The Macroeconomic Consequences of Early Childhood Development Policies

Diego Daruich*

University of Southern California (Marshall)

June 5, 2020

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 as those of a short-run small-scale program, 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.

*daruich@marshall.usc.edu. I am grateful to my advisors Raquel Fernandez and Gianluca Violante. For helpful comments, I thank Christopher Flinn, Julian Kozlowski, and Pau Roldan as well as participants at seminars and conference sessions at New York University, Federal Reserve Board, Federal Reserve Bank of St. Louis, Econometric Society, 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 NBER Summer Institute. 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] Garcia et al. (2019), for example, estimate that for every dollar invested in 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.[2] 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 that gap by incorporating early childhood development into a standard macroeconomic model, which shows that underinvestment in children's development is an important source of inequality and social immobility, 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 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 those 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 et al. (2010).[3]

[1] "Early childhood" refers to the period when children are under 4 or 5 years old, depending on the source. 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 et al. (2015) summarize the economics literature on the importance of early childhood environment.

[2] A back-of-the-envelope calculation suggests that financing a large-scale high-quality early childhood program similar to the one studied by Garcia et al. (2019) may require the government to increase its tax revenue by 6%. If taxes are distortive, the deadweight losses associated with raising taxes may be important. Moreover, Garcia et al. (2019) also find a large effect on the college graduation rate (i.e., increasing by 15 percentage points [p.p.] in the population they study). Thus, if these changes take place on a large scale, we may expect the wages of higher-educated individuals to be reduced, hence affecting the program gains (as well as inequality).

[3] Some evidence suggests potentially important differences of effects across genders (e.g., Garcia et al., 2019). We abstract from differences across genders, however, since this would require enlarging the model in a non-trivial way with some assumptions that can generate the impact differences across genders in our model. There are several possible model alternatives that, to the best of our knowledge, we may not yet be able to distinguish empirically. One could introduce gender differences in the skills production function—which is not part of the Cunha et al. (2010) skills production function

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 Garcia et al. (2019) randomized controlled trial (RCT) to test the validity of the model's predictions on the effects of government investments towards early childhood development. In this study, Garcia et al. examined the effects of two programs in which a small group of disadvantaged children were brought to high-quality early childhood development centers in North Carolina from the age of 8 weeks to 4 years old. The program's cost was approximately $13,500 per child-year. An equivalent program in the model implies introducing government expenditures on early childhood development of $13,500 per child-year, 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 low-income high-school educated parents. 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 Garcia et al..

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 9% in consumption equivalence units.[4] Moreover, the early childhood investments program is associated with a lifetime-earnings inequality reduction of 8% and an increase in intergenerational mobility of 20%, approximately half as large as the difference between US and Canadian levels of immobility and inequality.

Compared to the case in which such a program is permanently implemented, we find that if such government investments were introduced for only one generation and on a small scale they would achieve less than one-half of its welfare gains. We interpret this as evidence that an RCT is likely to underestimate the long-run benefits of such a policy. Even though the policy is self-financing in partial equilibrium (PE), in GE—by lowering the wage of college graduates and therefore the return on those investments—taxes need to be increased to finance the additional government expenditures, reducing

estimation that we use—as well as gender-specific preferences for education and job types. Alternatively, gender biases or higher probability of women quitting the labor force (e.g., due to child birth) may also help explain some of the differences. Although interesting, this is beyond the scope of this paper since our focus is on evaluating long-run GE effects of early childhood policies and comparing them to short-run PE ones. Given this limitation, our results, therefore, may be interpreted as referring to the average across genders.

[4]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. In Appendix G.1, we also evaluate alternative policies that use part of the resources to invest in older children as well but find that welfare gains only increase slightly. Gains are smaller if a large share of resources is used in older children because the child's skills production function as estimated by Cunha et al. (2010) implies that skills are more malleable at younger ages.

welfare gains by almost one-tenth. It is general equilibrium forces, however, that also generate most of the reduction in inequality: Increasing skills augments the share of college graduates which reduces the wages of college-graduates relative to high-school graduates. At the same time, the long-run change in the distribution of parental characteristics generates more than one-half 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.[5]

Even though benefits take time to accrue, our evaluation of transition dynamics shows that if the policy were implemented permanently, every new generation would be better off and more than 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.

In addition to our baseline policy analysis of government investments, we evaluate several model extensions and alternative policies to check the robustness of our results. The baseline model misses that the elements required to produce this early childhood development input may be scarce, particularly when the policy is scaled up. In Section 5.1.4, we provide a simple extension to the model in which the early childhood input is actually hours spent with a college graduate—which is in line with the costs reported for the RCT program on which we base our main analysis. On the one hand, the reform will now drive up the cost of early childhood programs as college-graduate labor is a scarce input. On the other hand, over time the policy itself will increase the share of college graduates, hence driving the cost down. In the long-run we find that both effects almost compensate each other and welfare gains are almost unchanged.[6]

We also evaluate another popular policy regarding childhood development: parenting education programs. Rather than investing in children's development directly, such programs train parents on how to promote children's development. We estimate the costs and returns of running a parenting education program based on the RCT evidence from Gertler et al. (2014) and Attanasio et al. (2016).[7] Similar to our baseline policy, we find that welfare improvements in the long-run GE framework are larger than if the policy is implemented for only one generation and on a small scale.

