Opportunity & Inclusive Growth Institute
FEDERAL RESERVE BANK *of* MINNEAPOLIS

The Macroeconomic Consequences of Early Childhood Development Policies

Diego Daruich
Federal Reserve Bank of St. Louis

System Working Paper 18-18
December 2018

This paper was originally published as *Working Paper 2018-029A* by the Federal Reserve Bank of St. Louis. This paper may be revised. The most current version is available at https://doi.org/10.20955/wp.2018.029.

RESEARCH DIVISION

Working Paper Series

The Macroeconomic Consequences of Early Childhood Development Policies

Diego Daruich

Working Paper 2018-029A
https://doi.org/10.20955/wp.2018.029

October 2018

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

The Macroeconomic Consequences of Early Childhood Development Policies

Diego Daruich*
Federal Reserve Bank of St. Louis

September 26, 2018
Click here for latest version

Abstract

To study long-run large-scale early childhood policies, this paper incorporates early childhood investments into a standard general-equilibrium (GE) heterogeneous-agent overlapping-generations model. After estimating it using US data, we show that an RCT evaluation of a short-run small-scale early childhood program in the model predicts effects on children's education and income that are similar to the empirical evidence. A long-run large-scale program, however, yields twice as large welfare gains, even after considering GE and taxation effects. Key to this difference is that investing in a child not only improves her skills but also creates a better parent for the next generation.

JEL Classifications: J13, J24, J62.
Keywords: Inequality, intergenerational mobility, early childhood development.

*ddaruich@nyu.edu. I am grateful to my advisors Raquel Fernandez and Gianluca Violante. For helpful comments, I thank Christopher Flinn, Julian Kozlowski, and Pau Roldan. I also received valuable feedback from participants at seminars and conference sessions at New York University, the Federal Reserve Board, the Federal Reserve Bank of St. Louis, the Econometric Society, the SED, Federal Reserve Bank of Minneapolis, University of Virginia, University of Wisconsin-Madison, Northwestern University, CEMFI, the World Bank, University of Maryland, UC San Diego, USC Marshall, University of Toronto (Rotman), University of Rochester, and the NBER Summer Institute. The views expressed in this paper are solely my responsibility and should not be interpreted as reflecting the views of the Federal Reserve Bank of St. Louis. This paper was previously circulated under the title "From Childhood to Adult Inequality: Parental Investments and Early Childhood Development."

1 Introduction

Early childhood environment has been shown to significantly impact adult outcomes.[1] García, Heckman, Leaf, and Prados (2017), for example, estimate that for every dollar invested on an early childhood program, children's lifetime labor income increases by 1.3 dollars. This evidence suggests that if these programs were scaled up they would increase welfare, reduce inequality, and increase intergenerational mobility. A large-scale program, however, would be associated with taxation and general equilibrium (GE) effects that cannot be accounted for in small-scale empirical studies. Macroeconomic models of inequality and mobility are well suited to study such effects, but they generally ignore the role of endogenous early childhood development. This paper fills this gap by incorporating early childhood development into a standard macroeconomic model. It shows that underinvestment in children's development is an important source of inequality and social mobility, and that large welfare gains can be obtained by large-scale government policies that target young children directly.

The model has two main building blocks. The first is that parental choices are important to a child's subsequent outcomes. An individual's education choice (college/no-college) and earnings depend on her assets, skills, and her taste for education. The key element here is that the level of these skills is determined by (money and time) investments made by her parents during her early childhood. College can be financed either with parental transfers (which are endogenous) or through working and borrowing. The second building block is the GE life-cycle Aiyagari framework in which these investments and intergenerational linkages are embedded. This GE framework allows aggregate education and skills to affect prices. It also includes endogenous labor supply which is important for the financing of policies to have distortionary effects. Both building blocks are important to the welfare evaluation of large-scale policies that target children.

The model is estimated using simulated method of moments to match evidence from the US in the 2000s. In addition to matching standard moments (e.g., average hours worked and the share of college graduates), we target ones that are informative about parental investments. The latter, along with the child's current skills and parental skills, are inputs into the child's future skills as in Cunha, Heckman, and Schennach (2010). Our model requires us, moreover, to specify explicitly how time and money aggregate to form "parental investments." We do this via a CES aggregator and estimate the parameters of this function by matching the average amount of "quality" time parents spend with their children, the average expenditures on child care and education, and the correlation between time and expenditures.

We use the evidence from a randomized control trial (RCT) to test the validity of the model's predictions on the effects of government investments towards early childhood development. García, Heckman, Leaf, and Prados (2017) study, using an RCT framework, two programs in which a small group

[1] Early childhood refers to the period when children are under 4 or 5 years old. The literature on the importance of this stage is large. Thompson and Nelson (2001) summarize the research in developmental neuroscience that is pertinent to early brain development while Elango, García, Heckman, and Hojman (2015) summarize the economics literature on the importance of early childhood environment.

of disadvantaged children were brought to high-quality early childhood development centers in North Carolina. The program's cost was approximately $13,500 per child-year. An equivalent program in the model implies introducing government expenditures towards early childhood development of $13,500 per child-year, but with three specific characteristics to be comparable to the RCT. First, the RCT focused on a small group of children so prices and taxes in the economy would not be affected. Second, the experiment focused on children of low-educated and low-income parents, so in the model we focus on disadvantaged children of high-school educated parents whose income and wealth are among the bottom 20%. Finally, the RCT involved only one generation of children so we do the same in the model. We find that children's college graduation rate and future labor income in the model increase by similar amounts to those found by García, Heckman, Leaf, and Prados (2017).

We then evaluate a *universal* version of this childhood investment program taking into account the distortionary taxation costs and GE effects. Welfare gains, computed for newborns under the veil of ignorance, are 10% in consumption equivalence units.[2] Moreover, the childhood investments program is associated with an income inequality reduction of 7% and an increase in intergenerational mobility of 34%, approximately enough for the US to achieve Canadian or Australian levels of inequality and mobility. It is important to note that the welfare gains achieved by the early childhood program are twice as large as if the same resources were used to fund a lump-sum transfer.

Comparing to the case in which the program is permanently implemented, we find that if such government investments were introduced for only one generation and in a small-scale they would achieve less than one-half of its benefits on welfare gains. We interpret this as evidence that a randomized control trial is likely to underestimate the long-run benefits of such a policy. General equilibrium forces—by lowering the wage of college graduates and therefore the return to those investments—and raising taxes to finance the additional government expenditures reduce welfare gains by one-tenth each. At the same time, the long-run change in the distribution of parental characteristics generate over two-thirds of the gains—more than compensating for the GE and taxation effects. The key mechanism is that investing in a child today not only increases that child's education and income, but also creates a better parent (and hence better inputs in the skill formation technology) for the following generation.

Even though benefits take time to accrue, our evaluation of transition dynamics shows that if the policy was implemented permanently, every new generation would be better off and over three-fourths of the long-run welfare gains would be achieved after only one generation. Older generations alive at the time the policy is introduced, however, are not better off. These cohorts are paying higher taxes to finance the initial costs of the program but are receiving gains only indirectly through their children, which results in net losses of welfare for them on average. A mechanism, such as government borrowing, that manages to transfer the cost to the future generations can reduce the losses for the older generations. We study this form of fiscal adjustment in Appendix D.

[2]Our main analysis focuses on a universal policy that invests the same amount as the early childhood RCT in North Carolina. Moreover, long-run welfare gains with this level of expenditures are close to the maximum that is achievable using alternative levels of resources for this policy.

Why do government investments in childhood development increase welfare? While several factors play a role, the main channel for welfare improvement lies in the government's capacity to make up for a parent's inability to borrow against her child's future income created by her parental investments.[3] To illustrate this channel, consider a poor parent who, by investing in the early childhood development of her kid, would raise a high-skilled, high-income child. The parent would then want to smooth consumption *intergenerationally*. The fact that this investment must come at the cost of her own lifetime consumption reduces her incentive to invest. If the child could promise to compensate her parent in the future and parents could borrow against this future income, this problem would be avoided. Government investments in early childhood can be thought of as (imperfectly) replacing the missing compensation-borrowing mechanism via the power of taxation. The government invests directly in children and taxes them once they are adults.

The rest of the paper is organized as follows. Section 2 discusses the literature. Section 3 introduces the model. Section 4 presents our empirical findings on parental investments and returns to skills by education groups. Section 5 explains the model's estimation and validation exercises. The policy analysis exercises are presented in Section 6. Finally, Section 7 concludes. The Appendix contains additional details and policy counterfactuals. These include parenting education programs, college subsidies, and optimal tax progressivity—and how this changes when early childhood development is endogenous, relative to the standard case in which it is exogenously fixed and, hence, policy invariant.

2 Related Literature

Macroeconomic and policy analysis of inequality can be divided in two strands. One subset of the literature focuses on the top 1%, with a particular interest in wealth and bequest taxation (e.g., Piketty and Saez, 2003; Diamond and Saez, 2011; Saez, 2016). The other one focuses on the bottom 99%, typically looking at the role of skills and education (e.g., Katz and Murphy, 1992; Autor, Katz, and Kearney, 2008; Abbott, Gallipoli, Meghir, and Violante, Forthcoming). In addition to income taxation, some of these papers also study college-education policies (e.g., Bénabou, 2002; Abbott, Gallipoli, Meghir, and Violante, Forthcoming; Holter, 2015; Krueger and Ludwig, 2016).[4] Among quantitative analyses of inequality, the standard model is based on Aiyagari-style life-cycle models, focusing usually on adult income shocks and abstracting from endogenous initial conditions (e.g., Keane and Wolpin, 1997; Huggett, Ventura, and Yaron, 2011). We also use a standard macroeconomic Aiyagari-style life-cycle model but we introduce new intergenerational linkages that allow us to endogenize those initial conditions and evaluate policies that target young children. The closest model to ours is probably Abbott, Gallipoli, Meghir, and

[3]In addition, life-cycle borrowing constraints as well as uncertain returns to investments (together with risk-averse agents and lack of insurance) can inefficiently reduce parental investments. We use the model to provide an estimate of the role of each of these sources of inefficiency in determining the welfare gain. We find that the introducing a form of intergenerational borrowing (i.e., implemented as a compensation system) leads to the largest gains.

[4]Holter (2015) evaluates the importance of education stages before college as well. He, however, focuses on cross-country partial-equilibrium comparisons of government education policies after the early childhood stage.

Violante (Forthcoming), which studies optimal college borrowing and grants. Our borrowing and grants structure is not as flexible as theirs, but we introduce endogenous parental investments in the formation of skills. The dynamic interactions between borrowing constraints and parental investments in child development may be important since limited assets and borrowing can limit the capacity of parents to invest money towards their children, which, due to complementarities, may also reduce their incentives to invest time towards them. This affects the income and wealth of the next generation, which once again shapes their capacity and incentives to invest in their own children.

Previous literature on childhood development estimates the production function of children's skills (e.g., Todd and Wolpin, 2003; Cunha, Heckman, and Schennach, 2010; Del Boca, Flinn, and Wiswall, 2014). We use the skills production function (and estimates) of Cunha, Heckman, and Schennach (2010) in our model but, unlike them, we model explicitly how investments are chosen by parents. This is necessary to study how policies affect parental investment choices and welfare in an equilibrium framework. Previous papers have modeled parental investments (e.g., Del Boca, Flinn, and Wiswall, 2014; Abbott, 2016; Caucutt and Lochner, 2017), but have abstracted from general equilibrium forces (and saving decisions in the case of Del Boca, Flinn, and Wiswall) which limits their capacity to evaluate large-scale policies.[5] Cunha, Heckman, and Schennach highlight two properties regarding childhood development: *dynamic complementarity* (i.e., skills produced at one stage raise the productivity of investment at subsequent stages) and *self-productivity* (i.e., skills produced at one stage augment skills attained at later stages). Our model incorporates these properties and connects them with inequality and social mobility in an environment suitable for policy analysis.

Including parental investments in a quantitative Aiyagari-style life-cycle model allows us to evaluate large-scale policies that directly focus on childhood development—which may reduce inequality and promote intergenerational mobility. Previous theoretical papers have highlighted that an environment with intergenerational investments in skills can lead to inefficient investment in children. Loury (1981) and Baland and Robinson (2000) use partial-equilibrium models to show that borrowing and parental transfers constraints (i.e., parents cannot borrow against their children's future income) can lead to inefficiently low levels of investments, which the government can improve on by enforcing higher investments towards children. Aiyagari, Greenwood, and Seshadri (2002) shows that general equilibrium effects imply that a world with borrowing and parental transfers constraints may lead to higher parental investments than an economy with complete markets due to the effect on aggregate wealth and interest rates—though it is still the case that inefficiency in investments arises with incomplete markets. The interaction between imperfect capital markets and human capital investments has also been explored in a growth context (e.g., Galor and Zeira, 1993; Galor and Moav, 2004). We contribute to this literature by providing a model that is suitable to quantitatively evaluate the effect of government investments towards children, in an economy that takes into account uncertainty in the returns to investments, gen-

[5] Restuccia and Urrutia (2004) extend an overlapping-generations model to incorporate intergenerational investments in human capital. They do not focus, however, on early childhood development and endogenous labor supply (important for the cost of raising taxes to finance policies).

eral equilibrium effects (both through the interest rate and the wage of college educated workers), and the distortionary impact of the tax changes needed to finance policies.

To the best of our knowledge, there are only two papers that introduce early childhood development in a quantitative general equilibrium model with heterogeneous agents. First, Yum (2018) incorporates parental time investments into a general equilibrium model but, differently from his work, our model explicitly takes into account the findings of the empirical literature which highlights the need for multiple periods of parental investments. Moreover, we also include monetary inputs in the formation of skills as well as flexible parental transfers, and allow for borrowing as observed in the data—which is important to study policies that may be affected by borrowing constraints.[6] Second, Lee and Seshadri (Forthcoming) also study parental investments in a general equilibrium model. They focus, however, on reallocating subsidies across different development stages, while we explore alternative levels of government investments. Furthermore, we study the transition dynamics, which is the key exercise to show the importance for welfare gains of taking into account intergenerational dynamics (i.e., investing in a child today not only increases that child's education and income, but also creates a better parent for the following generation). Given the importance of this mechanism, our transition evaluation shows the relatively fast pace at which this policy can be expected to yield its large returns (i.e., most gains are accrued after only one generation).[7]

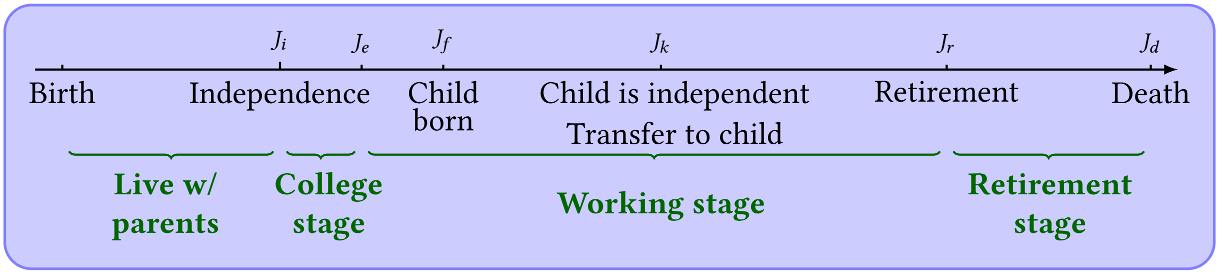
3 Model

The model has two main building blocks. The first is that parental choices are important to a child's subsequent outcomes. An individual's education choice (college/no-college) and earnings depend on her assets, skills, and her taste for education. Although all these are endogenously related to parental choices, the key element here is that skills are determined by parental investments (money and time) during her early childhood. The second building block is the GE life-cycle Aiyagari framework in which these investments and intergenerational linkages are embedded. This framework includes wage uncertainty and incomplete markets. Given our interest in studying costly policies, we want to take into account the cost of raising tax revenues, so we include endogenous labor supply and distortive taxation. Given the evidence that early childhood policies increase college graduation, general equilibrium is important to study the potential effect on the wage of college graduates. A representative firm combines the different types of labor (by education) and capital to produce the final consumption good. Finally, the government levies taxes on consumption, labor, and capital in order to finance some fixed exogenous expenses as well as provide a lump-sum transfer and retirement benefits.

[6] The set of policies is also different. In particular, including monetary investments allows us to interpret government investments in early childhood within the baseline framework of the model.

[7] Another difference with these papers is that we include both cognitive and non-cognitive skills which Cunha, Heckman, and Schennach (2010) highlight to be important for the estimation of the elasticity of substitution of the skill production function. Although our results are qualitatively similar in a model with only cognitive skills and one with both types of skills, we found that large differences in the magnitude of the effects emerge.

Figure 1: Life Cycle



3.1 The individual problem

There is a dynastic framework with four main stages (20 periods total): childhood, college, labor, and retirement. Figure 1 shows the life cycle of an agent, in which each period in the model refers to four years. Let j denote the age in periods (e.g., $j = 1$ refers to ages 0–3 while $j = 2$ to ages 4–7). From $j = 1$ until $j = J_i$ the child lives with her parents. At age $j = J_i$, individuals become independent (i.e., start making choices) after finishing high school with a level of skills, that depends on their parents' investments, as well as an amount of assets, also decided by their parents. Idiosyncratic uninsurable risk makes labor income stochastic. Individuals first choice is between going to college or remaining a high-school graduate. Once agents exit the education phase, they enter the third stage, which represents their labor market experience. Throughout their lives, agents choose their labor supply, savings, and consumption expenditures. They can borrow up to a limit, and save through a non-state-contingent asset. The framework is one of uncertainty both in earnings as well as child's skill development: Individuals choose how much time and money to invest in their child's development, but the final outcome is uncertain. Before the child is of college age, the agent decides the amount of monetary resources to transfer to them. The last stage is retirement. At this time, agents have two sources of income: savings and retirement benefits. Before going into more detail about these stages, we explain some elements that will make the rest of the model clearer.

Credit market Agents can only trade risk-free bonds, but interest rates are different for saving and borrowing. Agents with positive savings receive interest rate equal to r, while those borrowing pay interest rate equal to $r^- = r + \iota$, where $\iota \geq 0$. The wedge between interest rates is important to capture the cost of borrowing, which is a form of insurance relevant for the quantitative analysis. Individuals face borrowing limits that vary over the life-cycle. Student loans are explained in detail below. Young workers (i.e., under the age of 20) and retired households cannot borrow. Let $e \in \{1, 2\}$ be the level of education of the agent, which stands for high-school and college graduate, respectively. Workers with access to borrowing (i.e., after age 20) are subject to credit limit of $\underline{a}(e)$. Estimates of $\underline{a}(e)$ are based on self-reported limits on unsecured credit from the Survey of Consumer Finances.

Wage process The wage of individual i with education e at age j is given by $w^e \varepsilon_j^e(\theta, \eta)$ where w^e is the wage of education group e, $\varepsilon_j^e(\theta, \eta)$ is the efficiency units, and θ are the skills. $\varepsilon_j^e(\theta, \eta)$ includes the age profile for the education group, the returns to skills θ, and the idiosyncratic labor productivity given by η which evolves stochastically following $\Gamma_{e,j}(\eta)$. Notice that we allow for education- and age-dependent idiosyncratic shocks. The parametrization and estimation details are presented in Section 4.

Preferences The agent is risk averse and her preferences are represented by $u(c, h)$ which is increasing and concave in consumption c and decreasing and concave in hours worked h. The future is discounted by β. We model altruism à la Barro and Becker (1989), in which the agent cares about the utility of her child (i.e., this is *not* a warm-glow model).

3.1.1 Education stage

At $j = J_i$ (16 years old in the estimation), the agent has the option to go to college (for one period, until $j = J_e$). The individual state variables are savings a, skill θ, and school taste ϕ. The agent chooses whether to go to college or start working. All agents become independent as high-school graduates ($e = 1$). If an agent chooses to go to college, her education changes to $e = 2$. The education decision is irreversible. The monetary cost of education is p_e, but, as is common in the literature (e.g., Heckman, Lochner, and Todd, 2006; Abbott, Gallipoli, Meghir, and Violante, Forthcoming), we also allow for school taste ϕ to affect the total cost of education. Modeling school taste is necessary because the observed cross-section variation in resources available to finance schooling and in returns to education can only partially account for the variation in education patterns (e.g., the intergenerational persistence of education). Particularly, we assume that the school taste ϕ enters as a separate term in the value function, and we allow its value to be correlated with the education of the parents. After leaving school, ϕ is assumed not to affect any adult outcome.

College students face borrowing limits $\underline{a}^s$ for subsidized loans. They have access to subsidized loans at rate $r^s = r + \iota^s$ where $\iota^s < \iota$. To simplify computation, we assume that college student debt is refinanced into a single bond that carries interest rate r^-. $\tilde{a}^s(a')$ is the function performing this transformation. When making this calculation we assume that fixed payments would have been made for 5 periods (i.e., 20 years) following graduation.[8] Borrowing limit $\underline{a}^s$ and wedge ι^s are based on the rules for federal college loans, to be explained in detail in Section 5.

College students are allowed to work—providing high-school level labor—but we also allow for studying to take time $\overline{h}$. This reduces the number of hours worked by students in the model and is important for the quantitative analysis since otherwise too many students would work full-time while in college, hence reducing the importance of parental transfers or borrowing to finance education.

[8]Given the fixed payment nature of student loans and the assumption that they are repaid in 5 periods, we can transform college loans into regular bonds using the following formula: $\tilde{a}^s(a') = a' \times \frac{r^s}{1-(1+r^s)^{-5}} \times \frac{1-(1+r^-)^{-5}}{r^-}$.

Formally, let $V_j^s(a, \theta, e = 2, \phi)$ show the value of an agent of age j in college and with assets a, skills θ, and school taste ϕ.[9] It is defined by

$$V_j^s(a, \theta, e = 2, \phi) = \max_{c,a',h} u\left(c, h + \overline{h}\right) - \epsilon(\phi, \theta) + \beta E_{\eta|e} V_{j+1}^w(\tilde{a}^s(a'), \theta, e, \eta) \tag{1}$$
$$c + a' + p_e - y + T(y, a, c) = a(1 + r)$$
$$y = w^1 \varepsilon_j^e(\theta) h, \quad a' \geq \underline{a}^s, \quad 0 \leq h \leq 1 - \overline{h}, \quad \eta \sim \Gamma_{e,0}.$$

She can borrow up to the limit $\underline{a}^s$, and the return on positive savings is $1 + r$. However, if the agent decides to borrow she pays interest rates $r^s > r$. We denote as w^e the wage for an agent who is currently in school at level e. The disutility of working is included in u, while the disutility of going to school is in $\epsilon(\phi, \theta)$ which depends both on school taste and skills.

$V_j^w(a, \theta, e, \eta)$ is the value of work for an agent of age j with assets a, skills θ, education e, and stochastic labor efficiency η. It is defined by

$$V_j(a, \theta, e, \eta) = \max_{c,a',h} u(c, h) + \beta \mathbb{E} V_{j+1}(a', \theta, e, \eta'), \tag{2}$$
$$c + a' - y + T(y, a, c) = \begin{cases} a(1 + r) & \text{if } a \geq 0 \\ a(1 + r^-) & \text{if } a < 0 \end{cases}$$
$$y = w^e \varepsilon_j^e(\theta, \eta) h, \quad a' \geq \underline{a}_{e,j}, \quad 0 \leq h \leq 1, \quad \eta' \sim \Gamma_{e,j}(\eta).$$

The agent can borrow up to the limit $\underline{a}_{e,j}$, and the return on positive savings is $1 + r$. However, if the agent is borrowing she pays interest rates $r^- > r$. The return from working is the wage w^e scaled by $\varepsilon_j^e(\theta, \eta)$—a function of the worker's age, education, skills, and idiosyncratic labor productivity.

Then, V_j^{sw} is the value of an agent who can choose between working (as a high-school graduate) and going to college,

$$V_j^{sw}(a, \theta, \phi) = \max\left\{\mathbb{E}_{\eta|e} V_j(a, \theta, 1, \eta), V_j^s(a, \theta, 2, \phi)\right\}.$$

3.1.2 Working stage

From $j = J_e$ until $j = J_r$ (64 in the estimation), the agent works and her individual problem is equivalent to (2). However, the problem changes when the agent's child is born at the exogenously given fertility period $j = J_f$ (28). We assume that each agent has one child—or, alternatively, each household has one household offspring. For two periods the agent has to choose the number of hours τ and amount of money m to invest in the child's development of skills. Moreover, once the child become independent (at $j = J_k$, or 44 in the estimation), the agent chooses the size of the parent-to-child transfer φ.

[9] We assume that the initial draw of η takes place after going to school, thus the agents efficiency units ε are assumed to have a value of η at the mean, i.e., zero.

Investments in child's skills This is where the key novelties of the model are present. For two periods from $j = J_f$, the agent works and also invests directly in her child's development of skills θ_k. Hence, θ_k is added as a state variable at this stage. The initial distribution of θ_k is stochastic but is allowed to depend on parent's skills θ. In addition to standard choices of consumption, savings and labor supply, the agent now also chooses how much time τ and money m to invest in the child's development. Then, the skill development function—which consists of two nested CES functions—determines how these investments determine the evolution of θ_k. The outer CES is based on Cunha, Heckman, and Schennach (2010) but, differently from them, we model parental investments explicitly (to incorporate τ and m) in the inner CES.

Next-period child's skill θ_k' depends current child's skill θ_k, parents' skills θ, and parental investments I—as well as an idiosyncratic shock ν. The inner CES function shapes parental investments using both time τ and expenditures m towards the child.[10] In the main policy analysis we will assume that government investments into early childhood development and parental money investments are perfect substitutes—See equation (7). Thus, these government investments will crowd out parental money investments m (until $m = 0$). The reaction of time investments τ, however, will depend on how substitutable/complementary money and time are (i.e., will depend on γ).[11]

$$
\begin{aligned}
V_j\left(a,\theta,e,\eta,\theta_k\right) &= \max_{c,a',h,\tau,m} u\left(c,h\right) - v\left(\tau\right) + \beta \mathbb{E} V_{j+1}\left(a',\theta,e,\eta',\theta_k'\right), \\
& c + a' + m - y + T\left(y,a,c\right) = \begin{cases} a\left(1+r\right) & \text{if } a \geq 0 \\ a\left(1+r^-\right) & \text{if } a < 0 \end{cases} \\
& y = w^e \varepsilon_j^e\left(\theta,\eta\right) h, \quad a' \geq \underline{a}_{e,j}, \quad 0 \leq h+\tau \leq 1, \quad \eta' \sim \Gamma_{e,j}(\eta) \\
& m \in \left\{m_1, m_2, \ldots\right\}, \quad \tau \in \left\{\tau_1, \tau_2, \ldots\right\} \\
& \theta_k' = \left[\alpha_{1j}\theta_k^{\rho_j} + \alpha_{2j}\theta^{\rho_j} + \alpha_{3j} I^{\rho_j}\right]^{1/\rho_j} e^{\nu}, \quad \nu \sim N(0,\sigma_{j,\nu}) \\
& I = \bar{A}\left[\alpha_m m^{\gamma} + (1-\alpha_m)\tau^{\gamma}\right]^{1/\gamma}.
\end{aligned}
\tag{3}
$$

After two periods, child's skills are fixed, so the problem is equivalent to (2) but with the extra state variable θ_k.[12]

[10] The choice of time and money is made within a discrete set of possible alternatives for computational reasons. When solving the model we limit the number of options for time and money to 6 each, i.e. 36 total alternatives. We assume that time τ enters in a separable manner in the utility function because the cross-sectional data suggests that individuals that spend more time with their children reduce leisure time instead of hours worked.

