6), 2052–86.

Heckman, James J, & Mosso, Stefano. 2014. The economics of human development and social mobility. *Annu. Rev. Econ.*, **6**(1), 689–733.

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

Heckman, James J, Malofeeva, Lena, Pinto, Rodrigo, & Savelyev, Peter A. 2007. The effect of the Perry Preschool Program on the cognitive and noncognitive skills of its participants. *University of Chicago, Department of Economics. Unpublished*.

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. 1993. The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of economic Dynamics and Control*, **17**(5-6), 953–969.

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. 2019. On the intergenerational transmission of economic status. *Journal of Political Economy*, **127**(2), 855–921.

LINO, MARK. 2012. *Expenditures on Children by Families, 2011.* Center for Nutrition Policy and Promotion, US Department of Agriculture.

LOCHNER, LANCE, & MONGE-NARANJO, ALEXANDER. 2012. Credit constraints in education. *Annu. Rev. Econ.*, **4**(1), 225–256.

LOCHNER, LANCE J, & MONGE-NARANJO, ALEXANDER. 2011. The nature of credit constraints and human capital. *American economic review*, **101**(6), 2487–2529.

MCFADDEN, DANIEL. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 995–1026.

MORRIS, PAMELA, DUNCAN, GREG J, & CLARK-KAUFFMAN, ELIZABETH. 2005. Child well-being in an era of welfare reform: the sensitivity of transitions in development to policy change. *Developmental psychology*, **41**(6), 919.

NELDER, JOHN A, & MEAD, ROGER. 1965. A simplex method for function minimization. *The computer journal*, **7**(4), 308–313.

NORES, MILAGROS, BELFIELD, CLIVE R, BARNETT, W STEVEN, & SCHWEINHART, LAWRENCE. 2005. Updating the economic impacts of the High/Scope Perry Preschool program. *Educational Evaluation and Policy Analysis*, **27**(3), 245–261.

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

Pakes, Ariel, & Pollard, David. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, 1027–1057.

Price, Kenneth V. 2013. Differential evolution. *Pages 187–214 of: Handbook of optimization.* Springer.

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).

Schoellman, Todd. 2016. Early childhood human capital and development. *American Economic Journal: Macroeconomics*, **8**(3), 145–74.

Soytas, Mehmet, Golan, Limor, Gayle, George-Levi, *et al.* 2014. What accounts for the racial gap in time allocation and intergenerational transmission of human capital? *In: 2014 Meeting Papers.* Society for Economic Dynamics.

Storn, Rainer, & Price, Kenneth. 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, **11**(4), 341–359.

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

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

Todd, Petra E, & Wolpin, Kenneth I. 2007b. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital*, **1**(1), 91–136.

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

YUM, MINCHUL. 2020. Parental time investment and intergenerational mobility. *Available at SSRN 3862378.*