[5]These results are also qualitatively in line with the recent literature that finds positive second-generation effects of early childhood programs (Barr and Gibbs, 2019; Heckman and Karapakula, 2019).

[6]In the long run, the crucial question is whether the relevant input required for early childhood development can be "produced." Whether a higher-skilled population in fact makes the provision of such input easier or less costly is left for future research.

[7]A relevant caveat is that this evidence is from developing countries. To the best of our knowledge, this is the best evidence available on the costs and returns of such programs so we are limited by this fact, but we try to adjust for this limitation by showing how results change for alternative estimates of returns.

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 his parental investments.[8] To illustrate this channel, consider a poor parent who, by investing in the early childhood development of his 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 his 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 explains the model's estimation and validation exercises. Section 5 presents the policy analysis exercises. Finally, Section 6 concludes. The appendices contain additional details, model extensions, and policy counterfactuals.

2 Related Literature

Macroeconomic and policy analysis of inequality can be divided into 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, 2017). The other focuses on the bottom 99%, typically looking at the role of skills and education (e.g., Katz and Murphy, 1992; Autor et al., 2008; Abbott et al., 2019). In addition to income taxation, some of these papers also study college-education policies (e.g., Bénabou, 2002; Abbott et al., 2019; Holter, 2015; Krueger and Ludwig, 2016).[9] 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 et al., 2011). We also use a standard macroeconomic Aiyagari-style life-cycle model but 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 et al. (2019), 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

[8] 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 introducing a form of intergenerational borrowing (i.e., implemented as a compensation system) leads to the largest gains. Another potential source of underinvestments is that parental beliefs on the elasticity of skills to early childhood investments may be inaccurately low (Cunha et al., 2013). Our model does not incorporate this mechanism and focuses instead on incomplete markets as a source of underinvestments.

[9] Holter (2015) evaluates the importance of education stages before college as well. However, he focuses on cross-country PE comparisons of government education policies after the early childhood stage.

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 in their children's development, which, due to complementarities, may also reduce their incentives to invest time in them. This affects the income and wealth of the next generation, which will again shape 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 et al., 2010; Del Boca et al., 2014). Cunha et al. (2010) 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). We use the skills production function (and estimates) of Cunha et al. 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 et al., 2014; Abbott, 2016; Caucutt and Lochner, 2019), but have abstracted from GE forces (and saving decisions in the case of Del Boca et al.). Given that early childhood development policies can be expensive, this limits these studies' capacity to evaluate large-scale policies.[10] Our model incorporates the early childhood development properties suggested by Cunha et al. (2010) and connects them with inequality and social mobility in an environment suitable for large-scale policies, allowing us to study short-run and long-run effects as well as transition dynamics.

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 PE 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 investment, which the government can improve on by enforcing higher investments in children's development.[11] Aiyagari et al. (2002) shows that GE effects imply that it is possible for a world with borrowing and parental transfers constraints to have higher parental investments than an economy with complete markets because of 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

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

[11] Another potential source of underinvestments is that parental beliefs on the elasticity of skills to early childhood investments may be inaccurately low (Cunha et al., 2013). Our model does not incorporate this mechanism and focuses instead on incomplete markets as a source of underinvestments. Nevertheless, we highlight that our model is also in line with the result that parents tend to spend more quality time with their children once an early childhood development policy is introduced, due to the complementarity between time and money investments in the estimated production function of skills (6).

(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 in children's development, in an economy that takes into account uncertainty in the returns to investments, GE 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 three papers that introduce early childhood development in a quantitative GE model with heterogeneous agents, but all differ from this paper in substantive ways. First, to evaluate government policies that target different stages of development, Cunha (2013) studies the role of monetary early childhood investments in a model of incomplete markets. He finds that a policy that focuses on all stages of the development may lead to the largest gains, but he abstracts from endogenous labor supply, which may have important implications for the cost of raising taxes to pay for policy.[12] Second, Yum (2018) incorporates parental time investments into a GE model but, unlike 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, unlike Cunha (2013) and Yum (2018), we include both time and 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 in studying policies that may be affected by borrowing constraints.[13] Finally, Lee and Seshadri (2019) also study parental investments in a GE model. They focus, however, on reallocating subsidies across different development stages, while we explore alternative levels of government investments. Furthermore, unlike these three papers mentioned, we study in detail the differences between short-run small-scale and long-run large-scale policies, including the transition dynamics. This is a 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).[14]

Given the impressive results highlighted by some small-scale high-quality early childhood development programs (e.g., Gertler et al., 2014; Garcia et al., 2019), one challenge currently under study is how to scale up these programs while maintaining their quality and costs. Richter et al. (2017) summarizes the evidence on scaling up, suggesting that the approach considered to be the most promising involves linking the early childhood programs to pre-existing social programs. This was the approach taken in

[12] Cunha (2013) also abstracts from GE forces with labor inputs differentiated by educational levels, which our results suggest are one of the main drivers of the reduction in income inequality.