[11] We discuss the estimation of this parameter in Section 5 but, as a preview, we find that the evidence tends to suggest that time and money are imperfect complements.

[12] This assumption simplifies the solution but is also in line with the evidence on early childhood development. The literature tends to find that skills are significantly less malleable for older children. Similarly, the CDS data shows that while children under the age of 8 tend to transition between skill's ranks, this is not much the case for children above that age. In particular, if we assign children to terciles of the skills distribution and look at the transition matrix between those terciles over time, we find that the trace of the transition matrix grows from under 1 for children between 2 and 8 years old to over 1.6 for children between 8 and 12 years old.

Child becomes independent Just before reaching age $j = J_k$ (i.e., when the child is of age $j = J_i$) the agent needs to decide the size of monetary transfers φ to make to her child. We model this as a sub-period that takes place just before the child becomes independent, with a value for the agent defined by V_{Transfer}. Importantly, the transfer needs to be non-negative—i.e., the parent cannot leave debt to her child nor borrow against her future income. When making this choice, the parent already knows the realization of his income shock η, but is not aware of the school taste draw ϕ_k of her child.[13]

$$V_{\text{Transfer}}\left(a, \theta, e, \eta, \theta_k\right) = \max_{\varphi} V_{J_k}\left(a - \varphi, \theta, e, \eta'\right) + b\mathbb{E}V_{J_i}\left(\varphi, \theta_k, \phi_k\right), \tag{4}$$
$$\varphi \geq 0, \quad \phi_k \sim N(\bar{\phi}_e, \sigma_\phi)$$

Notice that unlike (3), the value function at this stage now includes the continuation value of the child V_{J_i}. This is the last period in which parents' choices affect their descendants. As the problem is written recursively, this implies that at every period in which parents' choices affect children's outcomes—i.e., all previous periods—the utility of their descendants is taken into account. This formulation embeds the parental altruism motive. The school taste of the child is stochastic but correlated with the parent's level of education—which is useful to match the intergenerational persistence of education. Moreover, recall that the effective school distaste from equation (1) may also depend on the skill level of the child. The functional form of the stochastic processes of skills and school taste are specified in Section 5. After the agent's child becomes independent, the individual problem is equal to (2).

3.1.3 Retirement stage

At $j = J_r$, the agent retires with two sources of income: savings and retirement benefits. For simplicity, retirement benefits are assumed to depend on the agent's education and skill level. Agents are assumed to provide no work at this stage, so $l = 0$. Unsecured borrowing is not allowed at this stage either. Formally, the problem at the age of retirement is

$$\begin{aligned} V_j\left(a, \theta, e\right) &= \max_{c, a'} u\left(c, 0\right) + \beta V^w_{j+1}\left(a', \theta, e\right), \\ & c + a' = \theta + \pi(\theta, e) + a\left(1 + r\right), \\ & a' \geq 0, \end{aligned} \tag{5}$$

where π are the retirement benefits, which depend on the education and skill level.[14]

[13] The assumption that the school taste is not perfectly known to parents helps make the problem smoother which is useful for computational reasons.

[14] We use education together with the skill level, as a proxy to approximate average lifetime income with which the retirement benefits are determined. See Section 5 for details.

3.2 Aggregate production function

We assume there is a representative firm with production technology $Y = K^{\alpha} H^{1-\alpha}$, where K is aggregate physical capital and H is a CES aggregator of the labor supply of the two education groups

$$H = \left[s H_1^{\Omega} + (1 - s) H_2^{\Omega} \right]^{1/\Omega}.$$

Capital depreciates at rate δ per period.

3.3 Definition of Stationary Equilibrium

The model includes J_d overlapping generations and is solved numerically to characterize the stationary equilibrium allocation. Stationarity implies that we study an equilibrium in which the cross-sectional distribution for any given cohort of age j is invariant over time periods. Particularly important is that the distribution of initial states is determined by the choices of the older generations. The equilibrium allocation requires that households choose education, consumption, labor supply, parental time and expenditure investments, and parental transfers such that they maximize their expected utility; firms maximize profits; and prices (wages of each education group and the interest rate) clear markets.[15] See Appendix B for details.

3.4 Role for Government

Why do government investments in childhood development increase welfare? While several factors play a role, the main channel for welfare improvement lies in the government's capacity to make up for a parent's inability to borrow against her child's future income created by her parental investments.

To illustrate this, consider a parent who is poor but invests enough to raise a high-skilled, high-income child. The parent would then want to smooth consumption *intergenerationally*. The fact that this investment must come at the cost of her own lifetime consumption reduces her incentive to invest. Suppose, for now, that individuals are free to borrow against their own future income. If the child could promise to compensate her parent in the future, the parent would not need to reduce her consumption and the problem would be avoided. This example shows that imperfect parental altruism ($b < 1$) is not the direct source of underinvestment. Even if parents were perfectly altruistic ($b = 1$), they may want to be compensated by (or borrow against) their children—particularly if they expect their children to be better off than themselves as in the example mentioned.[16] Lack of compensation is not the only reason

[15] The government is allowed to have other expenses which are wasteful since the estimation is designed such that the income redistribution is matched—rather than the government's budget being cleared. However, whenever we introduce a policy in Section 6 this is done such that these extra government expenses remain fixed.

[16] Even though altruism b is not the direct source of underinvestment, transfer constraints are more likely to bind if b is

for government intervention. Even if the child could compensate her parent by transferring resources, the timing of those transfers matters. If the compensation takes place once the child is past the development stage, borrowing constraints can prevent the parent from using the transferred money at the time the parental investments take place. Government investments in early childhood can be thought of as (imperfectly) replacing the missing compensation-borrowing mechanism via the power of taxation. Rather than children compensating parents for their investments, the government invests directly in children and taxes them once they are adults.

Borrowing constraints can also reduce investments if the parent is poor today but expects to be richer in the future. The parent would like to use part of her future income to invest in the child's development. Borrowing constraints, however, may prevent that allocation. Finally, in addition to transfers and borrowing constraints, our model is also one with uncertain returns to investments and lack of insurance, which can also lead to reduced parental investments since this uncertainty creates an extra incentive for agents to consume and invest in the safe asset rather than in children.

There is not a perfect way to evaluate the importance of each channel; but we use the estimated model to provide some information. Although these results depend on the model estimation, which is discussed in Section 5, we believe it is clearer to discuss this decomposition here. We evaluate the welfare gains, using consumption equivalence for newborns under the veil of ignorance, achieved by shutting down each friction independently. To do this we first introduce a small (i.e., prices are not affected) multi-generational family in our economy that is "special," in the sense that it is not (and, importantly, has never been) subject to one of these three sources of reduced investments.[17] We then estimate welfare gains by looking at how much extra consumption an agent would need to be indifferent between being born to a "normal" family rather than to this special family.

To capture the problem of lack of compensation across generations we would like to introduce a new market in which parents and children mutually decide their investments. This is beyond the scope of this paper so, instead, we focus on a limited form of compensation. We introduce a transfer system in which the government "taxes" high-skilled individuals and uses that same money to pay a reward to parents with a high-skilled child. Thus, instead of compensating for parental investments directly, this framework rewards skills which are an outcome of those investments. We evaluate different levels of taxes/rewards and pick the one that generates the largest welfare gains, i.e., of 9.4%. In comparison, reducing uncertainty leads to welfare gains of 3.8% while enlarging borrowing limits generates gains of 0.8%.[18] Thus, these suggests that lack of compensation (associated also with constraints in borrowing

low.

[17]It is important to allow for the dynamic effect on distributions to take place when removing closing each channel. This is why we highlight that this "special" family has never been subject to the source of the problem.

[18]Notice that this compensation mechanism implies that parents receive the money when the child is young but the child only pays once she is an adult. This is important for the results: if the child has to pay before accumulating savings and the parent receives the money only after the child is past the development stage, welfare gains are much smaller. In that case borrowing constraints limit the amount of money the child can use for the reward as well as the amount of money the parent can borrow at the time the parental investments take place. In this borrowing limits exercise we made the limits 5 times larger than in the baseline.

against future generations' income) is the leading source of underinvestment, followed by uncertainty and borrowing constraints.[19,20]

Loury (1981) provides a simpler partial-equilibrium model in which he can show that government investments towards childhood are welfare improving even when (lump-sum) taxes are needed pay for them.[21] In Section 6 we quantitatively evaluate such a policy in our richer model, which also takes into account distortive taxation and general equilibrium effects.

4 Data

This section presents evidence on parental time and money investments towards children. This evidence is certainly subject to endogeneity concerns, so we highlight that this evidence is only used to construct moments for the model's estimation and validation. The last part of this section also shows evidence that skills increase hourly wages.

4.1 Parental Investments

The Panel Study of Income Dynamics (PSID) survey follows a nationally representative sample of over 18,000 individuals living in the US. The Children Development Supplement (CDS) provides information on (a subsample of) children's test scores, time spent with parents, and parental expenditures. Among many assessments of child skills, two are the most commonly used to evaluate children's early cognitive capacities: Letter-Word (LW) and Applied Problem Solving (AP). We use both and findings are robust to the choice of the test. The most interesting feature about the CDS is that it provides detailed time diaries for each child. We can observe hour by hour what activity is being performed (e.g., reading or playing), if the activity is being performed with someone (e.g., the father is reading with the child), and if someone is around while the child is doing the activity (e.g., the father is working while the child is reading next to him). The time diary is available for both a weekday and a weekend day. We obtain these time diaries and test scores for the years 2002 and 2007. For details on the sample selection, see Appendix A.1. Using

[19] Importantly, we find that these welfare gains are concentrated among children born to low-skilled parents. Children of high-skilled and college educated parents would not gain from these changes.

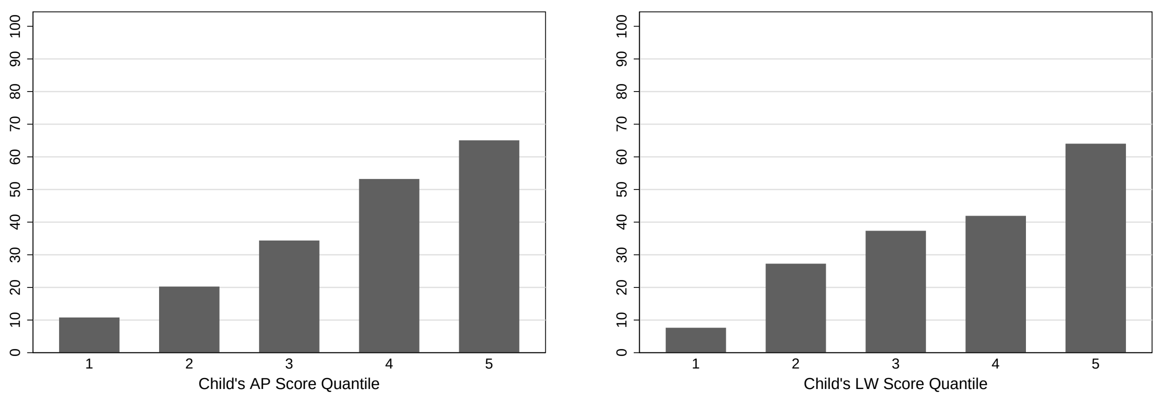
[20] There are alternative ways to try to measure the relevance of each channel. We also compared the investment choices this special families makes relative to a baseline family with similar characteristics regarding skills—results are similar if we compare families with similar income. We find that a low-skilled family with laxer borrowing constraints would invest between 20 and 30% more than a low-skilled family with the estimated borrowing constraints. In comparison, increasing borrowing limits for high-skilled families only increases their investments by less than 5%. On the other hand, eliminating the uncertainty in wages increases the investments of high-skilled families by 20% but has almost no effect on low-skilled families.

[21] Baland and Robinson (2000) also use a partial-equilibrium model of parental investments (though to study child labor) and highlight the theoretical role of two channels: potentially binding parental transfers constraints and potentially binding borrowing constraints. They find gains from increasing parental investments (reducing child labor in their interpretation) when either constraint binds.

this data, we first briefly highlight that skills are good predictors for college graduation and are related to parents' characteristics. Nevertheless, our main interest is in parental investments towards children's development. We show evidence that time with children and expenditures towards their development are associated with parents' education and income.

Children's test scores are good predictors of college graduation We first focus on children who were tested between the ages 11 and 13, and group children in quintiles according to their scores.[22] We then look at college graduation rates for these children. Importantly, individuals that were once children (and potentially included in the CDS) are then included in the main PSID survey. These allows us to connect young-age test scores from the CDS to college outcomes from the PSID. In order to reduce the concern that children may still be finishing their college education, we focus on children observed (at least once) in the main PSID sample after they are 24 years old. Figure 2 shows that young-age skills are positively correlated with later college-graduation rates. Children that were in the top quintile of the LW score distribution around age 12 are six times as likely to graduate college than children that were in the bottom quintile.

Figure 2: College graduation rate and young-age skills



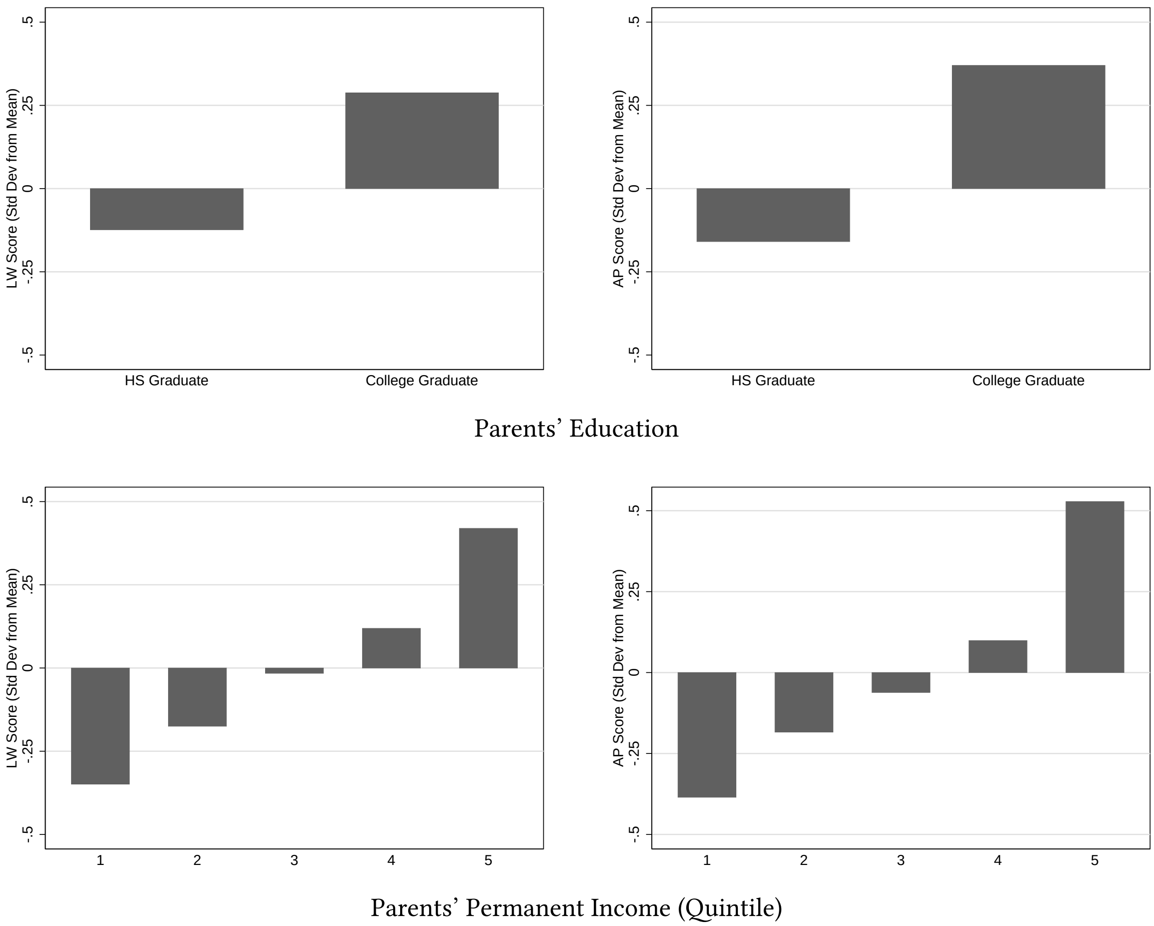
Source: CDS and PSID. We divide children into quintiles according to their Applied Problems (AP) and Letter Word (LW) test scores, respectively. For each quintile, we calculate the college graduation rate. Methodology is explained in the main text.

Children's test scores are associated with parents' characteristics Since these young-age skills are important for later outcomes, it is important to know where skill differences come from. Here we provide some evidence that children's skills are correlated with parents' socioeconomic group. Figure 3 reports average standardized scores (i.e., standard deviations below or above the mean) for children from parents with certain characteristics. Using the PSID data we are able to identify the education group of the parents as well as estimate their permanent income.[23] Test scores are positively correlated with parents' income and education, which is robust to the choice of the test. Children of college-educated par-

[22]Since age is an important determinant of test scores, we remove the age trend using a second-degree polynomial.
[23]Permanent income is estimated using all income parents had between while the child was under 16 years old.

ents are on average 0.5 standard deviations above children of high-school graduates. Similarly, children of high-income parents are approximately 0.7 standard deviations above those of low-income parents.

Figure 3: Children's skills and parents' characteristics



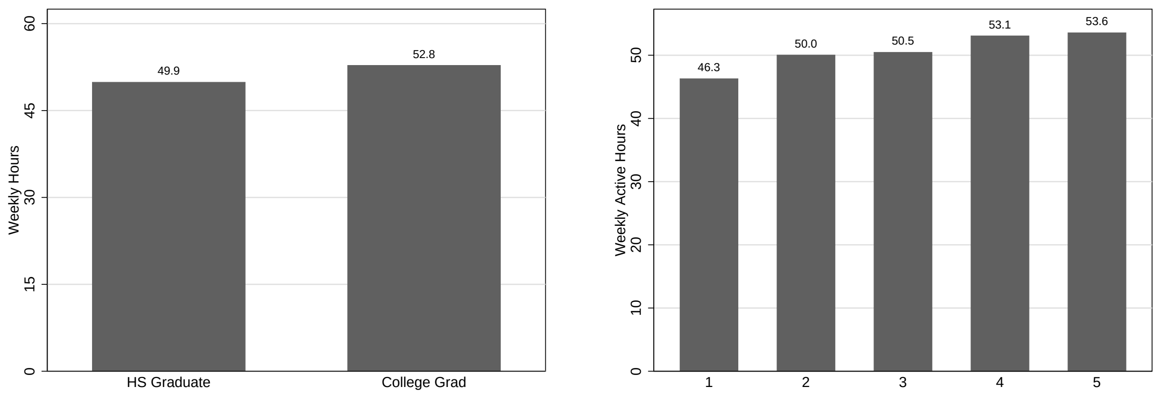
Source: CDS and PSID. We first remove the age component of the scores using a second-degree polynomial. Then, we standardize children's Applied Problems (AP) and Letter Word (LW) test scores. We compute average standardize scores for children born to families with different education (mother's education) or permanent income. Permanent income is computed using all income while the child was less than 16 years old.

Parental investments towards children's development are associated with parents' characteristics The main benefit of the CDS dataset is that we can observe detailed diaries on the time parents spend with their children. Using these time diaries, we define "time with parents" if the parent is doing the activity together with the child.[24] First, we add up all activities to estimate total active time with parents per week. Once again we remove the age trend and approximate the average for each child around age 4. Figure 4 suggests that that there are small differences in total time with children between

[24]If two parents are performing the activity, we interpret this as double the hours since time constraints must hold for the household.

parents with different education or income levels. However, this hides substantial heterogeneity in the kind of activities different groups of parents are doing with their children.

Figure 4: Total time with parents and parents' characteristics

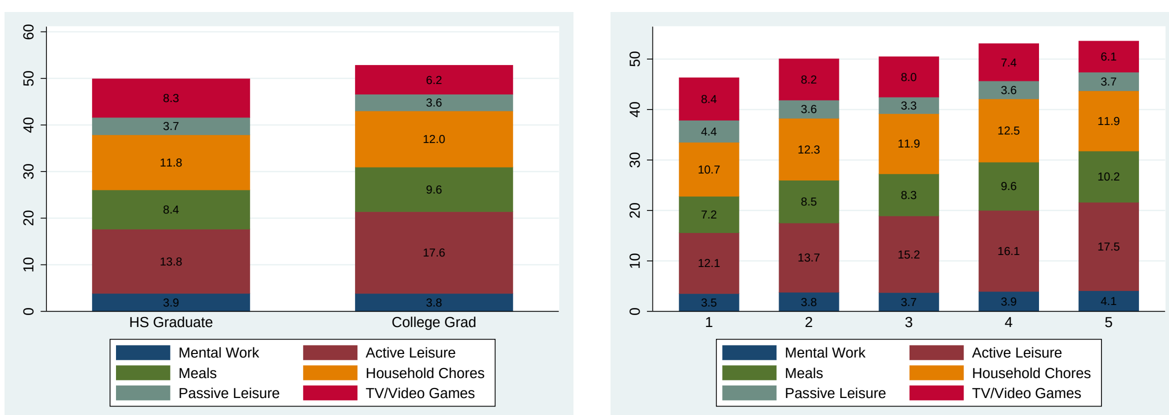


Parents' Education

Parents' Permanent Income

Source: CDS and PSID. We divide children according to the mother's education or parents' permanent income. For each group, we calculate the average total amount of time they spend with their parents. Methodology is explained in the main text.

Figure 5: Total time with parents by activity and parents' characteristics



Parents' Education

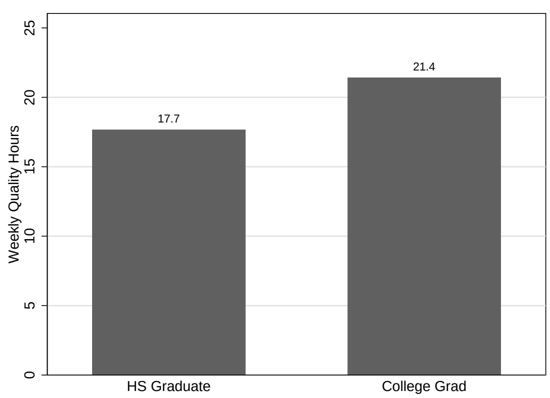
Parents' Permanent Income

Source: CDS and PSID. We divide children according to the mother's education or parents' permanent income. For each group, we calculate the average amount of time they spend with their parents doing each kind of activity. Methodology is explained in the main text.

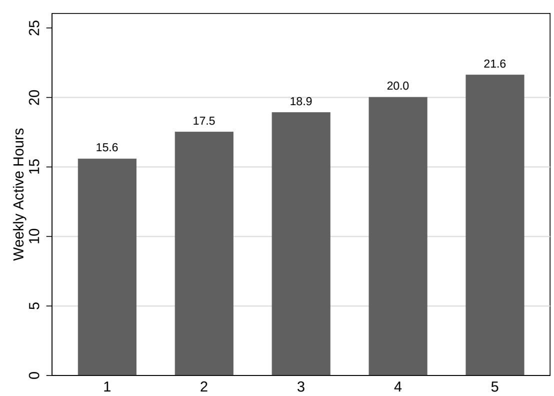
We group these detailed activities into six groups in order to facilitate the analysis. "Mental Work" refers to activities like reading, doing homework, or having conversations. "Active Leisure" groups activities like playing games or sports. "Meals" adds up the amount of time parents share meals with their children. "Household Chores" refers to basic activities like cooking, showering or dressing (in the case of young children, it is usually being showered or being dressed). "Passive Leisure" entails listening to the radio or laying around without doing a clear activity. Finally, "TV/Video Games" refers to times watching TV or playing video games. Figure 5 shows the distribution of these activities across parents from different socioeconomic groups. Even though lower-educated or low-income parents spend similar total amounts of time with their children as college-educated or high-income parents, a disproportionate larger share of that time is spent watching TV or playing video games, while a smaller share of time is allocated to active leisure or sharing meals.

These activities are different in the amount of interaction they entail between the parent and the child, and consequent skills development. Given the evidence that watching TV and playing video games is not typically associated with positive outcomes (Christakis, Zimmerman, DiGiuseppe, and McCarty, 2004; Swing, Gentile, Anderson, and Walsh, 2010), we exclude them from our definition of "quality time." We focus instead on playing and reading since these are suggested to be the most productive forms of interactions between parents and young children (Samuelsson and Carlsson, 2008).[25] Figure 6 shows that once we focus on quality time, parents from high socioeconomic groups spend more time with their children. College graduates spend 3.7 (21%) more hours per week than high-school graduates. Similarly, high-income parents spend 6 (38%) more quality hours per week with their children than low-income parents.

Figure 6: Total "quality" time with parents and parents' characteristics



Parents' Education



Parents' Permanent Income

Source: CDS and PSID. We divide children according to the mother's education or parents' permanent income. For each group, we calculate the average total amount of "quality" time they spend with their parents. "Quality" time includes activities like reading and playing games or sports. Methodology is explained in the main text.

These differences in the amount of quality time may help us explain the differences in skills developed by young children. Given typical concerns of endogeneity, the model in Section 3 provides a framework in which parents develop their children's skills and these skills are important for education and labor outcomes later in life. The technology for skill development used in the model is based on the estimates from Cunha, Heckman, and Schennach (2010), where the authors focus on identification and reducing

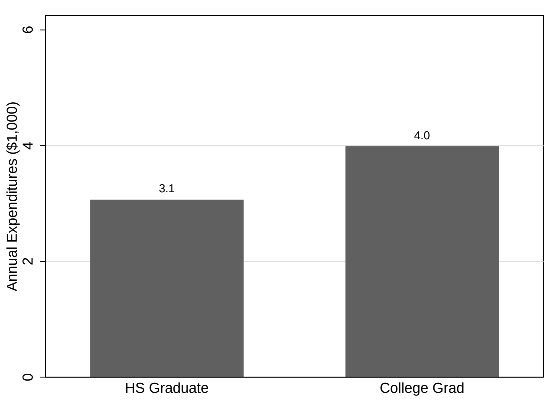
[25]The increased differences that emerge are robust to just removing time watching TV or playing video games.

concerns of endogeneity. Our model allows us to connect skill development with intergenerational mobility and inequality, and to analyze the dynamic interactions between parental investments and inequality and mobility.

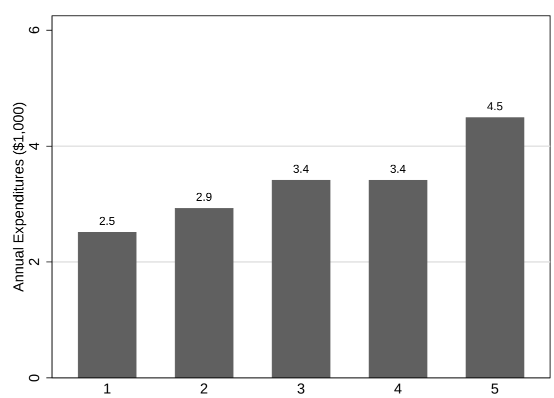
Monetary investments towards children Using PSID data it is also possible to obtain partial information on parental expenditures towards their children. We have information on expenditures towards children on toys, school supplies, clothes, food, medical, and vacations. Relevant expenditures on school fees (or costs included in the value of the house), extracurricular activities, museums, nannies, and others are not included, so this information should be read as a proxy for total expenditures, rather than as a perfect measure of them. We interpret them as more useful to compare different groups rather than to estimate average total expenditures. The results from this analysis are reported in Figure 7. High-income parents spend almost twice as much on their children as low-income families do. Similarly, college educated parents spend approximately 29% more than high-school educated parents.