[13] Including time investments is important since low-income parents can still invest time on their children. Including monetary investments is also important since it allows us to interpret government investments in early childhood programs within the baseline framework of the model.

[14] Another difference between our paper and these papers is that we include both cognitive and non-cognitive skills which Cunha et al. (2010) highlight to be important for the estimation of the elasticity of substitution of the skills 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.

several places including New Jersey (Lobman et al., 2005) and Colombia (e.g., Attanasio et al., 2014), among others. The Colombian case took advantage of a pre-existing conditional cash transfer program and a community network ("Madres Lideres") to try to replicate at scale the successful Jamaican early childhood development program studied by Gertler et al. (2014). In the case of New Jersey, the state supreme court ruled that 31 low-income school districts had to provide high-quality pre-school for all 3- and 4- year-olds, replicating the models of the Perry and Abecedarian programs. The public school system did not have the capacity to house all of the new high-quality preschool programs mandated by the court ruling, so they were allowed to contract with other child care centers (including Head Start ones) that were already offering preschool. To staff these early childhood programs, however, a large number of teachers were required to return to school. To address this, alternative teacher education programs were introduced and Lobman et al. (2005) found that they produced enough teachers to meet the extra demand. Even though it is still unclear whether these larger programs can maintain the quality of their small-scale counter-parts, these studies suggest methods on how to take advantage of pre-existing resources and show that it is in fact possible to quickly increase the size of such programs. Although this paper does not directly contribute to this empirical literature, the extension of the baseline model studied in Section 5.1.4 highlights that, in line with teacher education programs mentioned in the case of New Jersey, the crucial question for the long-run results is whether the relevant input required for early childhood development can be "produced." And whether a higher-skilled population (like the one a successful early childhood program might create) makes the provision of such input easier or less costly.

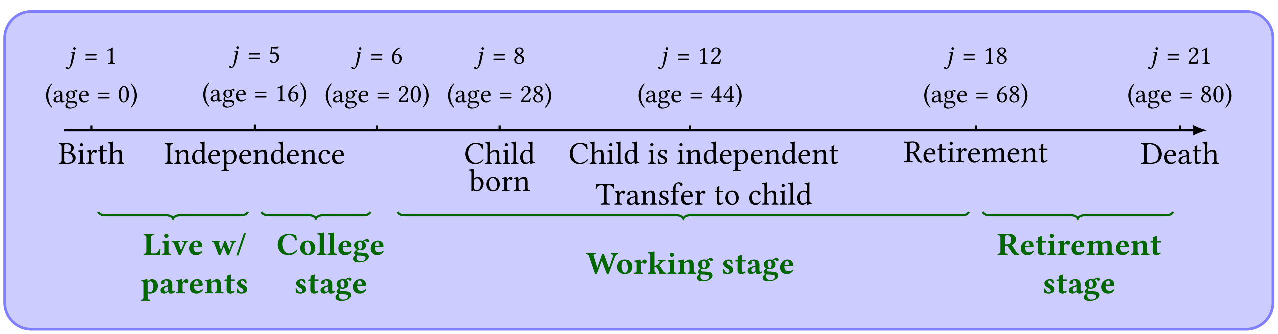
3 Model

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 school taste (or taste for college). 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 rates, GE is important to study the potential effect on the wage of college graduates. A representative firm combines 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.

3.1 The individual problem

There is a dynastic framework with four main stages (20 periods total): childhood, college, work/parenthood, 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 ($j = 1$ refers to ages 0–3, $j = 2$ to ages 4–7, etc.). From $j = 1$ until $j = 4$ the child lives with her parents and does not make any choices. For the purposes of our study, at $j = 5$ (age 16), the individual becomes independent (i.e., start making choices) after finishing high school with a level of skills, which depends on her parents' investments, as well as with an amount of assets, also decided by her parents. The agent's first choice is between going to college or remaining a high school graduate. Once she exits the education phase, she enters the third stage, which represents her working and parenting experience. Throughout their lives, agents choose their savings, consumption expenditures, and labor supply (idiosyncratic uninsurable risk makes labor income stochastic). They can borrow up to a limit, and save through a non-state-contingent asset. In period $j = 8$ (age 28), the agent becomes a parent and new decisions must be made. When her child is young, she chooses how much time and money to invest in his skill development, and before the child is of college age, she also decides the amount of monetary resources to transfer to him. 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.