As mentioned above, one missing element is the fact that in order to access good public education parents need to live in specific areas in the US, which may be associated with higher housing costs. Part of housing costs reflect the amenities that come with the local area, including the quality of schools available. In order to try to capture this public school quality expenditures, we calculate rent expenditures, imputing it for home owners. Given that higher income individuals might choose to live in bigger houses and all that is needed for a child to access public school is one room in the appropriate neighborhood, we divide the annual rental value of the house by the number of rooms. Figure 8 shows that high-income parents (annually) spend $1,600 (133%) more than low-income parents on rent per room. College-educated parents spend $800 (57%) more than high-school-educated parents.

Figure 7: Spending toward children and parents' characteristics



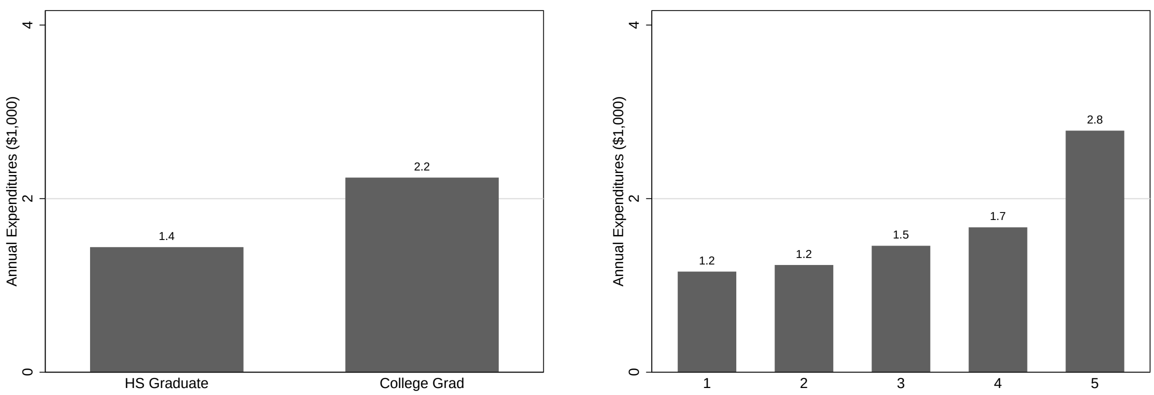
Parent's Education



Parent's Permanent Income

Source: CDS and PSID. We divide children according to the mother's education or parents' permanent income. For each group, we calculate the average amount of monetary expenditures towards children in the categories of toys, school supplies, clothes, food, medical, and vacations. Methodology is explained in the main text.

Figure 8: Rent cost per room and parents' characteristics



Parent's Education   Parent's Permanent Income

Source: CDS and PSID. We divide children according to the mother's education or parents' permanent income. For each group, we calculate the average amount of expenditures on rent (imputed rent for owners) per room. Methodology is explained in the main text.

Summary and other samples Table 1 presents summary evidence on parental investments in children. In the data, there is substantial heterogeneity in parents' characteristics. Given that the model used in this paper does not have fertility or marriage choices, we highlight that the moments used for the estimation are almost unchanged when we focus on a more homogeneous sample. The first column refers to the whole sample, as studied above. The other two columns focus on families for which the two parents live together while the child is under the age of 12. Moreover, the second column focuses on families with between 1 and 3 children, while the last column restricts the sample to cases with only two children.

Weekly hours with children is the moment we use in the estimation of the model, so it is important to note that even though married parents tend to spend more time with their children than non-married parents, the difference between sample averages is only 2 hours. Yearly expenditures (including rent) are not used in the estimation; another estimate that includes child care is used. Nevertheless, since that estimate does not control for potential single parenthood it is useful to note that differences between samples (across samples) are not too large. We interpret these findings as suggesting that the model estimation wouldn't change significantly if we made the sample selection more restrictive.

Table 1: Parental Investments: summary for alternative samples

	All	Parents Together 1–3 Children	Parents Together 2 Children
Sample Means			
Weekly Hours	19.1	21.5	21.1
	(0.3542)	(0.7090)	(0.7686)
Yearly Expenditures	5.0	5.6	6.2
	(0.1112)	(0.1808)	(0.2252)
Regression Coefficients			
Log(Hours) on College	0.189***	0.146**	0.132*
	(0.0401)	(0.0668)	(0.0743)
Log(Hours) on Log(Income)	0.178***	0.0104	0.0414
	(0.0333)	(0.0718)	(0.0815)
Log(Expenditures) on College	0.450***	0.221***	0.256***
	(0.0230)	(0.0586)	(0.0647)
Log(Expenditures) on Log(Income)	0.495***	0.514***	0.544***
	(0.0237)	(0.0464)	(0.0518)
Number of children	2,778	1,084	593

*Source: CDS and PSID. Robust standard errors in parentheses. *, **, *** denote statistical significance at the 10, 5, and 1 percent, respectively. Expenditures includes monetary expenditures towards children in the categories of toys, school supplies, clothes, food, medical, and vacations as well as the rent cost per room. Differences arise between these results and the figures above due to the cases with values zero—i.e., when the log is undefined—but differences are small. Methodology is explained in the main text.*

The bottom four rows summarize the cross-sectional evidence which is used in the validation of the model. Part of the heterogeneity of investments across children is reduced when we focus on more selective samples. For example, in the full sample children of college-educated parents tend to spend 19% more time with their parents than children of high-school educated parents, but if we focus on children of married parents this difference is reduced to between 13 and 15%. While the regression coefficient of time investments on parental income is insignificant for the married samples, the coefficient from regressing expenditures on income remains significant and stable around 0.5 for all samples.

4.2 Education and Returns to Skill

We evaluate the effect of skills on wages across different education groups. Abbott, Gallipoli, Meghir, and Violante (Forthcoming) show that skills increase hourly wages, and that this return is higher among college-educated individuals than among lower-educated groups. We contribute to this literature by highlighting that this pattern is robust to focusing on two-adult-households. For details on the sample selection, see Appendices A.2 and A.3.

We estimate the wage process for high-school and college graduates separately, allowing for differences across age and skills, to provide an estimate for the returns to skill. Heckman, Stixrud, and Urzua (2006) document that cognitive skills affect earnings around five times more than non-cognitive skills, so we make the simplifying assumption that only cognitive skills directly affect earnings in the labor market. The estimated model incorporates non-cognitive skills by assuming that they affect the (cognitive and non-cognitive) skills production function as well as education choices.

The wage process is an important element for the model since it determines the career profile—including the amount of uncertainty. We propose that the wage process of individual i with education e at age j is given by $w_t^e \varepsilon_{ij}^e$ where ε_{ij}^e are the efficiency units. These are defined by $\varepsilon_{ij}^e = \epsilon_j^e \psi_{ij}^e$ where ϵ_j^e is the age profile for the education group e and ψ_{ij}^e is the idiosyncratic labor productivity, which is specified as:

$$\begin{aligned} log\left(\psi_{ij}^e\right) &= \lambda^e log\left(\theta_i\right) + \eta_{ij}^e \\ \eta_{ij}^e &= \rho^e \eta_{ij-1}^e + z_{ij}^e, \quad z_{ij}^e \overset{iid}{\sim} N\left(0, \sigma_z^e\right). \end{aligned}$$

where θ_i is the level of cognitive skills and η_{ij}^e is the idiosyncratic shock. The initial value of productivity of an agent η_0^e is drawn from a normal distribution with mean zero and variance $\sigma_{\eta_0}^e$. The heterogeneity of the impact of skills on wages λ^e across education groups is particularly relevant for the education choices of agents with different abilities.

First, we use data from PSID to estimate the age profile ϵ_j^e as a second order polynomial in age, separately by education groups. Since the model has 4-years-long periods, we estimate this income process grouping observations over 4 years. We include year fixed effects to control for possible changes in average wages over time.[26] We use the PSID (instead of the National Longitudinal Survey of Youth, hereafter NLSY) because it includes a representative cross-section every year, so it avoids having the average age of the sample change directly with the calendar year.[27] Appendix Table A3 reports the results from this estimation, where the main finding is that age profiles are steeper for college-graduates than high-school graduates.

Second, we move to the NLSY to identify the effect of ability on wages. The NLSY is useful for this because it reports the Armed Forces Qualification Test (AFQT) score for these individuals—a typical measure of cognitive skills. For each household, we remove the appropriate age profile estimated from PSID, and estimate the returns to skill.[28] Consistent with previous evidence, Table 2 shows that returns to skill are higher for college graduates than high-school graduates.

[26] We include individuals from the over sample in our regression in order to improve the estimation power. We introduce a dummy to identify these individuals in our regression but results are robust to excluding them.

[27] Moreover, given the sampling methodology of the NLSY, it is still not possible to observe individuals over the age of 60. And, even if we were able to, it would be harder to distinguish age effects from year effects.

[28] In order to ease the interpretation of the results, we highlight that the standard deviation of log(AFQT) in the sample is approximately 0.05. Moreover, the average log(AFQT) is 5.19 and 5.38 for high-school and college graduates, respectively.

Table 2: Returns to skill by education group

	(1) High School	(2) College
log(AFQT)	0.533***	0.904***
	(0.0216)	(0.0456)
Observations	16,869	7,437
R-squared	0.104	0.147
# of households	3574	1479

*Source: NLSY. Robust standard errors in parentheses. *, **, *** denote statistical significance at the 10, 5, and 1 percent, respectively. log(AFQT) refers to the natural logarithm of the AFQT89 raw score. The regression includes dummies for over sample as well as head's gender and race. Methodology is explained in the main text.*

Finally, the residual after removing the age and skill component is used to estimate the process for the idiosyncratic shock η_j^e. We allow for measurement error and use a Minimum Distance Estimator, i.e., we use as moments the covariances of the wage residuals at different lags and age groups, separately for each education group. The results, reported in Table 3, suggest that shocks are persistent, particularly for higher educated groups.

Table 3: Income process estimation: idiosyncratic

	(1) High School	(2) College
ρ^e	0.893	0.945
σ_z^e	0.023	0.009
$\sigma_{\eta_0}^e$	0.044	0.054

Source: NLSY. A period is 4 years long. Methodology is explained in the main text.

5 Estimation

In this section we describe how we parametrize and estimate the model. The model is estimated using simulated method of moments to match standard moments as well as more novel ones (e.g., moments informative about parental investments) for the US in the 2000s. Some of the parameters can be estimated "externally," while others must be estimated "internally" from the simulation of the model. For these, we numerically solve the steady state of this economy, obtain the ergodic distribution of the economy, and calculate the moments of interest. Table 4 summarizes the parameters and moments used. After estimating the model, we validate the model using non-targeted moments as well as experimental evidence from an RCT that involved an early childhood program.

5.1 Preliminaries

Data and sample selection The model is estimated to match household level data so an agent in the model corresponds to a household with two adults in the data. Consequently, every household in the model has one household as offspring. We use three primary data sources: (i) the Panel Study of Income Dynamics (PSID); (ii) the Child Development Supplement (CDS) to PSID; and (iii) the 1979 cohort of the National Longitudinal Survey of Youth (NLSY79). We select a population for which our model can be taken as a reasonable approximation to household behavior and impose two main selection criteria on the data. First, as is standard in the literature (e.g., Huggett, Ventura, and Yaron, 2011), we drop household observations with income below a certain threshold. We choose this threshold as the one that corresponds to one person working 20 hours a week for the minimum wage (approximately \$8,000 total annual household income). There is no marriage decision in our model, so to avoid differences in income and time availability due to single parenthood, we keep only households with two adults. Details about sample selection are reported in Appendix.

Demographics A period in the model is four years. Individuals become independent at age J_i (equivalent to age 16), and they start with the equivalent of 12 years of education. They can go to college (one period), and so the maximum age for education is J_e (20). Parental time and money investment decisions are made at the time of (average first) birth J_f (28) and the period after. At age J_k (44), just before the agent's child becomes independent, she chooses the assets to transfer to her child. Retirement occurs at J_r (68). Death is assumed to occur for all agents at age J_d (80).

Prices Prices are normalized such that the average annual income of a high school graduate at age 48 is equal to one in the model. In the (PSID) data, this income is equal to $\$58,723$. The yearly price of college is estimated using the Delta Cost Project to be $\$6,588$.[29] All prices mentioned are in 2000 dollars.

Borrowing constraints Based on self-reported limits on unsecured credit by family from the Survey of Consumer Finances, we estimate the borrowing limits for working-age individuals $\underline{a}(e)$ to be $\{-24,000, -34,000\}$ for high-school and college graduates, respectively.

Taxes and replacement benefits The tax function is assumed to be $T(y, a, c) = \tau_y y + \tau_k a r \mathbf{1}_{a\geq 0} + \tau_c c - \omega$. Based on McDaniel (2007), we set $\tau_y = 0.22$, $\tau_k = 0.27$, and $\tau_c = 0.07$. The government's lump

[29] We take into account grants and scholarships, such that only private tuition costs are considered.

sum transfer to households ω is estimated to match the income redistribution observed in the data—as measured by the ratio of the variance of pre-tax income to after-tax income. Therefore, the government is allowed to spend some amount G, which is not valued by households and which is held constant in all our counterfactuals. The pension replacement rate is based on the Old Age, Survivors, and Disability Insurance federal program. We use education and skill level to estimate the average lifetime income on which the replacement benefit is based. See Appendix C.2 for details.

College loans College students have access to subsidized loans at rate $r^s = r + \iota^s$. According to the National Center for Education Statistics report "Student Financing of Undergraduate Education: 1999-2000," among the undergraduates who borrow, nearly all (97%) took out federal student loans, while only 13% took out non-federal loans. Moreover, the average loan value was similar for both federal and non-federal loans. Since average values were similar but federal loans were significantly more common, we focus on federal loans for our model estimation. Among federal loans, the Stafford loan program was the most common: 96% of undergraduates who borrowed took out Stafford loans. The second most common loans were Perkins loans, but they were much smaller: only 11% of borrowers used Perkins loans and average amounts were one quarter of average Stafford amounts. Therefore, we focus on Stafford loans. Stafford offers multiple types of loans so we use the weighted average interest rate to set $\iota^s = 0.009$. The borrowing limit while in college in the model is the set to match the cumulative borrowing limit on Stafford loans ($23,000).

School taste In this class of models it is difficult to match the education intergenerational persistence so we follow previous studies that introduced school taste, also known as psychic costs of education, (e.g., Abbott, Gallipoli, Meghir, and Violante, Forthcoming; Krueger and Ludwig, 2016). We assume that school (dis)taste in utility terms follows the function $\epsilon(\phi, \theta) = exp(\alpha + \alpha_\theta log(\theta) + \phi)$. This function allows for higher skilled individuals to have (on average) lower levels of school distaste if $\alpha_\theta < 0$. Then, ϕ is an idiosyncratic shock which is assumed to follow a normal distribution $N\left(\bar{\phi}_{e_p} - \frac{\sigma_{\phi^2}}{2}, \sigma_\phi\right)$ that has a possibly different mean depending on the parents' education. Without loss of generality, we assume that this mean is zero for children of high-school graduates. Even though all parameters are related, it is intuitive to think that α is estimated to match the college graduation share, α_θ is estimated to match the relation between college graduation and skills, $\bar{\phi}_{e_p}$ is estimated to match the intergenerational persistence of education, and σ_ϕ is estimated to match the variance in college graduation after controlling for skills.

Wages We assume that wages follow the process estimated in Section 4. Recall that this estimation implied that returns to skills are higher for college graduates.

Intergenerational transmission of skills Cunha, Heckman, and Schennach (2010) estimate children's future skills as dependent on children's current skills, parents' skills, and an index of parental investments—which is an unobserved factor in their estimation.[30] We assume that the child development function is of the nested CES form. The outer CES is based Cunha, Heckman, and Schennach—including their estimated values—but we parametrize and estimate the investment factor as an inner CES, with parental time and monetary expenditures as inputs. Hence, the functional forms are

$$\begin{aligned} \theta_k' &= \left[\alpha_{1j}\theta_k^{\rho_j} + \alpha_{2j}\theta^{\rho_j} + \alpha_{3j}I^{\rho_j}\right]^{1/\rho_j} e^{\nu}, \quad \nu \sim N(0, \sigma_{j,\nu}) \\ I &= \bar{A}\left[\alpha_m m^{\gamma} + (1-\alpha_m)\tau^{\gamma}\right]^{1/\gamma}. \end{aligned} \tag{6}$$

We use Cunha, Heckman, and Schennach's preferred estimation for the outer CES, which requires allowing parameters to vary with the age of the children.[31] Their main findings is that skills are more malleable when children are young, i.e., the elasticity of substitution determined by ρ_j is larger the younger the children. Moreover, they allow for two separate types of skills (cognitive and non-cognitive) so we enlarge our model to include both: θ and θ_k are vectors with a separate entry for each skill.[32] Cunha, Heckman, and Schennach highlight that abstracting from the two types of skills leads to estimates that suggest that investments on low-skilled children are much less productive (i.e., a more negative ρ_j).

Investments I are assumed to be unique and cannot be separated between skills—just as assumed by Cunha, Heckman, and Schennach. Then, we estimate α_m to match the average ratio of money to time investments, γ to match the correlation between the two investments, and $\bar{A}$ to match the efficiency of these investments such that the average level of cognitive skills in the estimated economy is one, since we estimated the income process under this normalization.

Agostinelli and Wiswall (2016a) use a Monte-Carlo simulation exercise to show that Cunha, Heckman, and Schennach's estimates may be biased. In particular, they suggest that estimates of ρ_j may be biased towards zero, and estimates of α_{1j} may be upward biased. Agostinelli and Wiswall (2016b) proposes another methodology but they are only able to apply it on children who are over 5 years old, hence limiting its use for our purposes. However, given the importance of the skill production function, in Section 6.2 we evaluate how our results would change if parameters are moved in the direction suggested by Agostinelli and Wiswall (2016a) and show that our main results are robust to relatively large changes in the parameters.

[30] Cunha, Heckman, and Schennach (2010) estimate it on a representative sample—rather than a sample of disadvantaged children—which is important for our purposes. The initial draw of θ_k is also estimated by Cunha, Heckman, and Schennach so we use their estimates on the covariances—Appendix Table 10-3 in their paper—of these initial draws in our model. In particular, given the importance of non-cognitive skills in the development function, we implement this by making children's draw of non-cognitive skills correlated to parents' non-cognitive skills. Then, the initial draw of cognitive skills are correlated to the initial draw of non-cognitive skills.

[31] Appendix Table C4 shows the parameter values and standard deviations.

[32] Similarly, α_θ is a vector that relates each skill to school taste and ν is a vector with independent shocks for each skill.

Preferences We specify the period utility over consumption and labor as

$$u(c,h)=\frac{c^{1-\gamma_c}}{1-\gamma_c}-\mu\frac{h^{1+\gamma_h}}{1+\gamma_h}.$$

We follow the literature and assume that $\gamma_c = 2$ and $\gamma_h = 3$ (i.e., the Frisch elasticity is 1/3).[33] μ is estimated to match average hours of labor. When parents choose their time with children τ, the disutility is assumed to be linear, i.e., $v(\tau) = \xi\tau$. ξ is estimated to match estimated average hours with children. Finally, the altruism factor b is estimated to match the average monetary transfers from parents to children, as estimated from the Rosters and Transfers supplement to the PSID. We estimate average transfers per age-group of children and obtain an estimate of total parental transfers per child of \$37,300—which correspond to 62% of average annual income.[34] The benefit of using this data is that we can keep the estimation sample consistent with the ones used for other moments, but an important caveat is that, given the data structure, we are unlikely to observe bequests or late-in-life transfers. Nevertheless, we find that our estimates of b are in line with the literature. [35]

Aggregate production function We assume there is an aggregate firm with production function $Y = K^{\alpha}H^{1-\alpha}$, where H is a CES aggregator of the labor supply of the two education groups

$$H=\left[sH_1^{\Omega}+(1-s)\,H_2^{\Omega}\right]^{1/\Omega}.$$

We set $\alpha = \frac{1}{3}$ and estimate the CES aggregator. We estimate $\Omega = 0.43$ and s to be 0.53. This leads to a elasticity of substitution between high-school and college graduates of $\frac{1}{1-\Omega} = 1.75$ which is close to previous estimates (e.g., Katz and Murphy, 1992; Heckman, Lochner, and Taber, 1998). See Appendix C for details on the estimation. The per-period capital depreciation rate δ is set such that the annualized depreciation rate is 6.5%.

5.2 Simulated Methods of Moments: Results

Thirteen parameters of the model are estimated using simulated method of moments. b relates to the degree of altruism, while μ and ξ are the disutility of labor and time spent with children, respectively. α,

[33]See Meghir and Phillips (2010) for a discussion on estimates of the Frisch elasticity.

[34]This procedure is similar to the one used to estimate total fertility rates which is useful since only information on recent transfers is typically available—i.e., not the history of all transfers.

[35]An alternative procedure to estimate b would be to use as a target moment an estimate of the total size of family transfers based on sources different from the PSID. For example, Brown and Weisbenner (2004) use a flow-to-stock conversion methodology to estimate that the share of wealth explained by parental transfers is around 29% (though this estimate depends substantially on the interest rate used). Obtaining a similar estimate in our model would require a share of total parental transfers equal to approximately 85%, which implies increasing the altruism factor by between approximately 0.05 and 0.08 as seen in Appendix C.4. Our robustness results in Section 6.2 show that b is important for our results, but a change of this magnitude is only likely to reduce the welfare gains of a government policy that invests in children's development—our main results—by less than one-third of our baseline estimates.

α_θ, $\bar{\phi}$, and σ_ϕ relate to the distribution of school taste and its relation with skills and parental education. $\bar{A}$, α_m, and γ relate to the effect of parental time and money investments in building skills. ω relates to the government's redistribution of income. Finally, ι is the wedge in the interest rate between saving and borrowing.

We implement a simulated method of moments procedure in two steps, based on a minimum distance estimator. In addition to our target moments calculated with the full sample (call these M_0), we calculate target moments M_n by bootstrap, for $n = 1, ..., N$. Second, we use a sobol sequence in order to estimate the model in a thirteen-dimensional hypercube in which parameters are distributed uniformly and over a "large" support. This provides a global method to find potentially good combinations of parameters. In Appendix C.4 we show how the information from this step can be used to justify the selection of each moment, i.e., why each moment is informative for each parameter. The drawback is that the distance between parameter sets may be large. We obtain the set of parameters P_n that best fit each moments M_n, for $n = 0, ..., N$. Table 4 shows the estimated parameters P_0 and the corresponding moments in the simulated economy. The standard deviation is obtained using P_n for $n = 1, ..., N$.

Table 4: Estimation: parameters and moments

Parameter	**Value**	**Std. Dev.**	**Description**	**Moment**	**Data**	**Model**
Preferences						
μ	1202	(31.2)	Mean labor disutility	Avg. hours worked	64.5	64.5
b	0.327	(0.005)	Altruism	Parent-to-child transfer as share of income	0.62	0.65
School Taste:						
α	6.24	(0.69)	Avg. taste for college	College share	29	29
α_{θ_c}	-0.70	(0.18)	College taste and cog. skills relation	College: cog skills slope	0.19	0.21
$\alpha_{\theta_{nc}}$	-0.40	(0.11)	College taste and noncog. skills relation	College: noncog skills slope	0.06	0.07
σ_ϕ	2.16	(0.31)	SD of college taste shock	College: residual variance	0.18	0.17
$\bar{\phi}$	-3.33	(0.61)	Draw of school taste: mean by parent's education	Intergenerational persistence of education	0.77	0.76
Skill Formation Productivity:						
ξ	0.19	(0.09)	Parental time disutility of time with children	Avg. hours with children	19.1	17.5
$\bar{A}$	271.0	(20.6)	Returns to investments	Average skill ratio	1.0	1.0
α_m	0.93	(0.03)	Money productivity	Ratio of money to hours	208	191
γ	-0.384	(0.16)	Money-time substitutability	Money-time correlation	0.88	0.90
Interest rate						
ι ($\times 10^2$)	2.42	(0.42)	Borrow-save wedge	Share of borrowers	5.0	4.6
Government						
ω ($\times 10^2$)	2.28	(0.08)	Lump-sum transfer	Income variance ratio: Disposable to pre-gov	0.69	0.70

Notes: Parent-to-child transfers, hours worked, skill formation moments and intergenerational persistence of education are estimated from PSID-CDS data. Share of borrowers is estimated from Survey of Consumer Finances. College share, college-skills slope and college residual variance are estimated using NLSY. Bootstrap standard deviations in parentheses. All moments matter for all parameters, but each line highlights the moments that is particularly informative for the corresponding parameter. See Appendix C.4 for more details.

The model provides a good fit of the data. The distribution of education, its relation with skills, and its intergenerational persistence are close to their data counterparts. Average time working and with children are successfully matched. The relation between money and time investments is well captured in the model. Finally, the share of borrowers in the simulated model is similar to the one found in the Survey of Consumer Finances. We also remark that the average productivity of parental investments is selected such that the average level of skills in the economy is around its normalized value of one. The income redistribution in the model, as measured by the ratio of the variances of log disposable-income and log pre-government-income, is close to its empirical estimate.

In Section 6.2 we test the importance of these parameters by looking at how much results change when each parameter is changed according to its standard deviation. Average parent-to-child transfers are slightly high in the model but we highlight that the estimated altruism factor $b = 0.327$ is in line with the literature (e.g., Manuelli and Seshadri, 2009; Abbott, Gallipoli, Meghir, and Violante, Forthcoming; Lee and Seshadri, Forthcoming). Moreover, it has a small standard deviation due to the observed size of parental transfers and how much those transfers change in the model when the parameter is moved. A second parameter of interest is the substitutability between time and money investments given by $\gamma = -0.384$. This suggests that the elasticity of substitution is equal to 0.72.[36] Even though the standard deviation for this parameter is not as tight as others, Section 6.2 shows that our main results are almost unaffected when we move γ within one standard deviation.

5.3 Validation Exercises

We test the validity of the estimated model in two ways. First, we look at relevant moments which are not directly targeted in the estimation. Then, we follow a more novel approach of using experimental evidence to test the model predictions when a policy related to childhood development is introduced.

Non-targeted moments

Table 5 summarizes the first validation results, i.e., those from non-targeted moments. As shown in Section 4, families from higher socioeconomic groups tend to invest more time and money towards their children. The estimated model displays similar qualitative features. The elasticities of log-hours to college and log-income are similar but slightly smaller in the model than in the data when using all types of families. These elasticities become even more similar if we use only families with two parents and two children (the closest to the model). Regarding the elasticities of log-expenditures, the model tends to produce larger elasticities than the data, which may be due to the fact that expenditures typically

[36] The key moment behind this parameter estimation is the correlation between time and money investments seen in the data. A relevant concern is that this is driven by effects other than the production function's substitutability. For example, heterogenous altruism levels (or other parental characteristics not considered in the model) could lead to correlation between the two. We find, however, that even if we focus on more homogeneous sets of parents (e.g., by education, income, number of children, or marital status), who are likely to be more similar in these other characteristics, the correlation between these two is always above 0.7, which would lead to similar levels of complementarity in the estimation.

associated with higher-income groups (e.g., child care and school fees) are not available in the data used to calculate these elasticities.

Chetty, Hendren, Kline, and Saez (2014) estimates intergenerational mobility, as measured by the income rank persistence across generations, to be between 0.26 and 0.29 for children of families with married parents (the closest to our agents in the model). In the model this persistence is 0.25, close to that range. Labor income inequality is also well captured by the model: both the Gini and top-bottom coefficients are below but similar to the data.[37] Regarding savings, the (annualized) capital-output ratio in the estimated model is 3.1 which is above but close to its typical empirical estimate of 3.

Table 5: Validation: Not targeted moments

Moment	Data		Model	Source
Investments towards children:				
	All Families	Homogeneous Families		
Hours to College	0.19	0.13	0.07	CDS
Expenditures to College	0.45	0.26	0.56	CDS
Hours Elasticity to Income	0.18	0.04	0.10	CDS
Expenditures Elasticity to Income	0.50	0.54	0.99	CDS
Intergenerational Mobility				
Rank-Rank coefficient	0.26–0.29		0.25	Chetty
Inequality				
Gini	0.32		0.30	PSID
Top-Bottom	3.9		3.4	PSID
Savings				
Capital-Output Ratio (annualized)	≈ 3		3.1	

All moments are computed using the estimated model in steady state. Moments on investments towards children were calculated for children around age 4 in the data. All families refers to all kids in the sample. Homogeneous families refers to households with two adults and two children. In the model we use the average between the corresponding moments for children of age 0–3 and of age 4–7. Chetty refers to estimates for children of married parents by Chetty et al (2016).

Using experimental evidence

We use experimental evidence to test the validity of the most important novelty in our model: childhood development. García, Heckman, Leaf, and Prados (2017) study a randomized control trial (RCT) in which a small group of disadvantaged children were introduced into two high-quality early childhood

[37]Top-bottom refers to the ratio of average incomes between the top 80–95 percentiles and the bottom 5–20 percentiles.

development programs (ABC and CARE in North Carolina) that cost approximately $13,500 per year.[38] The kids entered the program when they were around 8 weeks old and stayed for five years.

We introduce a similar policy in the model. From the steady state, we simulate a policy in which the government unexpectedly introduces money directly in the early development of some children. This involves adding money directly in the children's development function in equation 6—unexpectedly, when they are 0–4 years old, and for only one generation. In the model, this is introduced as the government spending g directly on the child's skills such that

$$I = \bar{A}\left[\alpha_m (m+g)^{\gamma} + (1-\alpha_m)\tau^{\gamma}\right]^{1/\gamma} . \tag{7}$$

We assume that g is a perfect substitute of m, as if both were used to acquire early childhood education goods available in the market. Although unexpected, parents are allowed to change their choices (including m and t) after g is introduced. Government investments g will crowd out parental money investments m (until $m = 0$) while the effect on time τ is not obvious since the estimation results imply that money and time are imperfect complements.[39]

The policy in the model is introduced with three specific characteristics in order to be comparable to the RCT. First, the RCT focused on a small group of children so prices in the economy would not be affected. Thus, when we introduce the policy in the model we abstract from wage and interest rate changes. Second, the experiment focused on disadvantaged children of low-educated and low-income parents. Therefore, we also study the effect of such a policy in the model on children of high-school educated parents whose income and savings are low (i.e., among the bottom 20%). We further focus on children with low initial draws of skills (i.e., in the bottom third) in order to further focus on disadvantaged children but results are similar if we abstract from this. We refer to this as the baseline target population but we also show results for other target populations. Finally, the children introduced into this RCT did not expect their own children to also participate in the program. Therefore, the policy in the model is also introduced as a temporary one-generation policy in which effects are evaluated on the targeted generation.

García, Heckman, Leaf, and Prados observe the education and income of these children at two specific ages (the latest being age 30). Table 6 shows that education and income gains in the model are in line with their findings. Income at age 30 increases by very similar amounts for intervened children in our model as does in their RCT. They also show that the policy led to an increase in the college graduation rate of approximately 13.5 percentage points. The model predicts large increases in this rate as well,

[38] They report the cost was $18,000 (in 2016 dollars) but adjusting prices by inflation (to obtain prices in 2000 dollars) brings the cost down to $13,500.

[39] A potential interpretation for m is that it is buying time with teachers in early childhood centers. Our baseline model does not include a market for this (which may be particularly relevant for the large-scale policy analysis in Section (6)), but in Section 6.3 we discuss an extension in which producing this early childhood input requires a college-educated individual's time. Thus, a large-scale government intervention may affect the cost of early childhood development goods. We find, however, that this is effect is counterbalanced by the increase in the supply of college graduates that the policy generates.

though slightly below the ones observed in the data. García, Heckman, Leaf, and Prados further use these effects on income and education, together with a life-cycle-income profile, to predict a return in lifetime earnings (in net present value) of 1.3 dollars for every dollar spent. Our model, due to differences in the life-cycle-income profile, predicts an smaller return of slightly above 1.2. Table 6 also shows the effect predicted by the model if the policy is targeted to alternative (less disadvantaged) groups of children. More advantaged groups children tend to have higher levels of parental investments so there is more room for crowding out when the policy is introduced. Therefore, gains from being introduced to this policy are smaller for these individuals.[40]

Table 6: Validation: Experimental evidence

	GHLP (2017)	**Model**				
Target Population's Characteristics		**Baseline**	**Other Populations**			
Parent's Education		Low	Low	Low	Low	All
Parent's Labor Income		Low	Low	Low	All	All
Parent's Savings		Low	Low	All	All	All
Child's Initial draw		Low	All	All	All	All
Share College Graduates (p.p.)	13.5	12.6	11.2	10.8	8.6	8.2
Labor Income Age 30	$4.7k	$4.7k	4.2k	4.1k	3.4k	3.3k
Lifetime Labor Income (NPV) Return	1.3	1.2	1.1	1.1	0.9	0.8

We use the estimated model (starting from steady state) to simulate experimental evidence on early childhood development in the spirit of the study of Garcia et al. (2017)—GHLP in the table. We simulate paying monetary expenses of a value of $13,500 per child-year in the first period of childhood development (unexpectedly, for only one cohort, and target to a small group of disadvantaged children). We evaluate alternative target samples regarding parent's characteristics and child's initial skills. In the model, low education refers to high-school, low labor income or savings refer to the bottom 20%, and low initial skills refers to the bottom third at the time of birth. We then compare our findings regarding the effect on children's education and labor income to those of Garcia et al.

In Section 6 we study the effects of a similar government investment policy (though at a large scale), so the success of these validation exercises gives us confidence in the results we obtain in those counterfactuals.

6 Policy

As reflected in the model, children cannot invest in their own early childhood or compensate their parents for doing so, which can lead to reduced levels of childhood investment relative to an economy in which the children can compensate their parents for their investments. In addition, borrowing constraints and risk aversion together with uncertainty in returns can limit investments—as explained in Section 3.4 and explored theoretically in simpler models (e.g., Loury, 1981; Baland and Robinson, 2000;

[40]Even though Table 6 shows that the lifetime labor income return can be smaller than one it is important to highlight that these are not the long-run gains which, as we will show in Section 6, are much higher.

Pouliot, 2006). In particular, Loury (1981) provides a partial-equilibrium model in which government investments towards childhood are welfare improving even when (lump-sum) taxes are needed pay for them. Here we quantitatively evaluate such a policy in our estimated model, taking into account potentially negative effects through distortive taxation and general equilibrium effects. Moreover, our model features multiple periods and forms of parental investments as well as estimated returns to those investments, making it more appropriate for quantitative evaluations.

We evaluate alternative levels of government investments in early childhood but we focus on the effects of introducing the same amount of investments g as in the case of the RCT studied by García, Heckman, Leaf, and Prados (2017). We now study, however, a large-scale and long-run version of this policy so we need to take into account general equilibrium and distortive taxation effects as well as intergenerational effects. After evaluating the long-run effects, we study the transition dynamics from which we obtain one of our main results: most of the long-run gains of early childhood programs are given by the fact that investing in a child's development not only increases that child's skills but also creates a better parent for the next generation.

Welfare measure: Consumption equivalence for newborn under veil of ignorance

When evaluating policies we are interested in inequality, intergenerational mobility, and average income. As a summarizing measure that allows us to compare policies we look at welfare. This is defined by the consumption equivalence under the veil of ignorance, in the baseline economy relative to the economy with the policy in place.

Let $P = \{0, 1, 2, \ldots\}$ denote the policy introduced, with $P = 0$ being the initial economy in steady state. We refer to consumption equivalence as the percentage change in consumption λ in the initial economy that makes agents indifferent between being born in the initial economy ($P = 0$) and the one in which the policy P is in place. In particular, let $\tilde{V}^P_{J_i}(a, \theta, \phi, \lambda)$ be the welfare of agents with initial states (a, θ, ϕ) in the economy P if their consumption (and that of their descendants) were multiplied by $(1 + \lambda)$:

$$\tilde{V}^P_{J_i}(a, \theta, \phi, \lambda) = E^P \left\{ \sum_{j=J_i}^{j=J_d} \beta^{(j-J_i)} u\left(c^P_j (1 + \lambda), h^P_j\right) + \beta^{J_c} b \tilde{V}^P_{J_i}(\varphi, \theta_k, \phi', \lambda) \right\}$$

where, for the sake of clarity, we have abstracted in the presentation from including the school distaste and highlighting that policy functions depend on the states. Notice that these policy functions are assumed to be unchanged when λ is introduced. For example, consumption c^P refers to the consumption chosen by individuals in economy P and is unchanged by λ. Then, for any λ we can obtain a measure of average welfare

$$\bar{V}^P(\lambda) = \int_{a,\theta,\phi} \tilde{V}^P_{J_i}(a, \theta, \phi, \lambda)\, \mu_P(a, \theta, \phi),$$

where μ_P refers to the distribution of initial states $\{a, \theta, \phi\}$ in the economy P. Then, we define the consumption equivalence λ^P to be the one that makes individuals indifferent between being born in the baseline economy and in the one with policy P in place, i.e.,

$$\bar{V^0}\left(\lambda^P\right) = \bar{V^P}(0).$$

Welfare gains, by definition, come from two sources: (i) changes in the expected discounted utilities at each state $\tilde{V}^P_{J_i}(a, \theta, \phi, 0)$ and (ii) changes in the probabilities of each state $\mu_P(a, \theta, \phi)$. Having explained how we measure welfare, we now move forward with the policy evaluation.

6.1 Government Investments towards Childhood Development

We simulate a policy in which the government invests money directly in the development of children. This involves adding government investments in the children's development function in equation 6. In the model, this is introduced as the government spending g directly on the child's skills such that $I = \bar{A}\left[\alpha_m(m+g)^\gamma + (1-\alpha_m)\tau^\gamma\right]^{1/\gamma}$. Parents may alter their parental investments (and other choices) as government spending g is introduced. In particular, we would expect that introducing g would lead to crowding out of expenses m. However, complementarities may lead to an increase in time investments τ.

Our validation exercise, as shown in Table 6, directly tested the mechanisms involved by introducing government investments as a randomized controlled trial and using experimental evidence to compare the results. Its success gives us confidence in the policy evaluations we perform now. Differently from that validation, we now introduce government investments in a permanent and universal manner, i.e., investments are introduced for all children and for ever. Moreover, we also take into account that revenues need to be raised to afford these investments: the government alters (labor) taxes τ_y such that the government budget is unchanged. Finally, we look at the long-run effect of such policy, taking into account general equilibrium effects on the interest rate and wages.

We focus here on the case in which the government invests directly only during the children's first period (i.e., between the ages 0–3), evaluating this policy for different amounts of resources available. We have also evaluated introducing government investments in the second period (i.e., between the ages 4–7). Gains are also obtained from allocating resources to that period but they are smaller. Earlier investments lead to larger gains because the child skill production function implies that skills are more malleable at younger ages. Although we focus here only on very early investments, Appendix F.1 shows the results for other alternatives.

Results

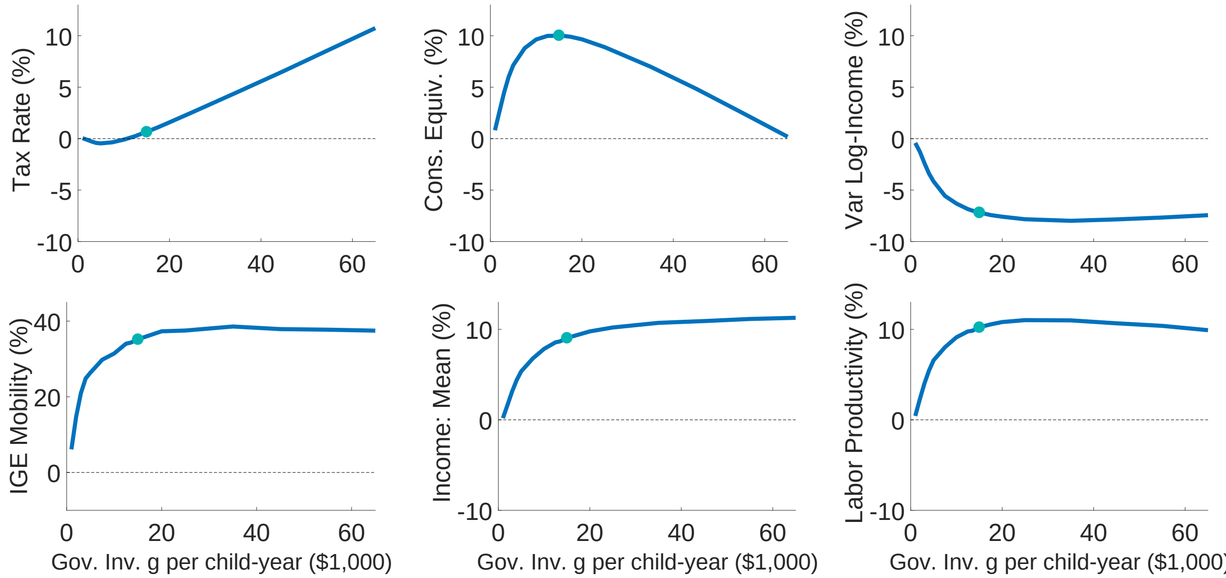
Figure 9 shows the results, with the horizontal axis referring to the amount of government early childhood investments g per child-year. Introducing government childhood investments leads to substantial welfare gains, of approximately 10% as measured by the consumption equivalence for a newborn under the veil of ignorance. The top-left figure shows that as expenditures per child are increased the tax rate needs to increase in order to keep the government's budget balanced. Investing in children raises the government's revenue since the tax base increases, but not enough to finance policies that require more than $10,000 per child-year.

As a reference point, the early childhood program evaluated by García, Heckman, Leaf, and Prados (2017) was estimated to cost approximately $13,500 per child-year (in 2000 dollar terms). Given that this is a level of investments that has actually been implemented and that the validation exercise was successful for such level, we will focus on this level of investments for the rest of our analysis. Moreover, as shown by the top-right figure, our estimates suggest that investing $13,500 is close to the long-run welfare maximizing level.

Agents prefer these resources to be used for childhood investments rather than to fund a government transfer. We evaluated another policy that uses the same resources to provide an initial transfer to every agent when they become independent at age 16. We found that such a policy would lead to a 4.3% long-run welfare gain, less than half of the one obtained using the same resources to fund the government investment program. This happens because the government can do something that these agents cannot do by themselves, i.e., invest in their childhood.[41]

[41]Funding a transfer program provides less welfare than using the same resources for childhood investments as long as the resources used are not too large. Once resources are over $120,000 per child (i.e., $30,000 per child-year if using it only in the first period), returns on those investments are small relative to the large cost of raising taxes to afford them.

Figure 9: Childhood investments



Notes: We simulate policies in which the government invests different amount of resources directly in the development of children ages 0–3. We evaluate this policy for different amounts of resources available. The horizontal axis refers to the investments per child-year. Then, for example, 15 refers to $15,000 per child-year, which equals $60,000 per child. Outcomes are reported in changes from the baseline steady state. Consumption equivalence is determined by newborns under the veil of ignorance. Inequality refers to the variance of log-labor-income while IGE mobility refers to minus the regression coefficient between children's and parents' income ranks.

The early childhood program is associated with a significant increase in income and intergenerational mobility as well as a reduction in income inequality. The model suggests that if such a universal investment policy were implemented it would lead to an increase in income of 8.7%, all of which is explained by the increase in labor productivity of 9.8%—as measured by the average product of wages and labor efficiency units. Intergenerational mobility—as measured by minus the rank-rank coefficient used by Chetty, Hendren, Kline, and Saez (2014)—would improve by 34%. This implies that the rank-rank coefficient is reduced by almost 0.1. Moreover, labor income inequality—as measured by the mean of the variance of log of pre-government-income by age—would be reduced by 7%. These last two changes are approximately large enough for the US to achieve Canadian or Australian levels of mobility and inequality.

Results decomposition: Long-run, general equilibrium, and taxation

Government early childhood investments achieve most of their effects on welfare and mobility through the long-run intergenerational dynamics: When the government invests in a child today, it not only creates better skills for that child but also creates a better parental background for the following generation.

Table 7: Results decomposition

Alternative Exercises			Change from Baseline (%)				
Long-Run	**General Equilibrium**	**Budget Balanced**	**Consumption Equivalence**	**Average Income**	**Labor Returns**	**Inequality**	**Mobility**
No	No	No	4.2	8.5	9.3	-1.4	18.0
Yes	No	No	11.7	13.1	15.7	-1.8	27.7
Yes	Yes	No	10.5	8.9	10.3	-7.0	34.7
Yes	Yes	Yes	**10.0**	**8.7**	**9.8**	**-7.0**	**34.3**

Notes: We simulate introducing the same level of investments as in Garcia et al (2017). Early childhood investments of $13,500 per child-year when children are between 0 and 3 years old are introduced. We simulate this policy closing down some channels to better understand the mechanisms. Long-run refers to looking at outcomes in the new long-run steady state. When this is deactivated we calculate the effect of a one-generation policy and evaluate the effect on that generation. General Equilibrium refers to adjusting wages and interest rates to clear the market. Budget Balanced refers to adjusting the labor income tax to keep the government's budget unchanged. Outcomes are reported in changes from the baseline steady state. The main welfare estimates refer to the case in which the three channels are activated. Consumption equivalence is determined by newborns under the veil of ignorance. Inequality refers to labor income inequality while intergenerational mobility refers to minus the regression coefficient between children's and parents' income ranks. Average income refers to mean labor income for young individuals (ages 28–31), to facilitate the comparison with the empirical evidence.

Table 7 decomposes the welfare gains by simulating the same policy applied in three alternative ways. The bottom row refers to the benchmark results (i.e., in the long-run taking into account tax changes and general-equilibrium effects). In the first row, we introduce government investments for only one generation, without balancing the government's budget or taking into account general equilibrium effects. Effects are evaluated on the generation that receives the intervention so this implementation can be compared to the one to be expected from an RCT which is typically of small scale and applied to only one generation. Differently from the case used in the validation, the policy is evaluated on a representative group of children, not on a disadvantaged group. We find that in this case welfare gains are only 4.2%, less than half of our benchmark. Mobility increases by only one-half of the benchmark increase, while inequality is reduced by only one-fifth of the benchmark reduction.

Next, we allow for long-run effects to take place: the policy is implemented permanently and the fact that by improving one generation's level of skills we are also improving the productivity of future investments is also taken into account. By permanently introducing this policy, the chances of children being born into a low-skilled family are reduced. Welfare gains increase by 7.5 p.p. (11.7 - 4.2), suggesting that long-run intergenerational dynamics generates three-fourths of the baseline welfare gains of 10.0%. Similarly, intergenerational mobility increases by one-fourth of the baseline mobility increase. The reduction in inequality, however, is still less than one-fourth of the baseline reduction.

It is general equilibrium forces that generate most of the reduction of inequality: Increasing skills augments the share of college graduates which reduces the wages of college-graduates relative to high-school graduates. Even though reducing inequality increases welfare, general equilibrium forces reduce welfare gains by one-tenth of the baseline value since reducing wages of college graduates reduces the labor productivity gains associated with this policy. Finally, taking into account that taxes need to be increased to finance this policy reduces welfare gains by approximately another one-tenth.[42]

To summarize, investing in children can lead to large welfare gains as well as changes in inequality and mobility. Long-run intergenerational dynamics—investing in a child today produces a better parent for the next generation—drive over two-thirds of the welfare gains and a sizable part of the increase of intergenerational mobility. This suggests that these gains may take a long time to accrue, but the transition dynamics studied below formally evaluates this concern. General equilibrium effects—by reducing the wages of college-graduates relative to high-school graduates—generate three-quarters of the reduction in inequality but reduce welfare gains by one-tenth since they also reduce labor productivity. Finally, raising taxes to pay for this policy reduces welfare gains by one-tenth as well.

Heterogeneous welfare gains

Recall that, by definition, welfare gains emerge from two sources: (i) changes in the value of an agent at each state $V_{J_i}(a, \theta, \phi)$, and (ii) changes in the distribution over those states $\mu(a, \theta, \phi)$. Parents are heterogeneous in their savings, education, skills, and idiosyncratic labor productivity, all of which affect the next generation's distribution over states (a, θ, ϕ). Consequently, welfare gains can be heterogeneous for children with different types of parents.

Children born to low socioeconomic groups benefit the most from this policy. Figure 10 shows the policy's welfare gains once the veil of ignorance is partially revealed. Here, we look at the consumption equivalence for an agent who knows her parent's skill level and education group, assuming this is unchanged between the baseline economy and the one in which the policy is in place.[43] Children of high-school educated parents benefit the most, with a consumption equivalence of up to 8%. On the other hand, agents that know they are going to be born to college-educated and high-skilled parents benefit much less, with a consumption equivalence of approximately 1%.

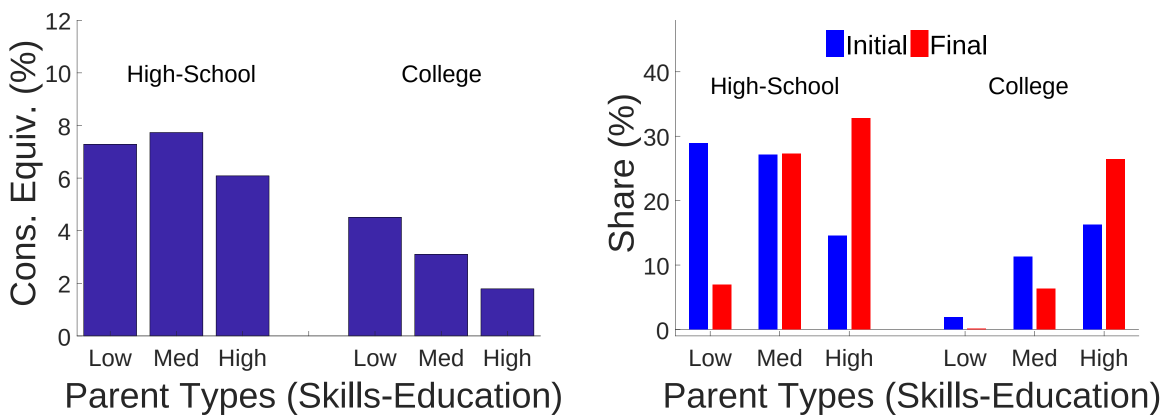
The welfare gains for each of these groups are smaller than the general welfare gains under the full veil-of-ignorance since a large part of the gains is driven by the long-run reduction in the share of children being born to low-skilled parents. The right panel of Figure 10 shows the share of children born to each of these groups in both economies. It is clear that once the policy is introduced the distribution of skills moves upward, reducing significantly the share of individuals born to low-skilled parents. This

[42] We find that most long-run welfare gains are driven the level effect on average consumption. As discussed by Bénabou (2002) and elaborated for non-homothetic utility functions in Abbott, Gallipoli, Meghir, and Violante (Forthcoming), total welfare change of the policy reform can be split into three components: (i) a level effect of the reform on the level of average consumption, (ii) an uncertainty effect on the volatility of the agents' consumption paths that affects welfare because of risk aversion and incomplete markets, and (iii) an inequality effect on the equilibrium distribution of initial conditions. We find that over 8/10th of the total welfare gains are driven by the level effects.

[43] Notice that there still is heterogeneity within each of these groups since parents' assets and idiosyncratic labor productivity can vary. We allow these two states to change when computing welfare gains for children of each group: Children know they are going to be born a parent with a given education and skill group in both economies, but they also know that the parent's distribution over assets and idiosyncratic labor productivity (conditional on education-skills) is different between those two economies.

reduction leads to changes in the distribution $\mu(a, \theta, \phi)$ towards states associated with higher utility $V_{J_i}(a, \theta, \phi)$. To understand the importance of this effect, we can recalculate welfare gains in two alternative ways: (i) fixing the distribution μ and only taking into account the changes in V_{J_i}; and (ii) fixing the values V_{J_i} and only considering the changes in the distribution μ. If we fix the distribution μ to the original steady-state, welfare gains are 2.4%, i.e., one-fourth of the total gains. On the contrary, if we fix the values V_{J_i} to the original steady-state, we find welfare gains 7.8%, i.e., three-fourths of the total gains. Clearly most welfare gains are driven by the fact that once the policy is introduced more children are born with states associated with higher utility—and not as much by the change in utilities at each state.

Figure 10: Heterogeneous effects



Notes: The left figure shows the welfare gains computed for children of parents with different cognitive skills and education levels, assuming each child knows she will be born to the same type of parent both in the original steady state and the one in which the policy is active. The right figure shows the share of each children born to each type of parent.

Transition dynamics

Given that we have shown that a substantial part of the benefits is driven by the long-run change of distributions, a logical concern is that a government investment policy may take too long to accrue this level of welfare gains and possibly go through periods in which welfare is reduced. We evaluate this by looking at the transition dynamics. [44]

We assume that the government investments are unexpectedly introduced at the previously defined level of $13,500 per child-year, together with the associated labor income tax change, and are known

[44]Studying transition dynamics, however, is complicated because there are many ways to implement the policy change (e.g., how the policy is introduced and/or financed in the transition). To the best of our knowledge, no paper has studied the optimal transition with such a degree of flexibility. Bakış, Kaymak, and Poschke (2015) studies a constrained form of optimal transition, in which the policy is assumed to be implemented immediately but they take into account the transition to define optimality. This would be interesting in our richer model but is beyond the scope of this study. Here we focus on studying the pace of the transition and highlighting that this policy can be welfare improving for every new generation in the transition.

to remain in place for ever. This change in labor income tax may not be enough to balance the government's budget (since the pool of skills in the economy takes time to change) so the government raises lump-sum taxes in the transition in order to balance its budget each period. Figure 11 shows the effects on all new cohorts, with cohort 0 being the first cohort to be intervened.