Figure 1: Model's life cycle, a dynastic framework with four main stages



Credit market Agents can only trade risk-free bonds, but we allow the interest rate to differ between saving and borrowing. Agents with positive savings receive interest rate r, while those borrowing pay interest rate $r^{-} = r + \iota$ (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 their life-cycle and by education. Student loans are explained in detail below. Young workers (i.e., those under the age of 20) and retired households cannot borrow. Workers with access to borrowing (i.e., workers age 20 and older) are subject to credit limits which we will estimate using self-reported limits on unsecured credit from the Survey of Consumer Finances (see Section 4 for details).

Wage process The wage of an individual depends on her education e and her efficiency units E. In particular, her wage is given by $w_e E_{e,j}(\theta, \eta)$, where w_e is the wage of education group e and $E_{e,j}(\theta, \eta)$ is the efficiency units, including the age profile for the education group, the return 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 $j = 5$ (16 years old), the agent has the option to go to college (for one period, until $j = 6$) with the alternative being to start work-life as a high school graduate. The individual state variables are savings a, skills θ (a vector that includes cognitive skills θ_c and non-cognitive skills θ_{nc}), and school taste ε. All agents become independent as high-school graduates ($e = 0$). If an agent chooses to go to college, her education changes to $e = 1$. The education decision is irreversible and entails a monetary cost of education p_e. In addition, as is common in the literature (e.g., Heckman et al., 2006; Abbott et al., 2019), we also allow for school taste to affect the utility of going to college as a linearly separable element in the value function. [15] After leaving school, school taste is assumed not to affect any adult outcome.

Agents can finance college by using their assets (i.e., parental transfers), by working, or with loans. College students have access to subsidized loans at rate $r^s = r + \iota^s$, where $\iota^s < \iota$ and with borrowing limits $\underline{a}^s$. 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 4. To simplify computation, we follow Abbott et al. (2019) and assume that college student debt is refinanced into a single bond that carries interest rate r^-, where $\tilde{a}^s(a')$ is the function performing this transformation. When making this calculation, we assume that fixed payments would have been made for five periods (i.e., 20 years) following graduation.[16]

College students are also allowed to work—providing high-school level labor—but their total available time is reduced by fixed study time $\overline{h}$. [17] Thus, let $V_j^s(a, \theta, e = 1)$ show the value of an agent in college

[15]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).

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

[17]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.

and with assets a and skills θ.[18] This value is defined by

$$V_j^s(a,\theta,e=1) = \max_{c,a',h} u\left(c, h+\overline{h}\right) + \beta \mathbb{E}_{\eta|e} V_{j+1}\left(\tilde{a}^s(a'), \theta, e, \eta\right) \tag{1}$$

$$c + a' + p_e - y + T(y,a,c) = a(1+r)$$

$$y = w_0 E_{e,j}(\theta) h, \quad a' \geq \underline{a}^s, \quad 0 \leq h \leq 1-\overline{h}, \quad \eta \sim \Gamma_{e,0}.$$

She can borrow up to the limit $\underline{a}^s$, repaying at interest rate $r^s > r$ or save with a rate of return r. We denote as w_e the wage for an agent with education level e. This student has not finished college yet, so she is assumed to be providing high-school level labor. The flow utility, including the disutility of working, is given by u. Agents pay net taxes $T(y,a,c)$ on their labor income y, assets a and consumption c.

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

$$V_j(a,\theta,e,\eta) = \max_{c,a',h} u(c,h) + \beta \mathbb{E} V_{j+1}(a',\theta,e,\eta'), \tag{2}$$

$$c + a' - y + T(y,a,c) = \begin{cases} a(1+r) & \text{if } a \geq 0 \\ a(1+r^-) & \text{if } a < 0 \end{cases}$$

$$y = w_e E_{e,j}(\theta,\eta) h, \quad a' \geq \underline{a}_{e,j}, \quad 0 \leq h \leq 1, \quad \eta' \sim \Gamma_{e,j}(\eta).$$

The agent can borrow up to the limit $\underline{a}_{e,j}$, paying an interest rate $r^- > r$, or save at return r. The return from working is the wage w_e scaled by $E_{e,j}(\theta,\eta)$—a function of the worker's age, education, skills, and idiosyncratic labor productivity.

Thus, at the beginning of period $j=5$, V_j^{sw} is the value of an agent who chooses between working (as a high-school graduate) and going to college,

$$V_j^{sw}(a,\theta,\varepsilon) = \max\left\{\mathbb{E}_{\eta|e} V_j(a,\theta,0,\eta), V_j^s(a,\theta,1) - \kappa(\varepsilon,\theta)\right\}.$$

The disutility of going to college is $\kappa(\varepsilon,\theta)$, which depends both on school taste ε (whose distribution may depend on parental education) and skills.