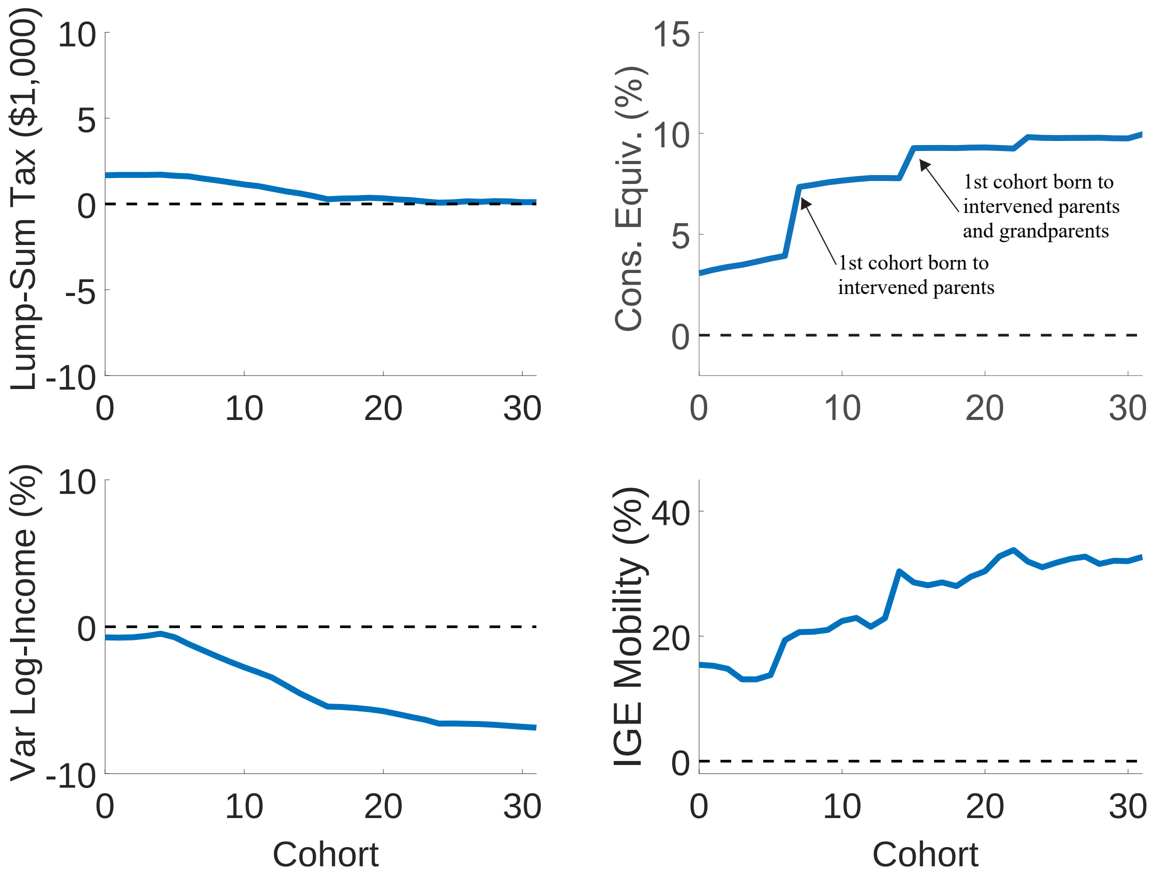
The first cohort to receive the government investments, as shown by the top-right panel of Figure 11, obtains a welfare gain of 3%, about one-third of the gains obtained by cohorts born in the new steady state. Welfare gains grow slowly cohort-by-cohort until a jump in the gains is observed for the first cohort born to intervened parents (i.e., those born 28 years after the policy is introduced). They obtain over 7% of welfare gains, i.e., around three-fourths of the final gains. Welfare gains grow slowly once again until a second jump is observed, i.e., for the first cohort born to intervened parents and intervened grandparents. These jumps in the consumption equivalence gains clearly show the mechanism behind the long-run intergenerational dynamics gains: investing in a child today creates a better parent for the next generation.[45]

The policy is associated with a steady decrease in income inequality that is paralleled by the change in wages. Appendix Figure D2 shows that wages of college graduate steadily decrease by up to 6%, while those of high-school graduates increase by up to 4%. Inequality takes time to be reduced since the pool of workers (with their distribution of education and skills) in the economy takes time to adjust. Intergenerational mobility also takes time to increase, with the first cohort displaying only half of the increase displayed by cohorts born in the new steady state.

The government needs to raise a lump-sum tax of $2,000 per household-year early on so that, together with the increase in labor taxes, its budget remains balanced. This lump-sum tax is slowly reduced such that after approximately 80 years it is eliminated. Appendix Figure D2 also shows that as the policy is introduced interest rates increase because agents optimally reduce savings, since future generations are more likely to be better off—due to both higher skills and lower lump-sum taxes.

[45]The small changes in gains for cohorts born between the jumps are due to the slow adjustment in prices and lump-sum taxes.

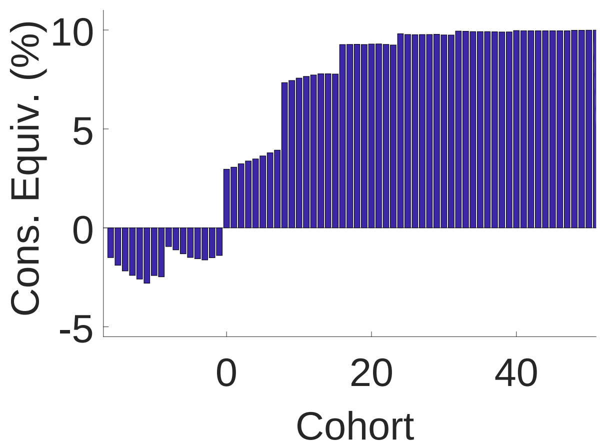
Figure 11: Transition dynamics



Notes: The policy (including the investments and labor income tax change) is introduced unexpectedly. We compute the transition introducing a lump-sum tax such that the government's budget balances every period. Consumption equivalence is shown for a newborn from the cohort defined by the horizontal axis. Cohort 0 is the first cohort to receive the government investments. Intergenerational mobility refers to minus the regression coefficient between children's and parents' income ranks. It is calculated for the generation born in such cohort and their parents. All values are relative to the initial steady state.

Older generations alive at the time the policy is introduced lose. Figure 12 shows that the welfare losses for each cohort alive at that time are between 1 and 3%. Individuals aged 44 and above (i.e., cohorts under -10) are not receiving any (non-pecuniary) gain from this policy since their children—and any future generation—are not included in their utility, but they are paying higher taxes. Individuals between 32 and 40 years old benefit indirectly through their grandchildren (who are going to receive the government investments). Individuals under 28 years old benefit through their children. Nevertheless, these gains are not enough to compensate the losses coming from higher taxation. Older generations are paying for the gains that are being accrued mostly by future generations, so a policy such as government borrowing that manages to pass the cost to these future generations may be able to reduce the losses for the older generations. Alternatively, a slow introduction of the government investments may also help older generations since initial costs would be reduced. More research on optimal transitions is necessary, but in Appendix D we show that government borrowing and slow introduction of the policy can help make welfare gains more common across cohorts. In particular, we find that a combination of both government borrowing and slow introduction of the policy is able to achieve welfare gains for all future generations and most of of the old individuals alive at the time the policy is introduced. This way to finance the transition highlights, once again, the main missing market in the economy: The government borrows to invest in early childhood development and finances this investments by raising the taxes these intervened children need to pay as adults.

Figure 12: Welfare gains including older cohorts



Notes: Welfare gains are reported for cohorts born after the policy is introduced (i.e., cohorts from 0 on) as well as for cohorts already alive at such time (i.e., cohorts less than 0). For the first group, welfare gains are computed for newborns. For the cohorts already alive at the time the policy is introduced, welfare gains are computed for agents with the appropriate age. For example, cohort -10 was born 40 years before the policy is introduced, so its welfare gains are computed according agents of age 40 at the time.

6.2 Robustness

We now evaluate how sensitive our main results on welfare gains are to changes in parameters. We start by moving the estimated parameters according to their standard deviations as reported in Table 4. We move each parameter one-by-one from the baseline estimation and recalculate the original steady state. Then, we introduce the government investments towards early childhood just as in the previous section and calculate the welfare gains. Table 8 reports the gains in the short-run (i.e., for the first generation) partial-equilibrium case when taxes are not adjusted—similar to an RCT applied to a small representative family—as well as for the long-run general-equilibrium case when labor income taxes are adjusted such that the government's budget is re-balanced.

Welfare gains do not change by more than one-tenth when parameters are moved in either direction. Moreover, we estimate that the total window of possible changes to the gains—by moving each parameter between plus/minus one standard deviation—is at most 12% of the total gains. However, it is instructive to analyze the effect of some parameters. First, the larger the parental altruism b the smaller the welfare gains, perhaps because underinvestment is less likely to occur—possibly because it is less likely that the parent-to-children transfer constraint binds as in Baland and Robinson (2000). A second important set of parameters involves the college distaste. Larger values of α, related to the average college distaste, and smaller σ_ϕ, related to its standard deviation, are associated with larger welfare gains. This is probably because agents are more likely to be low educated when either of these occur and gains are larger for those agents. Regarding the child's skill investment function, a larger $\bar{A}$ reduces the welfare gains of introducing government investments since the original parental investments are already more productive than in the baseline case. Interestingly, moving γ, the parameter controlling the elasticity of substitution between parental time and money investments, within one standard deviation does not seem to affect welfare gains by more than one-twentieth.

Table 8: Welfare gains robustness to estimated parameters

	Cons. Equiv. Change from Baseline					
	Short-Run PE			**Long-Run GE**		
	Down	Up	**Total**	Down	Up	**Total**
b	0.04	-0.05	0.09	0.06	-0.16	0.23
μ	-0.00	0.00	0.01	-0.05	0.03	0.08
α	-0.00	-0.04	0.04	-0.41	0.64	1.05
α_{θ_c}	0.05	-0.06	0.11	0.01	-0.01	0.03
$\alpha_{\theta_{nc}}$	0.00	-0.00	0.01	-0.01	0.03	0.04
$\bar{\phi}$	-0.01	-0.00	0.00	-0.03	0.19	0.21
σ_ϕ	-0.03	-0.01	0.02	0.72	-0.42	1.13
$\bar{A}$	0.05	-0.09	0.14	0.38	-0.40	0.78
α_m	-0.08	0.05	0.13	-0.32	0.35	0.67
γ	0.03	-0.06	0.09	0.13	-0.36	0.49
ξ	-0.01	-0.01	0.00	-0.03	-0.01	0.02
ι	-0.00	-0.00	0.00	-0.03	-0.03	0.00
ω	0.02	-0.03	0.05	0.08	-0.03	0.11
Baseline		**4.2**			**10.0**	

Starting from the baseline estimation, we move each parameters according to its standard deviation as reported in Table 4: Up (Down) refers to the the estimated value plus (minus) one standard deviation. Total reports the absolute value of the difference in reported results between Up and Down, i.e., a measure of by how much may each parameter affects the results. We solve the model for each given parameter set, introduce the same policy from the previous Section, and report the consumption equivalence welfare gains. Short-Run PE refers to the short-run partial-equilibrium gains without adjusting taxes and Long-Run GE refers to the long-run general-equilibrium gains adjusting labor income taxes such that the government's budget is balanced.

Given the importance of the child's skill production function in the model we also the study how sensitive our results are to changes in those parameters. Recall that the values used here are from Cunha, Heckman, and Schennach (2010), so we move parameters according to their reported standard deviations. We move each parameter one-by-one, re-estimate the model—particularly to guarantee that the average set of skills remains normalized—and calculate the original steady state. Table 9 reports the change in welfare gains from introducing the same government investment policy.

Table 9: Robustness to child's skill production function

	Change from Baseline					
	Cons. Equiv. SR-PE			**Cons. Equiv. LR-GE**		
	Down	Up	**Total**	Down	Up	**Total**
α_1	0.02	-0.79	0.81	0.20	-1.93	2.13
α_2	0.47	-0.94	1.41	0.79	-1.89	2.68
α_3	-0.20	-0.16	0.04	-0.29	-0.33	0.04
ρ	-0.67	0.39	1.06	-1.17	0.66	1.82
σ_v	-0.14	-0.22	0.08	-0.17	-0.55	0.37
$Var\left(\theta_{k_0}\right)$	-0.14	-0.14	0.01	-0.31	-0.23	0.08
$Corr\left(\theta, \theta_{k_0}\right)$	-0.12	-0.00	0.12	-0.40	0.51	0.90
Baseline		**4.2**			**10.0**	

We move each parameter by one standard deviation as reported by Cunha, Heckman, and Schennach (2010): Up (Down) refers to the the estimated value plus (minus) one standard deviation. Total reports the absolute value of the difference in reported results between Up and Down, i.e., a measure of by how much may each parameter affects the results. We re-estimate the model for each given set of parameters, introduce the same policy from the previous Section, and report the consumption equivalence welfare gains. SR-PE refers to the short-run partial-equilibrium gains without adjusting taxes and LR-GE refers to the long-run general-equilibrium gains adjusting labor income taxes such that the government's budget is balanced.

We find that results are more sensitive to changes in these parameters than to the ones internally estimated as reported in Table 8. However, moving parameters by one standard deviation never affects welfare gains by more than one-fifth, and the total welfare changes resulting from parameter changes between one standard deviation above and below are at most one-third, i.e., keeping long-run welfare gains always above 7%. As suggested by Cunha, Heckman, and Schennach (and the early childhood development literature in general), the elasticity of substitution between children's skills, parents' skills and investments is among the most important parameters. According to our results, increasing the elasticity of substitution parameter ρ by one standard deviation would increase welfare gains by approximately one-tenth. The larger the elasticity of substitution, the easier it is for investments to help children with bad initial conditions at birth—either because of the initial draw of skills or parent's characteristics—making investments more effective.

We find that two other parameters are as important as the elasticity parameter ρ. The larger α_1 the more persistent skills are, so it is harder for investments to be effective. Similarly, the larger α_2, the more important parental characteristics are so the more ineffective government investments are.[46] An

[46]This intuition is based on an initial elasticity of substitution above one.

additional takeaway from this exercise is that, at least according to our model, the elasticity of substitution is not the only important parameter for the magnitude of gains from investing in children. The persistence of initial skills and the importance of parents in the skill's production function may be as important as the elasticity of substitution.

Agostinelli and Wiswall (2016a) use a Monte-Carlo simulation exercise to show that Cunha, Heckman, and Schennach's estimates may be biased. In particular, they suggest that estimates of ρ may be biased towards zero and estimates of α_1 may be upward biased. This critique implies that inputs may be more substitutable in the production of cognitive skills—given that the baseline ρ_1 for cognitive skills when children are young is above zero—and that skills may not be as persistent as in our baseline estimation. According to Table 8, both of these effects would suggest welfare gains may be larger than in our baseline results. Nevertheless, the bias in the substitutability parameter also implies that inputs may be less substitutable in the production of non-cognitive skills—given that the baseline ρ for non-cognitive skills is below zero—which would suggest gains may be smaller than in our baseline estimation. The net effect is ambiguous but Table 8 shows results change by at most one-third for a two-standard-deviation change in any single parameter, so only very large biases are likely to significantly affect our main results.

We also estimated our model in an economy with only cognitive skills—using the estimates reported in the Appendix of Cunha, Heckman, and Schennach (2010)—and found long-run welfare gains to be twice as large while short-run gains were almost unchanged. We believe that this effect is driven by two reasons. First, in the world with cognitive and non-cognitive skills the estimation suggests that it is non-cognitive skills that matter relatively more for the productivity of parental investments, while it is cognitive skills that matter relatively more for income. This introduces more equality in the original steady state in the economy with two skills relative to the economy with only one. Second, the estimated values in the only-cognitive-skills economy suggest that the elasticity of substitution is lower and parental skills are more important than in the two-skills case. Since the long-run effect of government investments improves the distribution of parental skills, investments become particularly more productive in the world with a low elasticity of substitution and a large role for parental skills. More details on these results are available upon request, but we believe more research should be done on how welfare gains in this style of models are affected by increasing the variety of skills included.

6.3 Extension: With Early Childhood Education Market

In the baseline model we assume that the early childhood money input is equal to the good produced in the economy. This constant returns to scale assumption misses that the elements required to produce this early childhood development input may be scarce. In this section we provide a simple extension to the model in which this early childhood input is actually hours with a college-educated individual—which is in line with the costs reported for the RCT program in which we base our main analysis

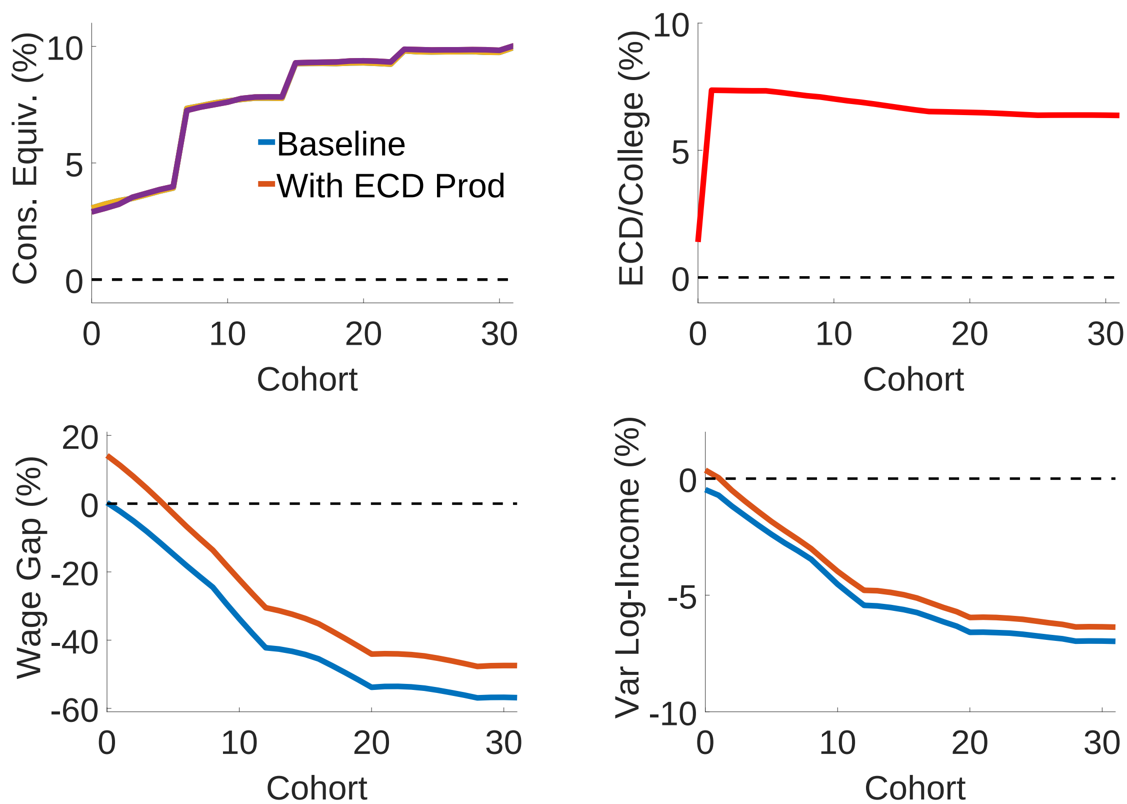
(García, Heckman, Leaf, and Prados, 2017).

The price of early childhood is now given by the wage of college-educated individuals, hence transforming the investment function I to

$$I = \bar{A}\left[\alpha_m \left(\frac{m+g}{w_2}\right)^{\gamma} + (1-\alpha_m)\tau^{\gamma}\right]^{1/\gamma}.$$

We re-estimate this model (see Appendix E) and introduce the same baseline policy in which the government invests \$13,500 per child-year. Figure 13 shows the effects of this policy, highlighting the main differences relative to the baseline case.

Figure 13: Baseline vs. early childhood production function extension: transition differences



Notes: The policy (including the investments and labor income tax change) is introduced unexpectedly. We compute the transition introducing a lump-sum tax such that the government's budget balances every period. Consumption equivalence is shown for a newborn from the cohort defined by the horizontal axis. Cohort 0 is the first cohort to receive the government investments. ECD/College refers to the share of college labor that is used in early childhood development. Wage gap refers to gap between the equilibrium wages for college and non-college workers ($w_2 - w_1$). All values are relative to the initial steady state.

On the one hand, the reform will now drive up the cost of early childhood as college-educated labor is a scarce input. On the other hand, over time the policy itself will increase the share of college-educated individuals, hence driving the cost down. In the long-run we find that both effects almost compensate each other and welfare gains are almost unchanged. During the transition, however, the effects are different, particularly regarding inequality. The increase in the demand for college workers drives up their wages, hence increasing the wage gap and inequality early on. This makes welfare gains slightly more homogeneous during the transition as well, since now the children of high-income parents (who were not likely to win much from the intervention) gain from the increase in their college wages (since these children are more likely to be college graduates).[47]

6.4 Other Policies

Parenting education In Appendix F.2, we evaluate another popular policy regarding childhood development: parenting education. These programs focus on teaching parents techniques and games to solve discipline problems and to foster confidence and capability. The key difference is that here, rather than investing towards children directly, parents are trained on how to promote children's development. We estimate the costs and returns of running a parenting education program based on the randomized control trials evidence from Gertler, Heckman, Pinto, Zanolini, Vermeersch, Walker, Chang, and Grantham-McGregor (2014) and Attanasio, Fitzsimons, Grantham-McGregor, Meghir, and Rubio-Codina (2016). A relevant caveat is that this evidence is from developing countries, but we try to control for that (see Appendix for details). We implement this program as a government policy (both in a partial-equilibrium framework similar to an RCT and in general equilibrium) as well as a new education good that parents can acquire in a private market. Our results suggest that such a program has the potential to increase welfare by between 7 and 8% as well as substantially reduce inequality and increase mobility. Importantly, once again we find that welfare improvements in the long-run general-equilibrium framework are larger than if we apply the policy as an RCT. The long-run change in the distribution of parental characteristics is important to obtaining all the benefits, since a higher-skilled distribution of parents provides better conditions for children.

College subsidies In Appendix F.3, we evaluate introducing government-funded college subsidies into our estimated model. This implies that the private cost of college is now reduced from p_e to $p_e(1-s_e)$ where s_e is the subsidy rate. Abbott, Gallipoli, Meghir, and Violante (Forthcoming) study this type of policies in a framework in which college subsidies have more flexibility than ours. However, differently from theirs, in our model college subsidies can affect skills through endogenous parental investment choices. We evaluate alternative values of s_e, increasing labor income tax to keep the government's budget balanced. We find that there are welfare gains from subsidizing college, and this tend to be larger for larger values of s_e. Full college subsidies (i.e., $s_e = 1$), however, are associated with an increase in college graduation rates from 29% to 33% and welfare gains of 1.7%, i.e., less than one-fifth of the one obtained

[47] In the long run, the important assumption is whether the relevant input required for early childhood development can be "produced." And whether a higher skilled population makes the provision of such input easier or less costly. This is left for future research.

by the program that funds early childhood development investments. Although this policy is associated with increases in parental investments and average skills, we find these changes to be relatively small (at least for this size of college subsidies and change in college graduation rates), suggesting that the results from Abbott, Gallipoli, Meghir, and Violante (Forthcoming) may not be substantially affected if they introduced endogenous parental investments towards early childhood skills development.

(Constrained) Optimal tax progressivity Finally, in Appendix F.4, we study the importance of endogenous childhood development for optimal taxation—within our constrained tax function. Macroeconomic analysis of inequality often focuses on progressive taxation but most of the models used abstract from endogenous intergenerational links such as childhood development. In addition to the traditional trade-off between equality and efficiency of labor, endogenous childhood skills lead to a new trade-off. On the one hand, higher progressivity may let poor parents increase investments towards child's skills. On the other hand, such progressivity would increase insurance and reduce the after-tax returns to skills, thus reducing the incentive to invest towards children's skills. In general, the question that emerges is: *Do tax policy evaluations change once we include endogenous parental investments?* The model introduced here is useful to answer this question as it adds endogenous childhood development and parental transfers to a standard life-cycle macroeconomic model.

In order to evaluate the importance of endogenous intergenerational links, we compare the effects in our model with endogenous links to the effects obtained by the same model but with exogenously fixed links. The model with exogenous childhood development is equivalent to the original model, but where the intergenerational transition matrix of skills is fixed to be equal to one obtained (endogenously) in the original steady state. This matrix defines that the distribution of children's skills depends (exclusively) on parents' skills and education group. Hence, the tax system cannot affect the development of skills directly, though it may affect it through education choices. The model with exogenous childhood skills predicts small welfare increases from increasing tax progressivity, in line with the literature that uses this kind of models to evaluate tax progressivity (e.g., Floden and Lindé, 2001; Conesa and Krueger, 2006; Heathcote, Storesletten, and Violante, 2017). On the other hand, the full model developed here predicts welfare gains of approximately 1% in consumption equivalence terms from substantially *reducing* tax progressivity. By increasing the incentives to invest towards children, a less progressive taxation can increase welfare in the long run, though this can be associated with a very costly transition. We remark that a more flexible tax function may allow for conditional transfers that increase progressivity, allowing poor parents to invest without harming incentives. We leave this for future research.

7 Conclusion

Doepke and Tertilt (2016) argue that there is a potentially large role for family economics within macroeconomics. This paper moves in that direction by combining a macroeconomic model that is appropriate for policy analysis on income inequality and intergenerational mobility, with the findings on childhood development (where family background is crucial). We use a standard macroeconomic Aiyagari-style life-cycle general-equilibrium model and introduce parental investment in the skills of children which, in turn, are later associated with endogenous education and labor outcomes. Parents build children's skills by investing both time and money during multiple periods. These skills make education easier and are also rewarded by the market.

This paper shows that underinvestment in children's development is relevant for the macroeconomic analysis of inequality and social mobility, and can be improved by government policies that target childhood directly. Introducing universal government investments towards early childhood (e.g., mandatory schools for children under the age of 4) leads to a long-run reduction in income inequality of 7% and an increase in intergenerational mobility of 34%. These changes for inequality and mobility are large enough for the US to reach Canadian or Australian levels. This policy yields long-run welfare gains (in consumption equivalence terms) of 10%.

These welfare gains are twice the ones obtained by introducing the same early childhood program as a short-run partial-equilibrium policy—similar to an RCT. Although general equilibrium and taxation effects reduce the gains by one-tenth each, the long-run change in the distribution of parental characteristics more than compensates for those reductions. Key to this welfare gain is that investing in a child not only improves her skills but also creates a better parent for the next generation. Although this suggests that these gains may take a long time to accrue, the transition dynamics analysis shows that the second generation to receive the government investments would already obtain over two-thirds of the final welfare gains.

We made several simplifying assumptions in order to keep our analysis computationally feasible. Incorporating richer family heterogeneity (e.g., endogenous marriage/divorce and fertility) would allow us to investigate additional potential effects of early childhood policies. We believe, however, that our main result of long-run effects being larger than short-run ones would remain true in models with richer family heterogeneity (as long as parents remain to be important for early childhood development). It would also be interesting to link early childhood development with college major choices, since Arcidiacono, Aucejo, and Hotz (2016) show that college majors are associated with pre-college skills. Finally, choosing optimal early childhood policies taking fully into account the transition would be particularly interesting given that welfare gains are heterogeneous by cohorts. We explored some alternatives including government borrowing and slow introduction of early childhood investments, but we believe more research is necessary.

References

ABBOTT, B. (2016): "The Effect of Parental Composition of Investments in Children," Working paper.

ABBOTT, B., G. GALLIPOLI, C. MEGHIR, AND G. L. VIOLANTE (Forthcoming): "Education Policy and Intergenerational Transfers in Equilibrium," *Journal of Political Economy*.

AGOSTINELLI, F., AND M. WISWALL (2016a): "Estimating the Technology of Children's Skill Formation," Working Paper 22442, National Bureau of Economic Research.

——— (2016b): "Identification of Dynamic Latent Factor Models: The Implications of Re-Normalization in a Model of Child Development," Working Paper 22441, National Bureau of Economic Research.

AIYAGARI, S. R., J. GREENWOOD, AND A. SESHADRI (2002): "Efficient investment in children," *Journal of Economic Theory*, 102(2), 290–321.

ARCIDIACONO, P., E. M. AUCEJO, AND V. J. HOTZ (2016): "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California," *The American Economic Review*, 106(3), 525–562.

ATTANASIO, O., E. FITZSIMONS, S. GRANTHAM-MCGREGOR, C. MEGHIR, AND M. RUBIO-CODINA (2016): "Early Childhood Stimulation, Nutrition and Development: A Randomised Control Trial," Working paper.

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2008): "Trends in U.S. Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, 90(2), 300–323.

BAKIŞ, O., B. KAYMAK, AND M. POSCHKE (2015): "Transitional dynamics and the optimal progressivity of income redistribution," *Review of Economic Dynamics*, 18(3), 679–693.

BALAND, J.-M., AND J. ROBINSON (2000): "Is Child Labor Inefficient?," *Journal of Political Economy*, 108(4), 663–679.

BARRO, R. J., AND G. S. BECKER (1989): "Fertility Choice in a Model of Economic Growth," *Econometrica*, 57(2), 481–501.

BÉNABOU, R. (2002): "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?," *Econometrica*, 70(2), 481–517.

BROWN, J., AND S. WEISBENNER (2004): "Intergenerational transfers and savings behavior," in *Perspectives on the Economics of Aging*, pp. 181–204. University of Chicago Press.

CAUCUTT, E., AND L. LOCHNER (2017): "Early and Late Human Capital Investments, Borrowing Constraints, and the Family," Working Papers 2017-040, Human Capital and Economic Opportunity Working Group.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014): "Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States," *The Quarterly Journal of Economics*, 129(4), 1553–1623.

Christakis, D. A., F. J. Zimmerman, D. L. DiGiuseppe, and C. A. McCarty (2004): "Early Television Exposure and Subsequent Attentional Problems in Children," *Pediatrics*, 113(4), 708–713.

Conesa, J. C., and D. Krueger (2006): "On the optimal progressivity of the income tax code," *Journal of Monetary Economics*, 53(7), 1425–1450.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931.

Del Boca, D., C. Flinn, and M. Wiswall (2014): "Household Choices and Child Development," *The Review of Economic Studies*, 81(1), 137–185.

Diamond, P., and E. Saez (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25(4), 165–90.

Doepke, M., and M. Tertilt (2016): *Families in Macroeconomics*vol. 2 of *Handbook of Macroeconomics*, pp. 1789–1891. Elsevier.

Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2015): "Early Childhood Education," Working Paper 21766, National Bureau of Economic Research.

Floden, M., and J. Lindé (2001): "Idiosyncratic Risk in the United States and Sweden: Is There a Role for Government Insurance?," *Review of Economic Dynamics*, 4(2), 406–437.

Flood, S., M. King, S. Ruggles, and J. R. Warren. (2015): Integrated Public Use Microdata Series, Current Population Survey: Version 4.0. [Machine-readable database]. Minneapolis: University of Minnesota.

Galor, O., and O. Moav (2004): "From Physical to Human Capital Accumulation: Inequality and the Process of Development," *The Review of Economic Studies*, 71(4), 1001–1026.

Galor, O., and J. Zeira (1993): "Income Distribution and Macroeconomics," *The Review of Economic Studies*, 60(1), 35–52.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2017): "Quantifying the Life-cycle Benefits of a Prototypical Early Childhood Program," Working Paper 23479, National Bureau of Economic Research.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. M. Chang, and S. Grantham-McGregor (2014): "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, 344(6187), 998–1001.

Heathcote, J., K. Storesletten, and G. L. Violante (2017): "Optimal Tax Progressivity: An Analytical Framework," *The Quarterly Journal of Economics.*

Heckman, J. J., L. Lochner, and C. Taber (1998): "Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents," *Review of Economic Dynamics*, 1(1), 1–58.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006): *Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond.* vol. 1 of *Handbook of the Economics of Education*, chap. 7, pp. 307–458. Elsevier.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The effects of cognitive and non-cognitive abilities on labor market outcomes and social behavior," *Journal of Labor Economics*, pp. 411–482.

Holter, H. A. (2015): "Accounting for cross-country differences in intergenerational earnings persistence: The impact of taxation and public education expenditure," *Quantitative Economics*, 6(2), 385–428.

Huggett, M., G. Ventura, and A. Yaron (2011): "Sources of Lifetime Inequality," *American Economic Review*, 101(7), 2923–54.

Katz, L., and K. M. Murphy (1992): "Changes in Relative Wages, 1963–1987: Supply and Demand Factors," *The Quarterly Journal of Economics*, 107(1), 35–78.

Keane, M. P., and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), pp. 473–522.

Krueger, D., and A. Ludwig (2016): "On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium," *Journal of Monetary Economics*, 77(C), 72–98.

Lee, S. Y., and A. Seshadri (Forthcoming): "On the Intergenerational Transmission of Economic Status," *Journal of Political Economy.*

Loury, G. C. (1981): "Intergenerational Transfers and the Distribution of Earnings," *Econometrica*, 49(4), 843–867.

Manuelli, R. E., and A. Seshadri (2009): "Explaining International Fertility Differences," *The Quarterly Journal of Economics*, 124(2), 771–807.

McDaniel, C. (2007): "Average tax rates on consumption, investment, labor and capital in the OECD 1950-2003," Discussion paper.

Meghir, C., and D. Phillips (2010): "Labour supply and taxes," *Dimensions of tax design: The Mirrlees review*, pp. 202–74.

Piketty, T., and E. Saez (2003): "Income Inequality in the United States, 1913–1998*," *The Quarterly Journal of Economics*, 118(1), 1.

Pouliot, W. (2006): "Introducing uncertainty into Baland and Robinson's model of child labour," *Journal of Development Economics*, 79(1), 264 – 272.

Restuccia, D., and C. Urrutia (2004): "Intergenerational Persistence of Earnings: The Role of Early and College Education," *The American Economic Review*, 94(5), pp. 1354–1378.

Saez, E. (2016): "Taxing the Rich More: Preliminary Evidence from the 2013 Tax Increase," Working Paper 22798, National Bureau of Economic Research.

Samuelsson, I. P., and M. A. Carlsson (2008): "The Playing Learning Child: Towards a pedagogy of early childhood," *Scandinavian Journal of Educational Research*, 52(6), 623–641.

Swing, E. L., D. A. Gentile, C. A. Anderson, and D. A. Walsh (2010): "Television and Video Game Exposure and the Development of Attention Problems," *Pediatrics*, 126(2), 214–221.

Thompson, R., and C. Nelson (2001): "Developmental Science and the Media: Early Brain Development," *The American psychologist*, 56, 5–15.

Todd, P. E., and K. I. Wolpin (2003): "On the Specification and Estimation of the Production Function for Cognitive Achievement*," *The Economic Journal*, 113(485), F3–F33.

Yum, M. (2018): "Parental Time Investment and Intergenerational Mobility," Working paper.

A Empirical Findings: Details

A.1 Child Development Supplement:

The results presented in Section 4 are for the whole sample of children born to at least high-school educated mothers for which we have data on their time diaries. The summary statistics for these children are presented in Table A1.

Table A1: CDS Summary Statistics: Whole Sample

Age Group	**3-7**	**8-12**	**13-18**
Letter-Word Score	16.4	41.0	48.1
	(10.8)	(7.8)	(5.9)
Applied-Problems Score	16.7	34.0	41.2
	(7.8)	(6.2)	(6.8)
Child's Age	5.4	10.2	15.1
Mother's Age	32.3	37.4	42.3
Father's Age	35.4	40.3	45.1
Mother's Education (years)	14.2	14.1	14.1
Father's Educarion (years)	13.9	13.9	13.8
Mother Works	76.7	83.6	84.8
Father Works	70.3	62.5	63.5
Mother's Work Hours (weekly)	23.2	27.3	29.2
Father's Work Hours (weekly)	30.9	27.4	27.5
Mother's Hourly Wage	18.0	18.2	17.9
Father's Hourly Wage	25.3	27.7	28.5
Family's Total Income	1391.1	1433.5	1712.1
Number of Children	1312	2089	1810

Table shows summary statistics for sample used to study parental investments in Section 4.

As expected test scores grow with the age of the children. Moreover, we see that as children grow older it is more likely that the mother works. We now present the relevant differences when we focus on families with two parents and two children, the closest to our model agents.

Sample Selection: We start with all the children born to at least high-school educated mothers for which we can observe the variables of interest from the Child Development Supplement, i.e., 2,778 children. Given that in our model there are no fertility choices, we restrict the sample to families with two children, reducing the number of children in the sample to 1,423. Similarly, we focus on stable two-parent families since our model has no marriage choices. Restricting the sample to children whose parents remain married between their time of birth and age 12, reduces the sample to 701 children. Finally, we keep only biological children and children whose parents are at least high-school graduates. This leads to a final number of children of 667.

The summary statistics of this sample are shown in Table A2. It is seen that this sample is rather high-income and highly educated relative to a sample that includes families with more children and are not married. All fathers work for at least some period when children are young. Interestingly, while early on all fathers are working, this share goes down as the share of mothers working goes up. Even though this sample is more representative of a stable high socioeconomic group, it is shown in Section 4 that differences on parental investments are still sizable. Differences in parental investments including single parent households would also be interesting, but are out of the scope of the model used here.

Table A2: CDS Summary Statistics: Selected Sample

Age Group	**3-7**	**8-12**	**13-18**
Letter-Word Score	18.6	43.40	50.5
	(11.3)	(6.7)	(3.9)
Applied-Problems Score	18.6	36.1	44.3
	(7.1)	(6.1)	(5.7)
Child's Age	5.4	10.2	15.0
Mother's Age	34.9	39.3	44.2
Father's Age	37.0	41.4	46.2
Mother's Education (years)	14.9	14.8	14.8
Father's Educarion (years)	14.6	14.5	14.5
Mother Works	72.5	81.2	88.6
Father Works	100.0	96.4	90.6
Mother's Work Hours (weekly)	21.9	24.3	29.5
Father's Work Hours (weekly)	44.0	42.5	39.7
Mother's Hourly Wage	20.4	19.7	20.9
Father's Hourly Wage	27.9	32.8	32.1
Family's Total Income	1890.0	2222.8	2763.6
Number of Children	265	445	394

Table shows summary statistics for (most selective) sample used to study parental investments in Section 4.

A.2 Income Profile: PSID Sample Selection

We start with observations of individuals between the ages of 25 and 63. After dropping observations of wages below half the minimum wage or inconsistent information on hours and income, we have an initial number of individuals in the sample equal to 22,052. The distribution across education groups within this starting point is:

- High-school dropouts: 3,649 Individuals (40% from SRC), with a total number of observations of 32,495.
- High-school graduates: 13,091 Individuals (53% from SRC), with a total number of observations of 104,595.

- College graduates: 5,001 Individuals (75% from SRC), with a total number of observations of 44,704.

Restricting to households with two adults, the number of individuals is reduced to 17,324. We further restrict observations to those with positive hours of labor in the household (but lower than 10,000 annually). We also drop individuals who at least once report hourly wages under $1 or above $400. This reduces the number of individuals to 16,563. Finally, we keep individuals with at least 8 observations of income and who do not report extreme changes of income (i.e., annual growth above 400%, or reduction by 66%). This leads to a final number of individuals of 5,350. These are distributed as follows:

- High-school dropouts: 877 Individuals (40% from SRC), with a total number of observations of 4,294.
- High-school graduates: 3,034 Individuals (61% from SRC), with a total number of observations of 17,147.
- College graduates: 1,431 Individuals (80% from SRC), with a total number of observations of 8,880.

A.3 Income Profile: NLSY Sample Selection

We start with 12,686 individuals, with a total of 317,150 observations. We exclude observations in the army, and restrict to those between the ages of 25 and 63. This reduces the number of individuals to 12,683 (217,570 observations). We drop observations with top-coded earnings, and drop individuals who change education groups (after age 25) or who have missing information on their AFQT score. This reduces the number of individuals to 11,213 (191,301). We further restrict observations to those with positive hours of labor in the household (but lower than 10,000 annually). We also drop individuals who at least once report hourly wages under half the minimum wage or above $400. We keep individuals with at least 8 observations of income. This reduces the number of individuals to 6,729 (94,727 observations). After grouping observations in 4 year periods (like the model), we eliminate observations with wages above $400 and who do not report extreme changes of income (i.e., annual growth above 400%, or reduction by 66%). This leads to a number of individuals of 6,694. Restricting to households with two adults leads to a final number of individuals of 5,607. These are distributed as follows:

- High-school dropouts: 554 Individuals, with a total number of observations of 2,350.
- High-school graduates: 3,574 Individuals, with a total number of observations of 16,960.
- College graduates: 1,479 Individuals, with a total number of observations of 7,552.

A.4 Additional Figures and Tables

Table A3: Age profile of wages by education group

VARIABLES	(1) HS Grad	(2) College
Age	0.0333***	0.0716***
	(0.00273)	(0.00431)
Age^2	-0.000326***	-0.000740***
	(3.33e-05)	(5.19e-05)
Over Sample	-0.0260***	-0.0406**
	(0.00968)	(0.0189)
Constant	2.066***	1.633***
	(0.0578)	(0.0911)
Observations	17,071	8,851
R-squared	0.114	0.136
# of households	3031	1430

*Robust standard errors in parentheses. *, **, *** denote statistical significance at the 10, 5, and 1 percent, respectively. Source: PSID. Methodology is explained in the main text.*

B Stationary Equilibrium

We introduce some notation to define the equilibrium more easily. Let $s_j \in S_j$ be the age-specific state vector of an individual of age j, as defined by the recursive representation of the individual's problems in Section 3. Let the Borel sigma-algebras defined over those state spaces be $\mu = \{\mu_j\}$. Then, a stationary recursive competitive equilibrium for this economy is a collection of: (i) decision rules for education $\{d^e(s_{J_e})\}$, consumption, labor supply, and assets holdings $\left\{c_j(s_j), h_j(s_j), a'_j(s_j)\right\}$, parental time and money investments $\left\{d_j^{\zeta}(s_j), d_j^m(s_j)\right\}$, and parental transfers $\{\varphi(s_j)\}$; value functions $\left\{V_j(s_j), V_j^s(s_j), V^{sw}(s_j)\right\}$; (iii) aggregate capital and labor inputs $\{K, H_1, H_2\}$; (iv) prices $\{r, w^1, w^2\}$; (v) tax policy $\{\tau_c, \tau_y, \tau_k, \omega\}$; and (vi) a vector of measures μsuch that:

1. Given prices, decision rules solve the respective household problems and $\left\{V_j(s_j), V_j^s(s_j), V^{sw}(s_j)\right\}$ are the associated value functions.

2. Given prices, aggregate capital and labor inputs solve the representative firm's problem, i.e. it equates marginal products to prices.

3. Labor market for each education level clears.

For high-school level:

$$H_1 = \sum_{j=J_e}^{J_r} \int_{S_j} \varepsilon_j^1(\theta)\, h_j\left(s_j | e=1\right) d\mu_j + \sum_{j=J_e}^{J_e} \int_{S_j} \varepsilon_j^e(\theta)\, h_j\left(s_j | e=2\right) d\mu_j$$

where the first summation is the supply of high-school graduates while the second is that labor supply of college students.

For college level:

$$H_2 = \sum_{j=J_e+1}^{J_r} \int_{S_j} \varepsilon_j^2(\theta)\, h_j\left(s_j | e=2\right) d\mu_j.$$

4. Asset market clears

$$K = \sum_{j=J_e}^{J_d} \int_{S_j} a_j\left(s_j\right) d\mu_j.$$

5. Good market clears:

$$\sum_{j=J_e}^{J_d} \int_{S_j} c_j\left(s_j\right) d\mu_j + \delta K + \int_{S_{J_e}} p_e 1\left\{d_{J_e}^e\left(s_{J_e}\right) = 2\right\} d\mu_{J_e} + \sum_{j=J_f}^{J_f+1} \int_{S_j} m_j\left(s_j\right) d\mu_j = F(K, H)$$

where the last two term on the left hand side represent the expenditures on education and childhood development, respectively.

6. Government budget holds with equality

$$\sum_{j=J_r+1}^{J_d} \int_{S_j} \pi(\theta, e)\, d\mu_j + G = \sum_{j=J_e}^{J_r} \int_{S_j} T\left(y\left(s_j\right), k\left(s_j\right), c\left(s_j\right)\right) d\mu_j.$$

Government expenditures on retirement benefits and G equal net revenues from taxes—which include the lump-sum transfer.

7. Individual and aggregate behaviors are consistent: measures μ is a fixed point of $\mu(S) = Q(S, \mu)$ where $Q(S, \cdot)$ is transition function generated by decision rules and exogenous laws of motion, and S is the generic subset of the Borel-sigma algebra defined over the state space.

C Estimation: Details

C.1 Child Skill Production Function

Table C4: Child Skill Production Function: estimates from Cunha, Heckman, and Schennach (2010)

	Cognitive Skills		Non-Cognitive Skills	
	1st Stage	2nd Stage	1st Stage	2nd Stage
Current Cognitive Skills	0.479	0.831	0.000	0.000
	(0.026)	(0.011)	(0.026)	(0.010)
Current Non-Cognitive Skills	0.070	0.001	0.585	0.816
	(0.024)	(0.005)	(0.032)	(0.013)
Investments	0.161	0.044	0.065	0.051
	(0.015)	(0.006)	(0.021)	(0.006)
Parent's Cognitive Skills	0.031	0.073	0.017	0.000
	(0.013)	(0.008)	(0.013)	(0.008)
Parent's Non-Cognitive Skills	0.258	0.051	0.333	0.133
	(0.029)	(0.014)	(0.034)	(0.017)
Complementarity parameter	0.313	-1.243	-0.610	-0.551
	(0.134)	(0.125)	(0.215)	(0.169)
Variance of Shocks	0.176	0.087	0.222	0.101
	(0.007)	(0.003)	(0.013)	(0.004)

Standard errors in parentheses.

C.2 Replacement benefits: US Social Security System

The pension replacement rate is obtained from the Old Age Insurance of the US Social Security System. We use education level as well as the skill level to estimate the average lifetime income, on which the replacement benefit is based. We estimate the average life time income to be $\widehat{y}_j(\theta, e) = w^e \varepsilon_j^e(\theta, \overline{\eta}) \times \overline{h}$ with $\overline{\eta}$ and $\bar{h}$ referring to the average efficiency and hours worked. Then averaging over j, mean income $\widehat{y}$ is calculated and used in (8) to obtain the replacement benefits.

The pension formula is given by

$$\pi(\theta, e) = \begin{cases} 0.9\widehat{y}(\theta, e) & \text{if } \widehat{y}(\theta, e) \leq 0.3\bar{y} \\ 0.9(0.3\bar{y}) + 0.32(\widehat{y}(\theta, e) - 0.3\bar{y}) & \text{if } 0.3\bar{y} \leq \widehat{y}(\theta, e) \leq 2\bar{y} \\ 0.9(0.3\bar{y}) + 0.32(2 - 0.3)\bar{y} + 0.15(\widehat{y}(\theta, e) - 2\bar{y}) & \text{if } 2\bar{y} \leq \widehat{y}(\theta, e) \leq 4.1\bar{y} \\ 0.9(0.3\bar{y}) + 0.32(2 - 0.3)\bar{y} + 0.15(4.1 - 2)\bar{y} & \text{if } 4.1\bar{y} \leq \widehat{y}(\theta, e) \end{cases} \tag{8}$$

where $\bar{y}$ is approximately \$240,000 (\$70,000 annually).

C.3 Aggregate Production Function

Wages from PSID

In order to estimate the APF we need to first estimate the wage for each year and education group. For this, we return to the PSID data and remove the the age profile. We use first difference in order to remove the effect of ability. Then, we estimate wage growth for each year by running a fixed effect regression for each year. Normalizing wages in the year 2000 (taking into account average ability from NLSY for each education group) we can now obtain the wages for each year and education group.

APF estimation using CPS

The last part of the estimation is done using CPS since the sample is larger and representative of the cross-section in each year. We restrict the sample to include only salary workers between the ages of 20 and 60 with properly reported education groups. For each year we then calculate the total wage bill ω of each education group (high-school and college graduates) and use the PSID estimated wages to obtain the efficiency units of labor H of each group.

We assume that the production function is the following:

$$\begin{aligned} Y_t &= K_t^{\alpha} H_t^{1-\alpha} \\ H_t &= \left[s H_{1t}^{\rho} + (1-s) H_{2t}^{\rho} \right]^{1/\rho} \end{aligned}$$

We can then estimate the parameters s and ρ using the following equation:

$$log\left(\frac{\omega_{2t}}{\omega_{1t}}\right) = log\left(\frac{1-s}{s}\right) + \rho log\left(\frac{H_{2t}}{H_{1t}^{1}}\right)$$

We can estimate this using OLS or First Differences. Moreover, we also do IV using lags as instruments. This approach leads to estimates around 0.3 for ρ, and 0.5 for s—in line with the estimates from Katz and Murphy (1992) and Heckman, Lochner, and Taber (1998).

C.4 Simulated Method of Moments: Moments' Selection

We internally estimate $P = 13$ parameters in order to match P moments. Although the model is highly nonlinear, so that (almost) all parameters affect all outcomes, the identification of some parameters relies on some key moments in the data. Figure C1 shows the result of the following identification exercise. First, given an hypercube of the parameter space, we draw 100,000 candidate parameter vectors from uniform Sobol (quasi-random) points, and compute the implied moments in the model. Second, for each parameter we associate a relevant target moment. Third, for each parameter, we divide the vector of this particular parameter in 50 quantiles and compute the 25th, 50th, and 75th percentiles of the associated moment in each quantile.[48] Finally, we show these percentiles of the moment along with the value in the data. We claim that a moment is important for a parameter's identification if, as we move across quantiles, the percentiles of the associated moment change and cross the horizontal dashed line (i.e., the value of that moment in the data). The slope of each curve shows how important is that parameter for the associated moment (a steeper curve implies the moment is more informative). The difference between the 25th and 75th percentiles informs about the relative importance of the remaining parameters (other parameters are more important when the 75th and 25th percentiles are further apart).

The success of this exercise relies on finding a relevant moment for each parameter. For example, the data on transfers to children, hours worked, and hours with child identify the preference parameters related to altruism, disutility of work, and disutility of time with children, as shown by the first row of Figure C1. More precisely, there is a positive relation between the level of altruism (λ_n) and transfers to children. As parents value more their children (higher λ), they increase the transfers to them. Similarly, there is a negative association between the disutility of work (μ) and average hours worked. When $\xi = 0$, the average number of hours with children converges to the maximum allowed in the solution grid (i.e., 35 hours).

The rest of the figures can be interpreted in similar ways. The only moment that seems to be affected by (substantially) more than the parameter selected is the money-time correlation. The money-time substitutability parameter γ is important for this moment but so seem to be other parameters, as suggested by the wide gap between the 75th and 25th percentiles. This is due to the fact that when ξ approaches zero all parents put the same amount of time (i.e., the maximum) or when the share of money α_m approaches one time with children is reduced to very similar amounts by all parents. This also leads to a relatively large standard deviation for γ as shown in Table 4, but as we show in our robustness analysis in Section 6.2, results were almost unchanged by moving γ within such interval.

[48] Notice that for each quantile there are $P-1$ parameters that are randomly drawn from the uniform Sobol points, and, therefore, potentially far away from the estimated parameter value.

Figure C1: Identification

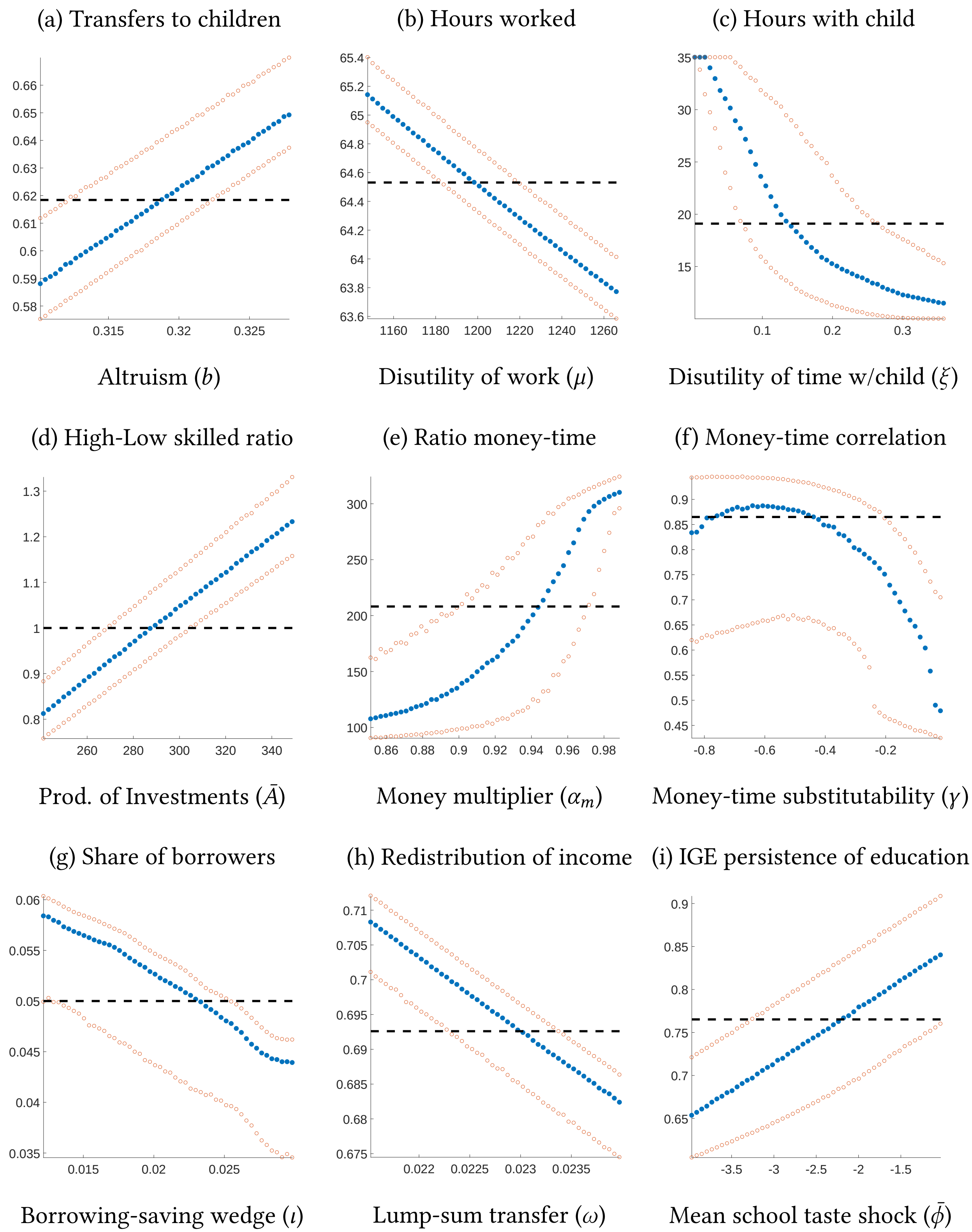
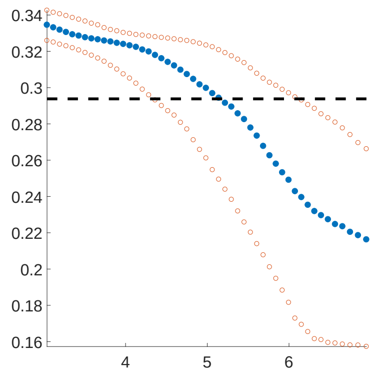


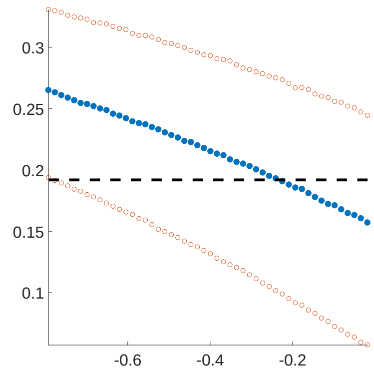
Figure C1 (cont.): Identification

(j) Share of college grads (%)



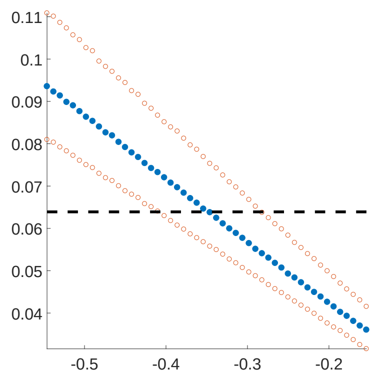
Mean school taste (α)

(k) College: cog skills slope



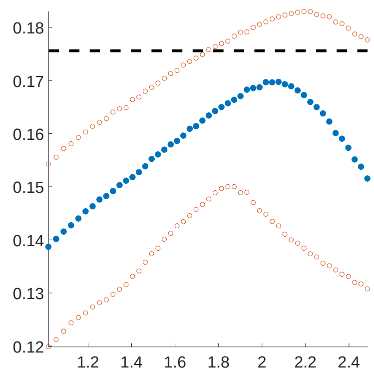
School taste-cog skill relation (α_c)

(l) College: noncog skills slope



School taste-noncog skill relation (α_{nc})

(m) College: residual variance



SD of taste shock (σ_ϕ)

Note: For each parameter's quantile, the (filled) blue dot shows the median while the (empty) red dots show the 25th and 75th percentiles of the assigned moment. The black dashed line shows the value of the moment in the data. Transfers to children are estimated as a share of income. Redistribution of income refers to the ratio of the variances of log-income after taxes and before taxes. Methodology is explained in the main text.