3.1.2 Working stage

From $j=6$ until retirement, which starts at $j=18$ (age 68), the agent works and her individual problem is mostly equivalent to (2). The problem changes when the agent's child is born at the exogenously given fertility period $j=8$ (age 28). We assume that each agent has one child—or, alternatively, each

[18] We assume that the initial draw of η takes place after going to school. Given the functional form assumptions, this implies we evaluate the efficiency units E at the mean value of η (i.e., zero).

household has one household offspring. Children are born with skills θ_k that are potentially correlated with the parent's skills.[19] Then, 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 is about to become independent, at $j = 12$ (age 44), the agent chooses the size of the parent-to-child transfer $\hat{a}$. We also introduce the choice for child's consumption c_k since these extra expenses associated with children may make the borrowing constraints more binding. We assume that children's consumption utility is evaluated using the same utility function u.[20] We model altruism à la Barro and Becker (1989), in which the agent cares about the utility of her child (including the child's consumption while young) with the altruistic weight δ.

Investments in child's skills This is where the key novelties of the model are present. In periods $j =$ 8 and 9, the agent works and also invests directly in her child's development of skills θ_k (a vector that includes cognitive skills $\theta_{c,k}$ and non-cognitive skills $\theta_{nc,k}$). 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 et al. (2010) but, differently from them, we model parental investments explicitly to incorporate τ and m in the inner CES.

Next-period child's skills θ_k' depend on current child's skills θ_k, parents' skills θ, and parental investments I—as well as on idiosyncratic shock ν. The inner CES function shapes parental investments using both time τ and expenditures m spent on the child.[21] 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).[22]

$$V_j\left(a,\theta,e,\eta,\theta_k\right) = \max_{c,c_k,a',h,\tau,m} u\left(c,h\right) + \delta u\left(c_k,0\right) - v\left(\tau\right) + \beta\mathbb{E}V_{j+1}\left(a',\theta,e,\eta',\theta_k'\right), \tag{3}$$

$$c + c_k + a' + m - y + T\left(y,a,c\right) = \begin{cases} a\left(1+r\right) & \text{if } a \geq 0 \\ a\left(1+r^-\right) & \text{if } a < 0 \end{cases}$$

[19]We assume that the initial draw of skills is related to parents' skills as an AR(1) process, calibrated using the estimates of Cunha et al. (2010) (see Section 4 for details on the estimation). Given that measuring skills at birth is particularly hard, we do several robustness exercises on the distribution of the initial draw of skills in Section 5.1.3.

[20]Given the parametric assumptions, we evaluate the children's utility flow u, which assumes that they are not working (i.e., at the value of $h = 0$).

[21]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 suggest that individuals who spend more time with their children reduce leisure time instead of hours worked.

[22]We discuss the estimation of this parameter in Section 4, but as a preview, we find that the evidence tends to suggest that time and money are imperfect complements.

$$
\begin{aligned}
&y = w_e E_{e,j}(\theta,\eta)\,h, \quad a' \geq \underline{a}_{e,j}, \quad 0 \leq h+\tau \leq 1, \quad \eta' \sim \Gamma_{e,j}(\eta) \\
&\theta'_{q,k} = \left[\alpha_{1,q,j}\theta_{c,k}^{\rho_{q,j}} + \alpha_{2,q,j}\theta_{nc,k}^{\rho_{q,j}} + \alpha_{3,q,j}\theta_{c}^{\rho_{q,j}} + \alpha_{4,q,j}\theta_{nc}^{\rho_{q,j}} + \alpha_{5,q,j}I^{\rho_{q,j}}\right]^{1/\rho_{q,j}} e^{\nu_q} \\
&\nu_q \sim N(0,\sigma_{q,j,\nu}), \quad q \in \{c, nc\} \\
&I = \bar{A}\left[\alpha_m m^{\gamma} + (1-\alpha_m)\tau^{\gamma}\right]^{1/\gamma} \\
&m \in \{m_1, m_2, \ldots\}, \quad \tau \in \{\tau_1, \tau_2, \ldots\}.
\end{aligned}
$$

After two periods, the child's skills are fixed, so the problem is equivalent to equation (2) but with the extra state variable θ_k.[23]

Child becomes independent Just before the agent reaches age-period $j = 12$ (i.e., when she is 44 and the child is 16 years old), she needs to decide the size of monetary transfers $\hat{a}$ 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}.

$$
\begin{aligned}
V_{\text{Transfer}}(a,\theta,e,\eta,\theta_k) = \max_{\hat{a}} V_{j=12}(a-\hat{a},\theta,e,\eta) + \delta\mathbb{E}V^{sw}_{j'=5}(\hat{a},\theta_k,\varepsilon), \\
\hat{a} \geq 0, \quad \varepsilon \sim N(\bar{\varepsilon}_e, \sigma_{\varepsilon}).
\end{aligned} \tag{4}
$$

Importantly, the transfer $\hat{a}$ 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 ε of her child. 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 taste from equation (1) may also depend on the skills level of the child.[24] Notice that unlike (3), the value function at this stage now includes the continuation value of the child $V_{j'=5}$, where j' stands for the age-period of the child. This is the last period in which parents' choices affect their chilren (and future descendants). As the problem is written recursively, this implies that at every period in which parents' choices affect their children's outcomes (i.e., all previous periods), the utility of their children (and future descendants) is taken into account. This formulation embeds the parental altruism motive. After the agent's child becomes independent, the individual problem is equal to equation (2).