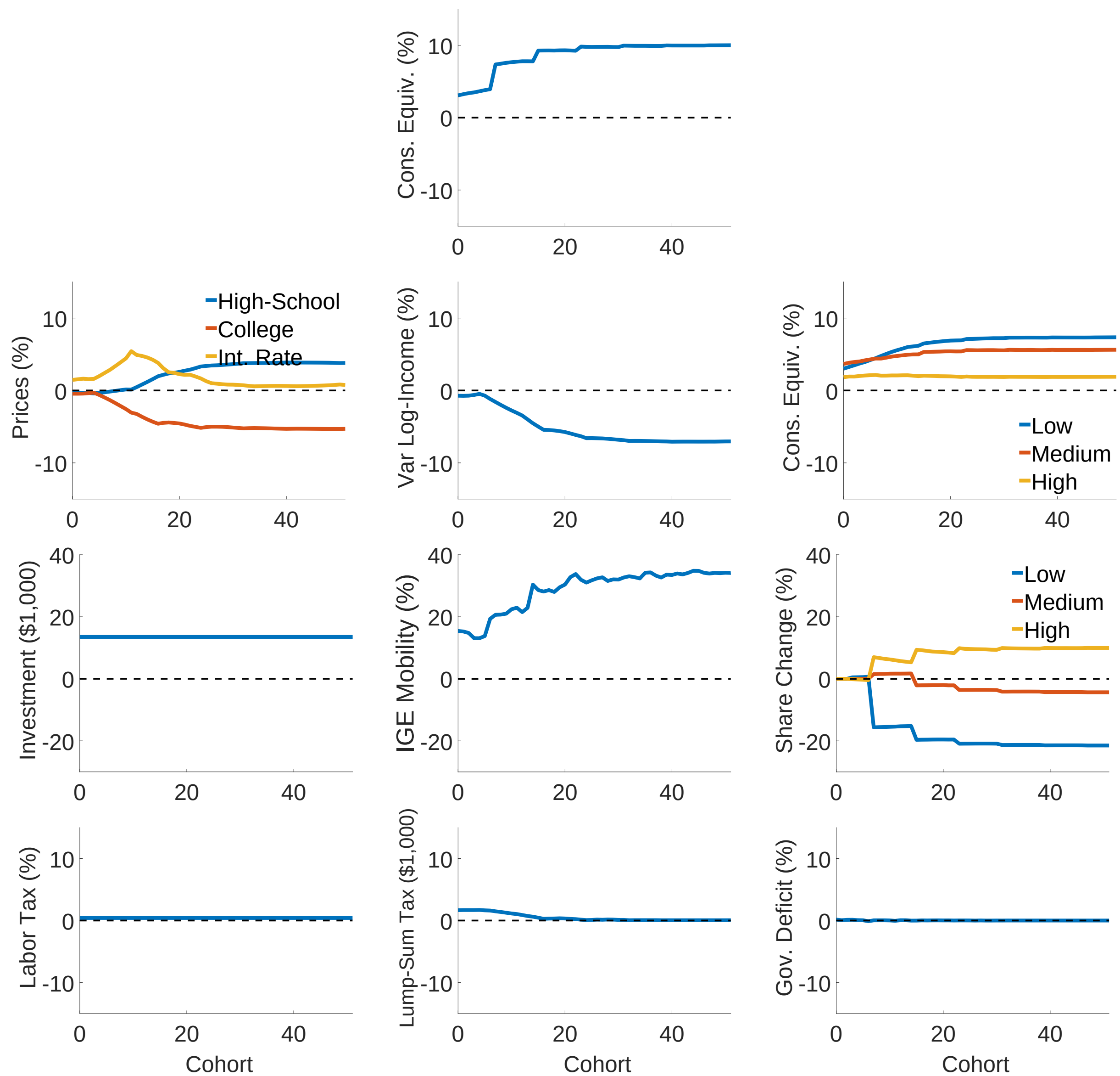
D Transition Details

Section 6 presents the main results for the case in which the economy transitions to the new steady state by introducing an extra lump-sum tax that balances the government's budget every period. Here we present more details on such transition as well as explore other ways to finance the transition that can lead to smaller welfare losses for older cohorts.

D.1 Government budget is balanced every period

Figure 11 showed the main results regarding welfare, inequality and mobility for the case in which the economy transitions to the new steady state by introducing an extra lump-sum tax that balances the government's budget every period. Figure D2 expands that analysis by also including information on price and tax changes as well as welfare changes for children born to different socioeconomic groups.

Figure D2: Transition dynamics: more details of balanced budget case



Notes: The policy (including the investments and labor income tax change) is introduced unexpectedly. We compute the transition introducing a lump-sum tax such that the government's budget balances every period. Consumption equivalence is shown for a newborn from the cohort defined by the horizontal axis. Cohort 0 is the first cohort to receive the government investments. Intergenerational mobility refers to minus the regression coefficient between children's and parents' income ranks. It is calculated for the generation born in such cohort and their parents. The consumption equivalence is also reported for children born to parents with different levels of cognitive skills. All values are relative to the initial steady state.

D.2 Other Alternatives

Figure 12, in the main text, shows that introducing the early childhood investment policy and forcing the government to keep its budget balanced every period leads to negative welfare effects on the older individuals at the time of the introduction. These individuals have to pay higher taxes, but most of the gains are obtained by later cohorts (who are subject to smaller tax increases). Here we evaluate two alternatives that transfer the cost of the policy to later cohorts to study whether, if the government is able to borrow temporarily, permanent government investments in early childhood can be welfare improving for every cohort (on average). We focus on the case in which the government is able to borrow at an annual 3% rate.[49]

The first alternative imposes that only intervened cohorts have to pay higher taxes. We find that this form of government borrowing alone is not enough to achieve welfare gains for most cohorts—at least at an interest rate of 3%. The second alternative adds a slower introduction of the government investments to the first alternative. Government investments start at $1,000 per child-year for the first cohort and grow by $1,000 for every new cohort, until they reach the target of $13,500. We find that this slow introduction, combined with the fact that only intervened cohorts pay higher taxes, leads to welfare gains for all new cohorts and most individuals alive at the time of the introduction.

D.2.1 Only Intervened Cohorts Pay

Figure D3 shows the transition dynamics to the baseline policy in which the government invests $13,500 per child-year, when only intervened cohorts have to pay higher taxes. To compensate for the smaller early increase increase in taxes, the government is allowed to borrow at an interest of 3%. But it has to use the later higher taxes to pay off its debt by the time cohort 50 is born.[50] We assume that higher taxes are introduced only after cohort 16 is born, as a way of reducing the costs even further to the earlier cohorts (since they accrue less gains than later cohorts).

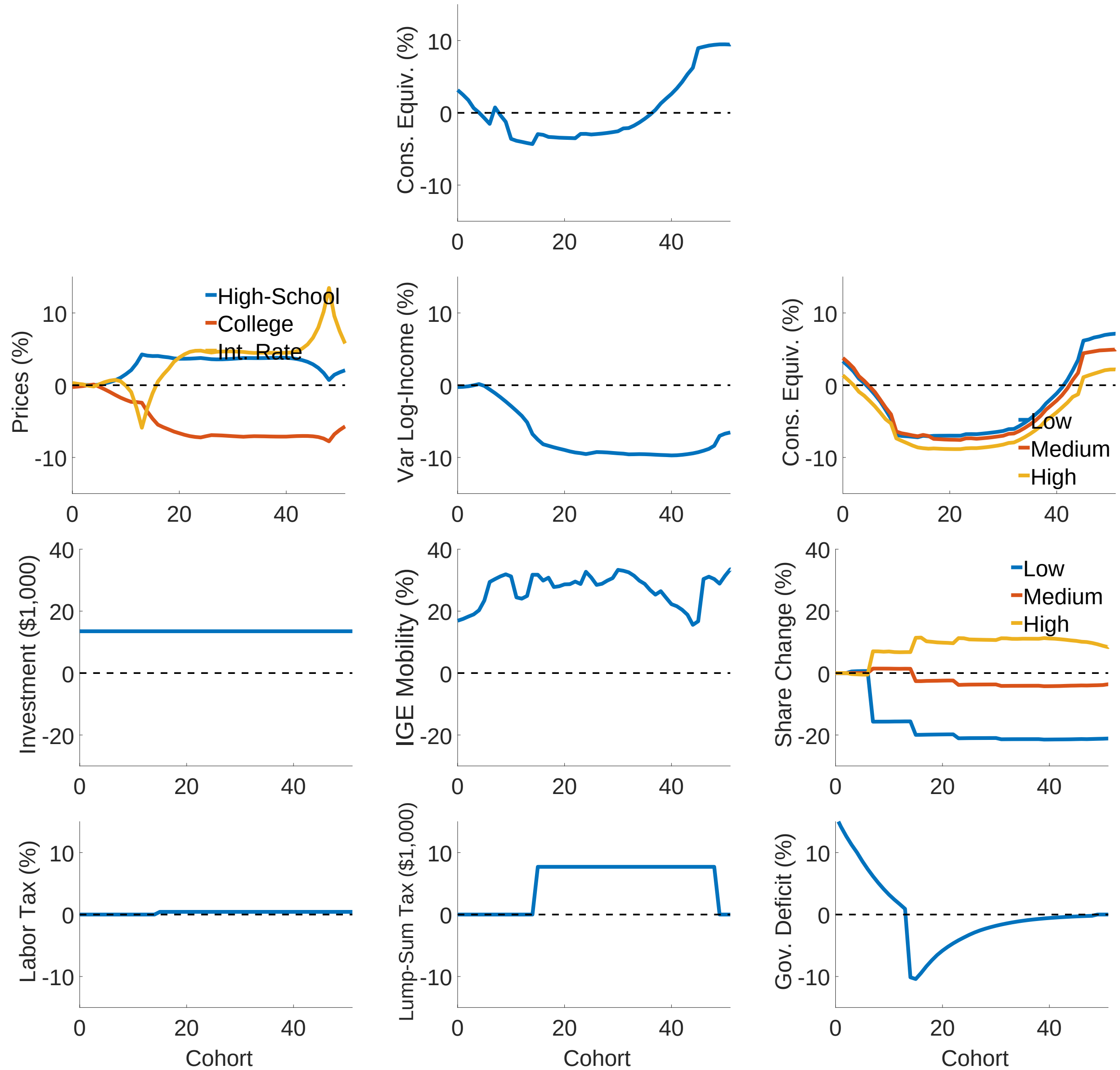
Figure D4 shows that this form of government borrowing alone is not enough to guarantee welfare gains for most cohorts—at least at an interest rate of 3%. Even though the first few cohorts born after the policy is introduced do have welfare gains (particularly because of the way taxes are introduced), later cohorts suffer welfare losses since they are forced to pay off large amounts of debts. Only after 40 cohorts are born do we observe welfare gains once again.[51]

[49] Smaller interest rates would make the policy easier to be welfare improving. 3% is likely to be on the upper bound of the rate at which the US government is able to borrow, so we can interpret this analysis as a lower bound on the welfare gains that can be achieved if the government uses its borrowing capacity. We limit to foreign borrowing here, i.e., government borrowing does not require funds provided by the agents in the model. Requiring the government to borrow locally is not theoretically difficult but would require an extra convergence step in the simulation.

[50] It is possible to allow for different interest rates and times in which the debt has to be repaid. Results are qualitatively similar, but smaller interest rates make the policy easier to afford. Longer times to full debt-repayment imply that earlier cohorts are better off but later cohorts are worse off.

[51] These results clearly depend on the assumptions of times of repayments and interest rates. However, they do show

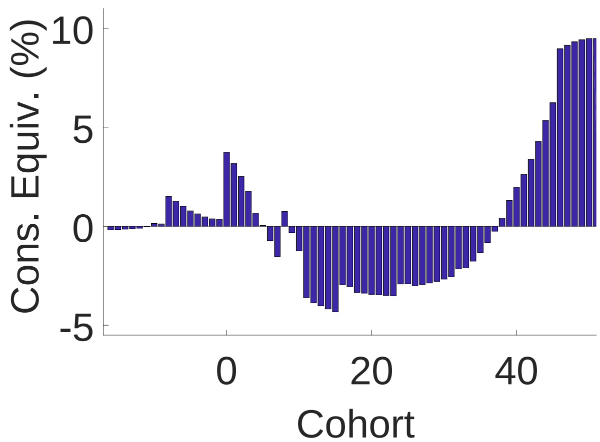
Figure D3: Transition dynamics: only intervened pay



Notes: The policy (including the investments and labor income tax change) is introduced unexpectedly but only intervened cohorts pay extra taxes. We compute the transition introducing a lump-sum tax paid only by intervened cohorts up to the time cohort 50 is born, such that the government's budget balances over the transition (assuming an interest rate of 3% annually). Consumption equivalence is shown for a newborn from the cohort defined by the horizontal axis. Cohort 0 is the first cohort to receive the government investments. Intergenerational mobility refers to minus the regression coefficient between children's and parents' income ranks. It is calculated for the generation born in such cohort and their parents. The consumption equivalence is also reported for children born to parents with different levels of cognitive skills. All values are relative to the initial steady state.

that government borrowing may not be enough.

Figure D4: Welfare gains including older cohorts



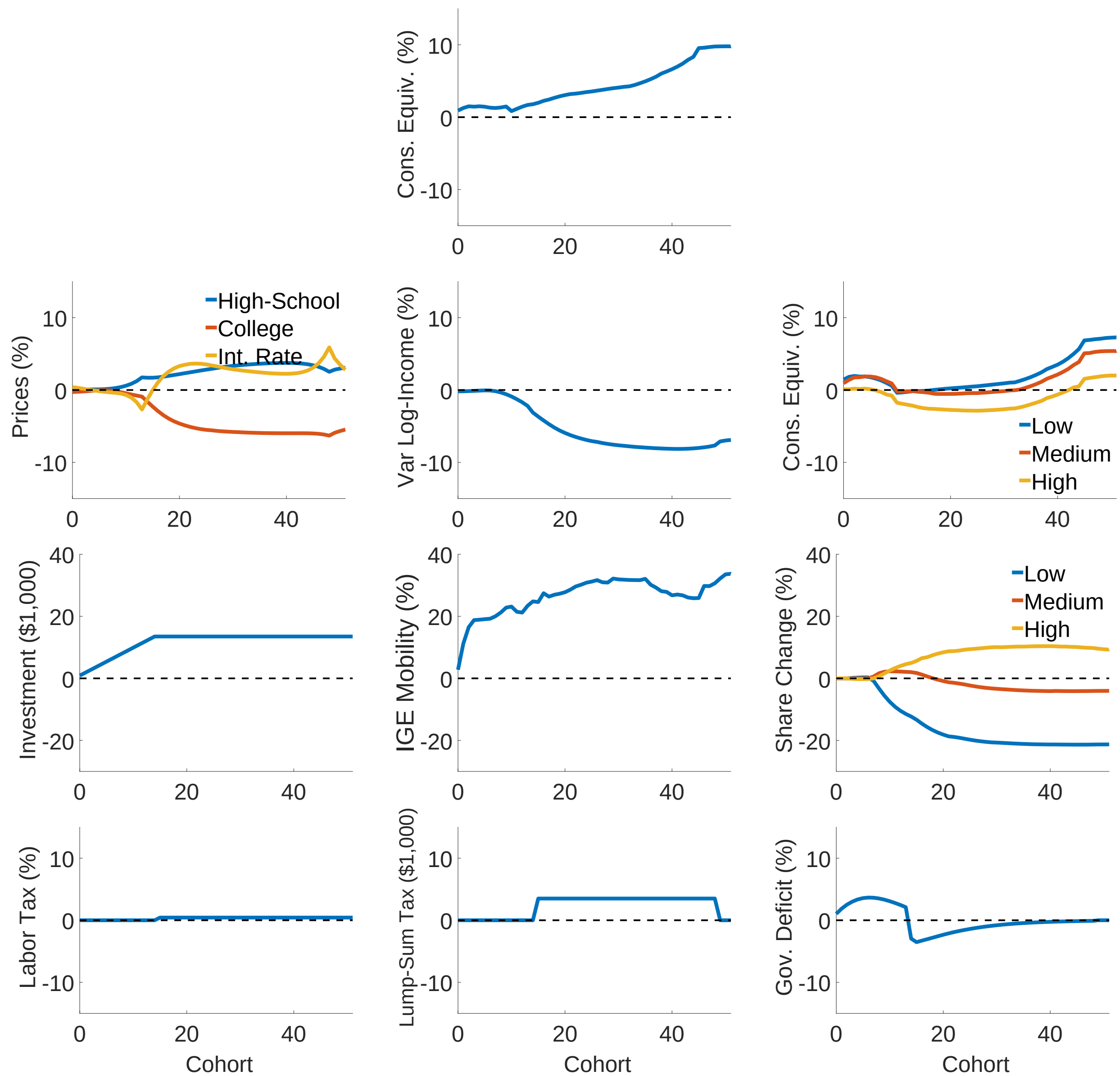
Notes: Welfare gains are reported for cohorts born after the policy is introduced (i.e., cohorts from 0 on) as well as for cohorts already alive at such time (i.e., cohorts less than 0). For the first group, welfare gains are computed for newborns. For the cohorts already alive at the time the policy is introduced, welfare gains are computed for agents with the appropriate age. For example, cohort -10 was born 40 years before the policy is introduced, so its welfare gains are computed according agents of age 40 at the time.

D.2.2 Only Intervened Cohorts Pay + Slow Introduction

Next we explore adding a slow introduction of government investments to the previous transition framework. Government investments start at $1,000 per child-year for the first cohort and grow by $1,000 for every new cohort until they reach the target of $13,500. Just like before, we assume that higher taxes are introduced only after cohort 16 is born and the government is allowed to borrow at an interest of 3%.

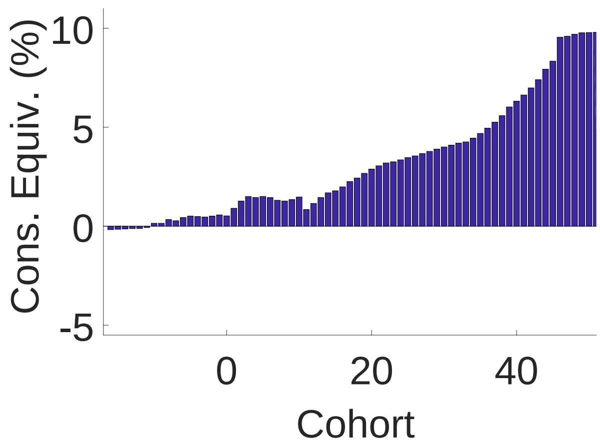
Figures D5 and D6 show that this combination of slow introduction and higher taxes only for intervened cohorts is able to achieve welfare gains for most cohorts. All new cohorts accrue welfare gains. And so do most cohorts alive at the time of the introduction. Only cohorts for whom their children have already grown up (and are not included in their utility functions) obtain welfare losses due to the changes in prices. These losses, however, are small and could be easily compensated using an age-dependent lump-sum transfer.

Figure D5: Transition dynamics: slow introduction + only intervened pay



Notes: The policy (including the investments and labor income tax change) is introduced unexpectedly but only intervened cohorts pay extra taxes. Government investments are introduced slowly, they start at $1,000 per child-year for the first cohort and grow by $1,000 for every new cohort until they reach the target of $13,500. We compute the transition introducing a lump-sum tax paid only by intervened cohorts up to the time cohort 50 is born, such that the government's budget balances over the transition (assuming an interest rate of 3% annually). Consumption equivalence is shown for a newborn from the cohort defined by the horizontal axis. Cohort 0 is the first cohort to receive the government investments. Intergenerational mobility refers to minus the regression coefficient between children's and parents' income ranks. It is calculated for the generation born in such cohort and their parents. The consumption equivalence is also reported for children born to parents with different levels of cognitive skills. All values are relative to the initial steady state.

Figure D6: Welfare gains including older cohorts



Notes: Welfare gains are reported for cohorts born after the policy is introduced (i.e., cohorts from 0 on) as well as for cohorts already alive at such time (i.e., cohorts less than 0). For the first group, welfare gains are computed for newborns. For the cohorts already alive at the time the policy is introduced, welfare gains are computed for agents with the appropriate age. For example, cohort -10 was born 40 years before the policy is introduced, so its welfare gains are computed according agents of age 40 at the time.

E Extension Estimation: With Early Childhood Education Production Function

The model with early childhood development production function, described in section 6.3, is re-estimated to match the same set of moments from the full model. The estimated parameters and moments are shown in Table E5.

Table E5: Extension estimation: with early childhood education production function

Parameter	Value	Description	Moment	Data	Model
Preferences					
μ	1016	Mean labor disutility	Avg. hours worked	64.5	67.0
b	0.33	Altruism	Parent-to-child transfer as share of income	0.62	0.64
School Taste:					
α	5.93	Avg. taste for college	College share	29	30
α_{θ_c}	-0.35	College taste and cog. skills relation	College: cog skills slope	0.19	0.19
$\alpha_{\theta_{nc}}$	-0.26	College taste and noncog. skills relation	College: noncog skills slope	0.06	0.06
σ_ϕ	2.08	SD of college taste shock	College: residual variance	0.18	0.18
$\bar{\phi}$	-4.96	Draw of school taste: mean by parent's education	Intergenerational persistence of education	0.77	0.76
Skill Formation Productivity:					
ξ	0.18	Parental time disutility of time with children	Avg. hours with children	19.1	17.8
$\bar{A}$	92.7	Returns to investments	Average skill ratio	1.0	1.0
α_m	0.96	Money productivity	Ratio of money to hours	208	190
γ	-0.21	Money-time substitutability	Money-time correlation	0.88	0.89
Interest rate					
ι ($\times 10^2$)	1.61	Borrow-save wedge	Share of borrowers	5.0	5.2
Government					
ω ($\times 10^2$)	2.33	Lump-sum transfer	Income variance ratio: Disposable to pre-gov	0.69	0.70

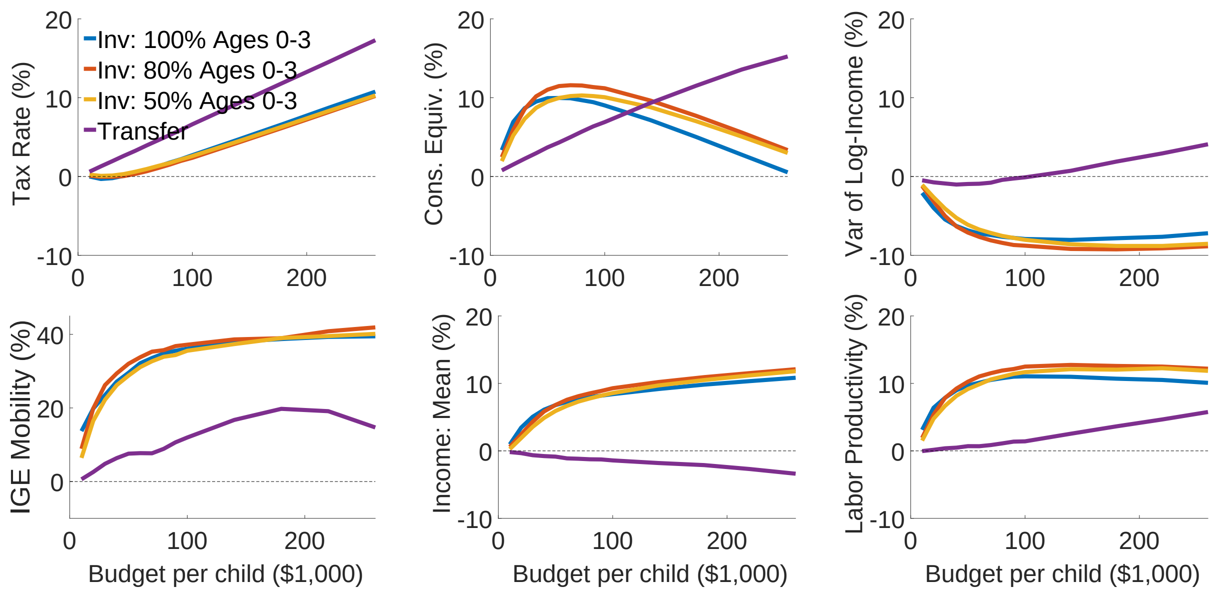
Notes: Parent-to-child transfers, hours worked, skill formation moments and intergenerational persistence of education are estimated from PSID-CDS data. Share of borrowers is estimated from Survey of Consumer Finances. College share, college-skills slope and college residual variance are estimated using NLSY. Bootstrap standard deviations in parentheses. All moments matter for all parameters, but each line highlights the moments that is particularly informative for the corresponding parameter—as explained intuitively in the main text and shown more formally in Appendix C.4.

F Other Policies

F.1 Early Childhood Investments: More Alternatives

We evaluate additional alternatives to the policy evaluated in the Section 6. Instead of using all resources to invest in children ages 0–3, we evaluate here alternatives that use part of those resources to also invest in older children (age 4–7). Moreover, we also evaluate using the same amount of resources to fund a pure government transfer program that provides a lump-sum transfer to all individuals at the age of 16. Figure F7 shows the results of these policies in the new steady-state, taking into account general equilibrium effects as well as adjusting the labor income tax such that the government's budget remains balanced. There are two main results.

Figure F7: Childhood investments



Notes: We simulate policies in which the government uses different amount of resources either to invest directly in the development of children or to fund a government transfer to all individuals (at the age of 16). Among childhood investment policies, we also evaluate different alternatives: (i) use all resources in children age 0–3 (as in the main text); (ii) use 80% of the resources for children 0–3 and 20% for children 4–7; and (iii) use 50% of the resources for children 0–3 and 50% for children 4–7. We evaluate these policies for different amounts of resources available. The horizontal axis refers to the investments per child. Then, for example, 40 refers to $40,000 per child, which equals $10,000 per child-year if using all resources for investments in children age 0–3. Outcomes are reported in changes from the baseline steady state. Consumption equivalence is determined by newborns under the veil of ignorance. Inequality refers to the variance of log-labor-income while IGE mobility refers to minus the regression coefficient between children's and parents' income ranks.

First, there are small gains from allocating part of the resources towards investing in older children. Using 20% of the resources to invest in children ages 4–7 and 80% for children ages 0–3, it is possible to obtain larger welfare gains than if investing all resources in children ages 0–3. However, the welfare gains differences are small (only 1 percentage point) and so are differences in other outcomes (inequality, mobility, and average income). Although not shown here, we find that gains are larger if a larger share of resources is used in children ages 0–3 than if a larger share is used on older children. Earlier investments lead to larger gains because the child skill production function implies that skills are more malleable at younger ages.

Second, agents typically prefer resources to be used for childhood investments rather than to fund a government transfer. This happens because the government can do something that these agents cannot do by themselves, i.e., invest in their childhood. Funding a transfer program provides less welfare than using the same resources for childhood investments as long as the resources used are not too large. Once resources are over $120,000 per child, returns on those investments are small relative to the large cost of raising taxes to afford them.

F.2 Parenting Education

Introducing parenting education is not trivial so we explain it in detail here. In these programs, parents are educated on techniques that promote children's development—including recommendations on reading, games, and ways to interact with children. We implement this in the model using θ_{pe} an extended version of children's development function 6

$$\theta_k' = \left[\alpha_{1j}\theta_k^{\rho_j} + \alpha_{2j}\max\left\{\theta, \theta_{PE}\right\}^{\rho_j} + \alpha_{3j}I^{\rho_j}\right]^{1/\rho_j} e^{\nu}, \quad \nu \sim N(0, \sigma_{j,\nu}) \tag{9}$$
$$I = \bar{A}\left[\alpha_m m^{\gamma} + (1-\alpha_m)\tau^{\gamma}\right]^{1/\gamma},$$

which increases the productivity of parental investments I. The parenting education program can be thought of as providing a minimum training on parenting techniques, which is most helpful for lower skilled parents. Alternatively, if parenting education were useful for everyone (e.g., if θ_{PE} entered as a perfect substitute for θ) we would expect general welfare benefits to be even larger. Therefore, our results here may be thought as a lower bound in that respect.[52]

Estimating the cost of and returns to parenting education (in terms of θ_{PE}) is not easy, so we take the following approach in order to estimate a lower bound on the benefits of such policy. We would like to estimate these from parenting education programs in the US but, to the best of our knowledge, this data is not available. In general, parenting education programs have been more popular in research studies from developing countries so we use that evidence instead. Moreover, even though we were not able to find evidence of costs and long-term impacts from the same study, we used evidence of two programs with similar curricula. We estimate the cost of running such policy in the US—based on the upper-bound available for Colombia (Attanasio, Fitzsimons, Grantham-McGregor, Meghir, and Rubio-Codina, 2016)—to be \$11,400 per family in the first period with children.[53] We also need to estimate the effectiveness of parenting education, i.e., θ_{PE}. In order to do this, we use experimental evidence from a parenting education program that was implemented in Jamaica and studied by Gertler, Heckman, Pinto,

[52] An alternative compelling interpretation of the effect of parenting education is to increase $\bar{A}$. Effectively, both alternatives increase the derivative of θ_k' to τ or m, i.e., the productivity of investments. The key element is how to benchmark the increase of productivity for either alternative. We present here the first approach since our benchmarking is made on the increased income of children from low-income parents. If we were to focus on increases in$\bar{A}$we would get higher returns for high-income individuals than if we follow our selected approach. This is in line with our objective of estimating a lower bound on the potential impact of parenting education.

[53] Running a similar policy in Colombia has been estimated to cost between \$450 and \$750 per child (Attanasio, Fitzsimons, Grantham-McGregor, Meghir, and Rubio-Codina, 2016). This program actively used a group of women ("Madres Lideres") with average education equivalent to a high-school degree. If we assume running the program in the use would use similar inputs we can try to estimate the costs, we can compare the salaries of similar individuals in the US and Colombia to estimate the cost in the US. In order to estimate an upper bound to such cost, we assume here that in the US they would employ college-educated women instead. In Colombia, \$450 represented approximately the average monthly salary of a high-school educated person in Colombia. Assuming this would require a college graduate in the US (whose average salary in the 2000s was approximately \$42,000), this would imply that in the US the cost of running a similar program would be between \$3,400 and \$5,700 per child. Disregarding potential returns to scale of running the program for two children per family, this would imply a cost of up to \$11,400 per family in our model.

Zanolini, Vermeersch, Walker, Chang, and Grantham-McGregor (2014). Parents of growth-stunted children were randomly selected to participate in the program when their children were between 0 and 2. Once children were approximately 22 years old, Gertler, Heckman, Pinto, Zanolini, Vermeersch, Walker, Chang, and Grantham-McGregor (2014) estimate that parenting education program led to 12% increase in the children's income.[54] As shown in Table F6, we choose θ_{PE} such that if a small share of poor families whose children had low initial levels of skills in our estimated economy were introduced to the parenting education program their children's income would increase by 12% as well.[55] This is obtained by θ_{PE} that is 0.8 standard deviations above the average θ. We refer to this value as the benchmark return to parenting education.

Table F6: Parenting Education: benchmarking θ_{PE}

θ_{PE} relative to		Change from Baseline (%)
Avg. θ	**Policy benchmark**	**Income Bottom**
-1.5 SD	-2.3 SD	0.00
-1.3 SD	-2.1 SD	0.97
-1.1 SD	-1.9 SD	1.97
-0.9 SD	-1.7 SD	3.11
-0.7 SD	-1.5 SD	4.28
-0.5 SD	-1.3 SD	5.49
-0.3 SD	-1.1 SD	6.50
-0.1 SD	-0.9 SD	7.48
+0.1 SD	-0.7 SD	8.48
+0.2 SD	-0.6 SD	9.03
+0.4 SD	-0.4 SD	10.09
+0.6 SD	-0.2 SD	11.20
+0.8 SD	**0.0 SD**	**12.09**
+1.0 SD	+0.2 SD	12.90

We use the estimated model (starting from steady state) to simulate experimental evidence on a parenting education program in the spirit of the study of Gertler et al. (2013). We simulate low-income, low-skilled parents with low-skilled children going through the program that increases their parenting skills as explained in the main text. We then evaluate the results on children's income at age 22. We define the benchmark program productivity as the level of skills (in standard deviation terms) required for this income to grow by 12% (i.e., as much as reported by Gertler et al. (2012))

We now evaluate parenting education in three steps. The first two steps estimate what the return would be a government run program that introduced parenting education. Here the government enrolls (and pays for) every agent to obtain parenting education, independently of whether it is ineffective for them

[54]We focus on the estimates for earnings on current job that exclude individuals that migrated to other countries (Table S.14 in the Appendix of Gertler, Heckman, Pinto, Zanolini, Vermeersch, Walker, Chang, and Grantham-McGregor (2014)). This estimate is smaller than the average finding for all individuals of 25%. Thus, consistently with our other choices, we are likely to obtain a lower bound on the gains of such programs.

[55]We focus on families whose children have a low initial draw of skills (to capture the idea of growth stuntedness in the model). Moreover, given that these families lived in poor neighborhoods we focus on non-college graduate and low-skilled parents, whose income is in the bottom 5%.

or not—i.e., the government cannot observe or use the skills of the agents to determine their enrollment. Our third step looks at whether such a program would need to be government enforced. We introduce the program as something that agents can purchase by themselves once children are born and study its effects. For most of the analysis we focus on measures of income inequality, intergenerational mobility, college-graduation rate, and average income. Finding policies that manage to improve these aggregate outcomes is of general interest. As a measure of general welfare, we also report on consumption equivalence.

F.2.1 Parenting Education as a Government Program

We introduce parenting education in the previous steady state and evaluate the effects in the relevant cohort receiving those benefits: children born to the generation receiving the parenting education. This environment is useful to understand the first-order and short-term effects of the policies. Moreover, studies from the empirical literature are more comparable to this environment as their experimental evidence is usually based on small-scale policies and effects are evaluated in the short term.

Table F7: Parenting Education Program

θ_{PE} relative to benchmark	Cons. Equiv.	Avg. Income	Inequality	Mobility	College	Tax Revenue	Tax Rate
	Change from Baseline (%)						
	Partial Equilibrium - Short-Run						
-1.8 SD	0.78	1.14	-0.20	5.66	3.79	2.17	0.00
-1.6 SD	1.33	1.87	-0.31	8.56	6.54	3.75	0.00
-1.4 SD	1.83	2.56	-0.39	10.72	8.95	5.15	0.00
-1.2 SD	2.21	3.19	-0.43	12.89	10.84	6.27	0.00
-1.0 SD	2.57	3.79	-0.47	15.23	12.57	7.31	0.00
-0.8 SD	2.89	4.33	-0.51	17.05	14.14	8.26	0.00
-0.6 SD	3.39	5.22	-0.56	18.72	16.56	9.74	0.00
-0.4 SD	3.84	6.04	-0.64	19.96	18.71	11.06	0.00
-0.2 SD	4.24	6.81	-0.71	21.98	20.68	12.27	0.00
Benchmark	**4.61**	**7.47**	**-0.80**	**23.66**	**22.44**	**13.36**	**0.00**
0.2 SD	4.94	8.10	-0.88	24.97	24.04	14.35	0.00
	General Equilibrium - Long Run						
-1.8 SD	1.44	1.30	-0.93	7.50	1.23	2.11	-0.05
-1.6 SD	2.54	2.09	-1.51	10.78	2.45	2.47	-0.28
-1.4 SD	3.49	2.84	-2.09	15.30	3.55	2.91	-0.46
-1.2 SD	4.26	3.40	-2.44	16.92	4.57	3.10	-0.63
-1.0 SD	4.85	3.78	-2.83	18.86	4.80	3.14	-0.75
-0.8 SD	5.35	4.17	-3.01	21.52	5.72	3.44	-0.85
-0.6 SD	6.19	4.76	-3.48	23.65	6.26	3.57	-1.02
-0.4 SD	6.89	5.25	-3.85	25.47	6.92	3.84	-1.14
-0.2 SD	7.48	5.61	-4.14	27.19	7.43	4.14	-1.24
Benchmark	**8.01**	**5.99**	**-4.37**	**29.08**	**7.94**	**4.29**	**-1.35**
0.2 SD	8.46	6.28	-4.68	30.36	7.98	4.30	-1.43

Notes: We simulate a policy in which the government sets up the parenting education program as explained in the main text. We simulate the program under different levels of efficiency, relative to the benchmark. The program is estimated to cost $11,400 per family for all cases. General equilibrium and long run refers to the case in which the policy is implemented permanently and we look at the effects in the new steady state, taking into account that wages and interest rates adjust to clear the market and the government adjusts the labor income tax to keep its budget balanced. The other case focuses on the effects on the children of the first (and only) cohort of parents is intervened, without considering changes in prices or taxes. This case is similar to an RCT applied to a small representative sample. Regarding the columns, consumption equivalence is determined by newborns under the veil of ignorance. Inequality refers to the variance of log-labor-income while IGE mobility refers to minus the regression coefficient between children's and parents' income ranks.

The top panel of Table F7 shows the results in the short-run partial-equilibrium case. Effects are reported as percent changes from the baseline economy. Each row shows the outcome changes for different levels of θ_{PE}. For example, -0.2 SD means that the effectiveness of the program is 0.2 standard deviations (of θ) lower than predicted by the benchmark estimate. Focusing on the consumption equivalence column, it is clear that even if this policy is not very efficient (e.g., 1 standard deviation below), parenting education stills seems to provide welfare gains. This policy can also generate new tax revenues. For example, by increasing the share of college graduates or the share of high-skilled individuals, the average income, consumption, and savings increase, and so does the tax revenue. A 10% increase in tax revenue is approximately equal to a $500 increase in the tax revenue per household each year. Thus, parenting education, assuming the policy is at least as effective than the benchmark, is even able to increase the resources available for the government in the short run (net of the extra expenditures required to run the program).

We now evaluate the parenting education program in a long-run general-equilibrium environment: it is introduced permanently and we consider the economy in the new steady state. We adjust government (labor) taxes so that its budget does not change. We evaluate what the long-run effects would be— taking into account the interactions between taxation, education, and parental investments towards children. The model now provides evidence that is harder to obtain empirically. The bottom panel of Table F7 shows the results.

Parenting education remains highly beneficial. Once again, even for the cases in which the policy is 1.6 standard deviation less effective than the estimated benchmark, we find the consumption equivalence measure for welfare to be larger than zero. The effect on intergenerational mobility is almost equivalent to the partial equilibrium case. If parenting skills can be improved as much as the literature suggests, it would lead to a decrease in the intergenerational mobility rank-rank persistence coefficient of 0.07 points. Similarly, the effect on average income is almost as strong as in the partial equilibrium case, with the reduction being driven by the wages adjustment. The effect on college graduation is considerably smaller than in partial equilibrium. Parenting education also proves to be a policy that would increase tax revenue in the long run. By increasing the share of high-skilled, the growth in income, consumption, and savings even allows the government to reduce labor tax rates.

F.2.2 Parenting Education Market

Next we look at whether such a program would need to be government enforced. We introduce the program as something that agents can purchase by themselves once children are born and study its effects. In other words, when their children is born they have one more choice to make: acquire parenting education at the price of $11,400 (same as the government estimate) or now. Table F8 shows the results for such exercise, for different levels of productivity θ_{PE} relative to the estimated benchmark. At the benchmark the consumption equivalence is 7.2%, which is associated with a 100% take-up among the low-skilled parents. Intergenerational mobility which would increase by almost 24%. Similarly, average income would increase by 5.6% while inequality would be reduced by 3.9%. Relative to the benchmark, the productivity of the program would need to be far below the benchmark for the take-up to be very small and aggregate effects to be minimal. For very low values of θ_{PE}, no agent acquires parenting education and hence the economy does not change relative to the initial steady state.

Table F8: Policy evaluation: comparison

	Change from Baseline (%)									
	Cons. Equiv.	Avg. Income	Inequality	Mobility	College	Tax Revenue	Tax Rate	Take-Up Low	Take-Up Medium	Take-Up High
-1.8 SD	0.62	0.55	-0.41	-0.97	0.40	0.23	-0.14	27.67	0.00	0.00
-1.6 SD	1.37	1.15	-0.75	1.39	1.10	0.36	-0.32	35.99	0.00	0.00
-1.4 SD	2.70	2.14	-1.43	8.83	2.57	0.81	-0.62	71.00	0.00	0.00
-1.2 SD	3.25	2.65	-1.84	12.49	3.10	1.00	-0.70	79.65	0.00	0.00
-1.0 SD	3.80	3.03	-2.10	13.43	3.62	1.17	-0.84	83.09	0.00	0.00
-0.8 SD	4.41	3.50	-2.38	14.58	4.44	1.34	-0.97	90.29	0.00	0.00
-0.6 SD	5.12	4.10	-2.90	17.46	5.08	1.58	-1.11	98.57	0.00	0.00
-0.4 SD	6.09	4.77	-3.43	20.75	5.85	1.81	-1.32	100.00	0.00	0.00
-0.2 SD	6.70	5.29	-3.71	22.40	6.84	2.15	-1.43	100.00	8.75	0.00
Benchmark	**7.21**	**5.62**	**-3.96**	**24.09**	**7.52**	**2.26**	**-1.53**	**100.00**	**14.72**	**0.00**
0.2 SD	7.82	6.06	-4.27	24.45	8.15	2.54	-1.64	100.00	40.94	0.00

Notes: We allow families to acquire the parenting education program as explained in the main text. We simulate the program under different levels of efficiency, relative to the benchmark. The program costs $11,400 per family for all cases. The table shows the results in the new steady state, taking into account that wages and interest rates adjust to clear the market and the government adjusts the labor income tax to keep its budget balanced. Regarding the columns, consumption equivalence is determined by newborns under the veil of ignorance. Inequality refers to the variance of log-labor-income while IGE mobility refers to minus the regression coefficient between children's and parents' income ranks.

There are differences between the case in which the program is government run (and every household is paid to go through it, even if it provides no benefits) and when parenting education is market good that households decide to acquire. On the one hand, the government program is wasteful in the sense that it pays for people who do not benefit from the program. On the other hand, the government alternative gets a larger take-up. Among these larger take-up, it is important to consider that some families may be constrained when deciding to acquire parenting education. For low values of θ_{PE} the larger take-up is important for low-skilled individuals. Once θ_{PE} is close to its benchmark value it also starts being beneficial for mid-skilled agents. Around the benchmark θ_{PE} the welfare gains of the parenting education program (in general equilibrium) and the parenting education market are similar. This suggests that, if parenting education is as effective as the evidence implies, it may not be necessary for the program to be provided by the government.

F.3 College Subsidies

We introduce government-funded college subsidies into our estimated model. This implies that the private cost of college is now reduced from p_e to $p_e(1-s_e)$ where s_e is the subsidy rate. Abbott, Gallipoli, Meghir, and Violante (Forthcoming) study this type of policies in a framework in which college subsidies have more flexibility than ours. However, differently from theirs, in our model college subsidies can affect skills through endogenous parental investment choices.

We evaluate alternative values of s_e, increasing labor income tax to keep the government's budget balanced. We find that there are welfare gains from subsidizing college, and this tend to be larger for larger values of s_e. Full college subsidies (i.e., $s_e = 1$) are associated with welfare gains of 1.7%, and an increase in college graduation rates from 29.4% to 32.5%. Income mobility increases by 20%. There are substantial changes in the wage gap between college graduates and high-school graduates.$w_2 - w_1$ is reduced by 38% and this is associated with a reduction in the variance of log-income of 5%. This general equilibrium effect on wages implies that children of high-skilled college-graduates parents are not better off after the subsidies are introduced. The welfare gains are concentrated on children of families with lower income who can now afford college more easily. Although this policy is associated with increases in parental investments and average skills, we find these changes to be relatively small (at least for this size of college subsidies and change in college graduation rates), suggesting that the results from Abbott, Gallipoli, Meghir, and Violante (Forthcoming) may not be substantially affected if they introduced endogenous parental investments towards early childhood skills development.

To summarize, college subsidies have the scope of increasing welfare but their effect is less than one-fifth of the one obtained by the program that funds early childhood development investments.We remark, however, that we could make the subsidy rate s_e larger than one and this is associated with larger welfare gains than the ones for $s_e = 1$. But the gains are limited. For example, introducing $s_e = 4$ only increases welfare by 4%, still much lower than those obtained by the early childhood investment program. We abstract from studying $s_e > 1$ in detail here since larger values of s_e are closer to a conditional cash transfer policy than to a standard college subsidy, but results for these cases are available upon request.

F.4 Endogenous Parental Investments and the Welfare Evaluation of Tax Progressivity

Most macroeconomic analysis of inequality focuses on progressive taxation. However, most of the models used for that analysis abstract from including endogenous intergenerational links like childhood development or parental transfers. Given that increasing the amount of income available for the poor (i.e., increasing progressivity) can have an effect on their parental investment decisions, it is possible that by excluding those intergenerational forces their welfare estimates of such policies may be biased. In general, the question that emerges is: *Do tax policy evaluations change once we include endogenous parental investments?*

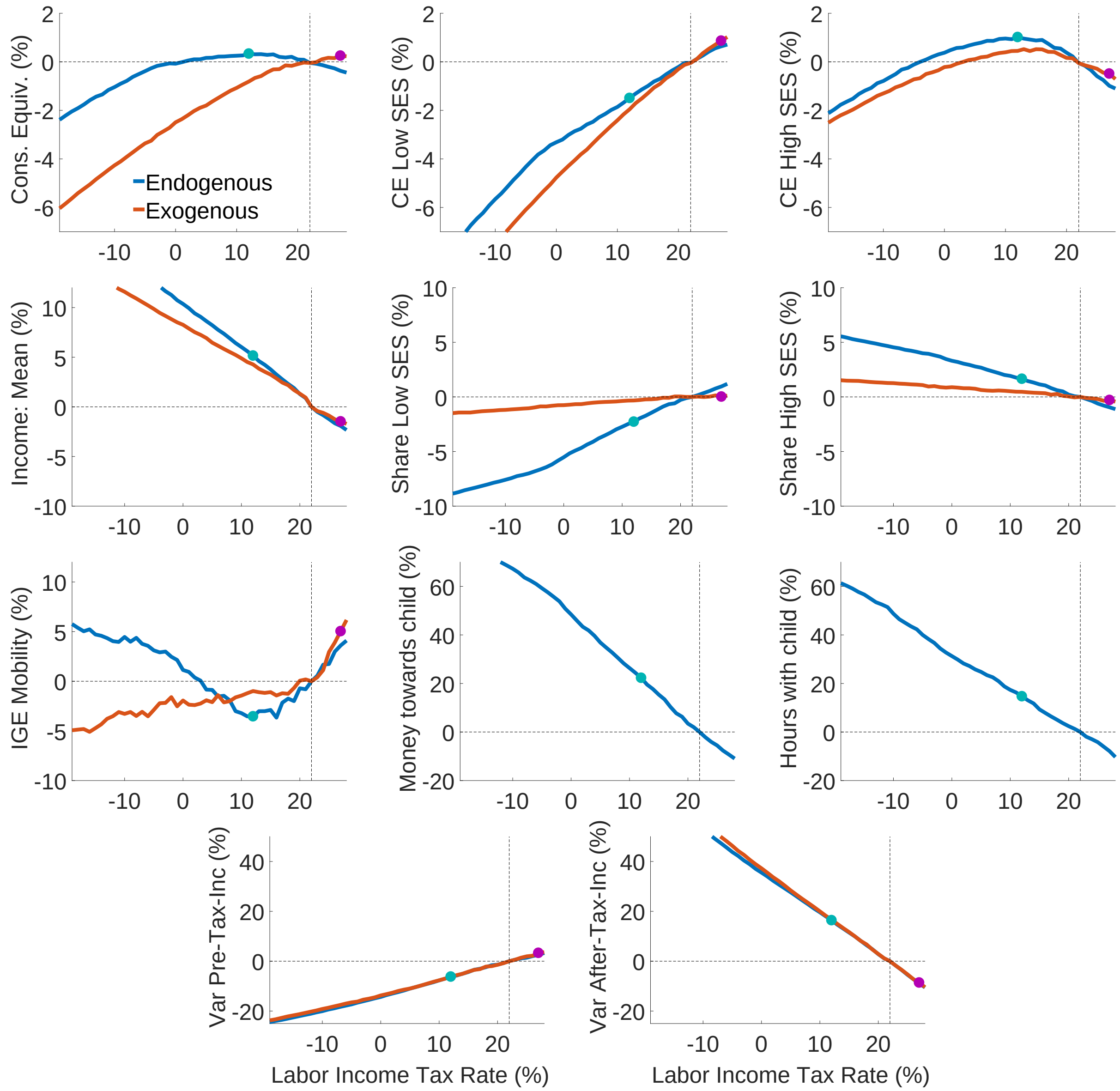
The model introduced here is useful to answer this question as it adds endogenous childhood development and parental transfers to a standard life-cycle macroeconomic model. In order to evaluate the importance of endogenous intergenerational links, we compare the effects in our model with endogenous links to the effects obtained by the same model but with exogenously fixed links. The model with exogenous childhood development is equivalent to the original model, but where the intergenerational transition matrix of skills is fixed to be equal to one obtained (endogenously) in the original steady state. This matrix defines that the distribution of children's skills depends (exclusively) on parents' skills and education group. Hence, the tax system cannot affect the development of skills directly, though it may affect it through education choices. The model with exogenous childhood development is re-estimated to match the same set of moments (excluding the childhood development related ones) from the full model. The estimated parameters and moments are shown in Table F9.

Table F9: Estimation: exogenous childhood development model

Parameter	**Value**	**Description**	**Moment**	**Data**	**Model**
Preferences					
μ	1153	Mean labor disutility	Avg. hours worked	64.5	64.6
b	0.32	Altruism	Parent-to-child transfer as share of income	0.62	0.62
School Taste:					
α	4.0	Avg. taste for college	College share	29	30
α_{θ_c}	-0.1	College taste and cog. skills relation	College: cog skills slope	0.19	0.19
$\alpha_{\theta_{nc}}$	-0.24	College taste and noncog. skills relation	College: noncog skills slope	0.06	0.07
σ_ϕ	1.3	SD of college taste shock	College: residual variance	0.18	0.18
$\bar{\phi}$	-1.8	Draw of school taste: mean by parent's education	Intergenerational persistence of education	0.77	0.76
Interest rate					
ι $(\times 10^2)$	2.3	Borrow-save wedge	Share of borrowers	5.0	4.9
Government					
$\omega(\times 10^2)$	2.4	Lump-sum transfer	Income variance ratio: Disposable to pre-gov	0.69	0.68

For both models, we evaluate modifying the labor income tax τ_y and adjusting the lump-sum transfer ω such that the government's budget remains constant. For the welfare analysis we focus on consumption equivalence under the veil of ignorance as defined in Section 6. Figure F8 shows the results. The model with exogenous childhood skills predicts small welfare increases from increasing tax progressivity, in line with the literature that uses this kind of models to evaluate tax progressivity (e.g., Floden and Lindé, 2001; Conesa and Krueger, 2006; Heathcote, Storesletten, and Violante, 2017). On the other hand, the full model developed here predicts welfare gains (in consumption equivalence terms) of slightly less than 1% from substantially *reducing* tax progressivity.

Figure F8: (Constrained) Optimal Tax Progressivity



Notes: We simulate policies in which the government changes the marginal labor income tax rate, adjusting the lump-sum transfer to balance its budget. The horizontal axis refers to the marginal labor income tax rate . Outcomes (except for the size of the transfer) are reported in changes from the baseline steady state. Consumption equivalence is determined by newborns under the veil of ignorance. CE Low SES refers to the consumption equivalence measured gains for children of low-skilled, non-college educated parents. CE High SES refers to the consumption equivalence measured gains for children of high-skilled college-educated parents. Inequality is shown using the variance of log-income (both pre- and after-tax) while IGE mobility refers to minus the regression coefficient between children's and parents' income ranks.

In addition to the traditional trade-off between equality and efficiency of labor, endogenous childhood skills leads to a new trade-off. On the one hand, higher progressivity may let poor parents increase investments towards child's skills. On the other hand, such progressivity would increase insurance and reduce the after-tax returns to skills, thus reducing the incentive to invest towards children's skills. We find that the second effect is stronger in the long-run. By increasing the incentives to invest towards children, a less progressive taxation can increase welfare in the long run, though this can be associated with a very costly transition. We remark that a more flexible tax function may allow for conditional transfers that increase progressivity, allowing poor parents to invest without harming incentives. We leave this for future research.