[23]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 in older children. In Appendix F, we also provide an extension in which we incorporate an additional period of skills development. After re-estimating the model, we show that long-run welfare gains are still sizable. More important for our main result, however, this extension leads to an even larger gap between long-run and short-run welfare gains since in a model with three periods of skill development, parents are even more important, as the gains from the early childhood program have more periods during which effects can fade out—particularly in the short-run analysis. In the long-run, however, the policy is able to create better parents, making the fading out of the results less prominent than in the short-run.

[24]The functional form of the stochastic processes of skills and school taste are specified in Section 4. The assumption that the school taste is not perfectly known to parents helps make the problem smoother, which is useful for computational reasons.

3.1.3 Retirement stage

At $j = 18$ (age 68), 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 skills level. Agents are assumed to provide no work at this stage, so $h = 0$. Unsecured borrowing is also not allowed at this stage. Formally, the problem at the age of retirement is

$$\begin{aligned} V_j\left(a,\theta,e\right) = \max_{c,a'} u\left(c,0\right) + \beta V_{j+1}\left(a',\theta,e\right), \\ c + a' = \pi(\theta,e) + a\left(1+r\right) - T(0,a,c), \\ a' \geq 0, \end{aligned} \tag{5}$$

where π are the retirement benefits, which depend on the agent's education and skills level.[25]

3.2 Aggregate production function

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

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

Capital depreciates at a fixed rate per period.

3.3 Definition of Stationary Equilibrium

The model includes 20 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-period 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.[26] (See Appendix B for details.)

[25]We use education together with the skills level, as a proxy to approximate average lifetime income, from which the retirement benefits are determined. See Section 4 for details.

[26]The government is allowed to have other expenses which are wasteful (or linearly separable in the utility) 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 5 this is done such that these extra government expenses remain fixed.

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 the child's future income that would be increased by current 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 parents were free to borrow against their own future income. If the child could promise to compensate his parent in the future, the parent would not need to reduce her consumption and the disincentive to invest problem would be avoided. This example shows that imperfect parental altruism ($\delta < 1$) is not the direct source of underinvestment. Even if parents are perfectly altruistic ($\delta = 1$), 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.[27] Lack of child-to-parent compensation is not the only reason 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 future money to be transferred at the time that parental investments in the child's development need to 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 their children.

There is no 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, detailed in Section 4, 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 into our economy that is "special," in the sense that this family is not (and, importantly, has never been) subject to one of these three sources of reduced investments (i.e., lack of a compensation

[27] Even though altruism δ is not the direct source of underinvestment, transfer constraints are more likely to bind if δ is low.

mechanism, uncertainty, and borrowing constraints).[28] We then estimate welfare gains by looking at how much extra consumption an agent would need to be indifferent to 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. That 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 parents directly for their investments, 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 36%.[29,30] In comparison, eliminating uncertainty (both in the income process as well as skills development) leads to welfare gains of 4.4% while enlarging borrowing limits (even up to 2 times the baseline value) generates almost no gains. Thus, these results suggest that lack of compensation (associated with constraints in borrowing against future generations' income) is the leading source of underinvestment, followed by uncertainty and borrowing constraints.[31]

Loury (1981) provides a simpler PE model in which he can show that government investments in childhood development are welfare improving even when lump-sum taxes are needed pay for them.[32] In Section 5 we quantitatively evaluate such a policy in our richer model, which also takes into account distortive taxation and GE effects.

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

[29]Notice that this compensation mechanism implies that parents receive the money when the child is young but the child only pays once he 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 smaller. In such a 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.

[30]This case is associated with increases in average parental time and money investments of over 200% and 400%, respectively. Intergenerational mobility within this special family, measured according to the rank-rank coefficient, is also 40% higher.

[31]There are alternative ways to try to measure the relevance of each channel. We also compared investment choices this special family makes relative to a baseline family with similar characteristics regarding skills—results are similar if we compare families with similar income. We find that doubling the borrowing limits would lead a low-skilled family to invest approximately 3% more than with the estimated borrowing constraints. In comparison, increasing borrowing limits for high-skilled families only increases their investments by less than 0.5%. On the other hand, eliminating the uncertainty increases the investments of high-skilled families by 27% but has a smaller effect (15%) on low-skilled families.

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

4 Estimation

In this section we describe how we parameterize 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.

4.1 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.[33] We use four primary data sources: (i) the Panel Study of Income Dynamics (PSID); (ii) the Child Development Supplement (CDS) to PSID; (iii) Consumer Expenditure Survey (CEX), and (iv) the 1979 cohort of the National Longitudinal Survey of Youth (NLSY). 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 PSID and NLSY data. First, as is standard in the literature (e.g., Heathcote et al., 2010), we drop household observations that report hourly wages below half the minimum wage. 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. Lastly, we drop individuals who do not have a high-school degree, which reduces the sample by approximately 8% and decreases the computational complexity as in this way we have only two education levels.[34] Details about sample selection are reported in Appendix A.

Demographics An age-period in the model is four years. For the purposes of this paper, individuals become independent at the start of age-period $j = 5$ (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 $j = 6$ (age 20). Parental time and money investment decisions are made at the time of (average first) birth $j = 8$ (age 28) and the period after. At age $j = 12$ (age 44), just before the agent's child becomes independent, she chooses the assets to transfer to her child. Retirement occurs at $j = 18$ (age 68). Death is assumed to occur for all agents at age $j = 21$ (age 80).

[33]Given that fertility is exogenous in the model, this implies that every household in the model has one household (i.e., two children) as offspring.

[34]High-school graduates are defined as individuals with 12 to 15 years of education, while college graduates are defined as individuals with at least 16 years of education.

Prices Prices are normalized such that the average annual household income of a high school graduate at age 48 is equal to 1 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.[35] 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 $\{-\$20,000, -\$34,000\}$ for high-school and college graduates, respectively.

Taxes and replacement benefits The tax function is assumed to be $T(y, a, c) = \tau_y y + \tau_k ar\mathbf{1}_{a\geq 0} + \tau_c c - \omega$. Based on McDaniel (2007), we set $\tau_y = 0.22$, $\tau_k = 0.27$, and $\tau_c = 0.07$. The government's lump 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). The pension replacement rate is based on the federal Old Age, Survivors, and Disability Insurance program. We use education and cognitive skill level to estimate the average lifetime income on which the replacement benefit is based.[36] Any amount raised in taxes beyond the expenses on transfers is ignored (i.e., not valued or valued in a linearly independent fashion), and this quantity is held constant in all policy experiments.

Intergenerational transmission of skills Cunha et al. (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). We assume that the child development function is of the nested CES form. The outer CES is based on Cunha et al. but we parameterize and estimate the investment factor as an inner CES, with parental time and monetary expenditures as inputs. Hence, the functional forms are

$$\theta'_{q,k} = \left[\alpha_{1,q,j}\theta_{c,k}^{\rho_{q,j}} + \alpha_{2,q,j}\theta_{nc,k}^{\rho_{q,j}} + \alpha_{3,q,j}\theta_{c}^{\rho_{q,j}} + \alpha_{4,q,j}\theta_{nc}^{\rho_{q,j}} + \alpha_{5,q,j}I^{\rho_{q,j}}\right]^{1/\rho_{q,j}} e^{\nu_q} \tag{6}$$
$$\nu_q \sim N(0, \sigma_{q,j,\nu}), \quad q \in \{c, nc\}$$
$$I = \bar{A}\left[\alpha_m m^{\gamma} + (1-\alpha_m)\tau^{\gamma}\right]^{1/\gamma}.$$

We use Cunha et al.'s preferred estimation (on a representative sample) for the outer CES, which requires allowing parameters to vary with the age of the children. We are able to use their estimates because they also report summary statistics for skill vectors θ_k and θ, particularly the estimated covariances.[37] Their main finding is that skills are more malleable when children are young, i.e., the elasticity of substitution

[35] We take into account grants and scholarships, such that only private tuition costs are considered.
[36] See Appendix C.2 for details.
[37] Appendix Table C1 shows the parameter values and standard deviations. We also need to transform their estimates from 2-year periods to 4-year periods. See Appendix C.1 for details.

determined by ρ_j is larger the younger the children. [38,39]

as assumed by Cunha et al., parental investments I are assumed to be unique and cannot be separated between skills. 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 log-cognitive skills in the estimated economy is zero, since this normalization approach is used by Cunha et al. and we estimate the income process under this normalization.[40]

Agostinelli and Wiswall (2016a) use a Monte-Carlo simulation exercise to show that Cunha et al.'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 are only able to apply it on children older than 5 years of age, hence limiting its use for our purposes. However, given the importance of the skills production function, in Section 5.1.3 we evaluate how our results would change if parameters were 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.

Wages and return to skills The wage process and return to skills are important elements for the model since they determine the career profile, including the amount of uncertainty. 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 return to skills.[41] Heckman et al. (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.[42]

We propose that the wage process of household i with education e at age j is given by $w_{e,t}E_{i,e,j}$, where $E_{i,e,j}$ are the efficiency units. These are defined by $E_{i,e,j} = \epsilon_{e,j}\psi_{i,e,j}$ where $\epsilon_{e,j}$ is the age profile for the

[38] Cunha et al. highlight that abstracting from the two types of skills (i.e., cognitive and non-cognitive) leads to estimates that suggest that investments on low-skilled children are much less productive (i.e., a more negative ρ_j).

[39] The initial draw of θ_k is also estimated by Cunha et al. so we use their estimates on the covariances—Appendix Table 10-3 in their paper—of these initial draws in our model. In particular, we assume that the initial draw of cognitive and non-cognitive skills is related to parents cognitive and non-cognitive skills, respectively. In particular, the initial draw of cognitive skills is given by $log\left(\theta_{ic}\right) = \rho_c log\left(\theta_{pc}\right) + v_{ic}$, where θ_{pc} is the cognitive skill level of the parent and idiosyncratic shock v_{ic} is normally distributed. The same function (with different parameter values) is used for non-cognitive skills. Parameter values are estimated to match the above mentioned covariances.

[40] Cunha et al. (2010) re-normalize skills every period, while our estimation procedure only targets that the normalization needs to hold at the end of the skills development process. Nevertheless, average skills (in logs) are close to zero (0.1) at the end of the first period of development and results are robust to making $\bar{A}$ age dependent so that skills are normalized every period. Results, available upon request, are almost unchanged if we make $\bar{A}$ age dependent and normalize both periods to have average log-skills of 0.

[41] As in the rest of the model, these are estimated using data from households with two adults. Details about sample selection are reported in Appendix A.

[42] Even though non-cognitive skills are assumed to not affect wages directly, the estimated model allows them to affect the (cognitive and non-cognitive) skills production function as well as education choices through the school taste function κ.

education group e and $\psi_{i,e,j}$ is the idiosyncratic labor productivity, which is specified as:

$$log\left(\psi_{i,e,j}\right) = \lambda_e log\left(\theta_{ic}\right) + \eta_{i,e,j}$$
$$\eta_{i,e,j} = \rho_e \eta_{i,e,j-1} + z_{i,e,j}, \quad z_{i,e,j} \overset{iid}{\sim} N\left(0, \sigma_{e,z}\right),$$

where θ_{ic} is the level of cognitive skills (one of the elements of θ_i) and $\eta_{i,e,j}$ is the idiosyncratic shock. The initial value of productivity of an agent $\eta_{e,0}$ is drawn from a normal distribution with mean zero and variance σ_{e,η_0}. 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 the PSID to estimate separately, by education group, the age profile $\epsilon_{e,j}$ as a second order polynomial. Since the model has 4-year-long periods, we estimate this income process by grouping observations over 4 years correcting for selection into work.[43] We include year fixed effects to control for possible changes in average wages over time. We use the PSID instead of 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.[44] 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 skills on wages. The NLSY is useful for this because it reports the Armed Forces Qualification Test (AFQT) score for sampled individuals, a typical measure of cognitive skills. For each household, we remove the appropriate age profile estimated from the PSID, and estimate the return to skills.[45] Consistent with previous evidence, Table 1 shows that these return is higher for college graduates than high-school graduates.[46]

Finally, the residual after removing the age and skill component is used to estimate the process for the idiosyncratic shock $\eta_{i,e,j}$. 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 time lags and age groups), separately for each education group. As reported in Table 2, the results suggest that shocks are persistent, particularly for higher-educated groups.

[43] To control for selection into work we use a Heckman-selection estimator. In particular, we construct Inverse Mills ratios by estimating the participating equation separately for each education group using number of children as well as year-region fixed effects.

[44] 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.

[45] 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.

[46] Abbott et al. (2019) also show that skills increase hourly wages, and that this return is higher among college graduates than among high school graduates. We contribute to this literature by highlighting that this pattern is robust to a focus on two-adult-households.

Table 1: Return to skills by education group

	(1) High School	(2) College
log(AFQT)	0.471***	1.008***
	(0.0335)	(0.0768)
Observations	7,015	3,378
R-squared	0.045	0.082
# of households	988	487

*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 year fixed effects. Methodology is explained in the main text.*

Table 2: Income process estimation: idiosyncratic

	(1) High School	(2) College
ρ_e	0.924	0.966
$\sigma_{e,z}$	0.029	0.046
σ_{e,η_0}	0.050	0.047

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

Parental Investments The PSID survey follows a nationally representative sample of over 18,000 individuals living in the US, and the CDS provides information on a subsample of children: their test scores, time spent with parents, and parental expenditures on them. The main benefit of the CDS dataset for this study is that it compiles detailed diaries on the time parents spend with their children. Using these diaries, we estimate the amount of "quality time" parents spend with their children, focusing on time spent playing and reading with them since these are suggested to be the most productive forms of interactions for young children's development (Samuelsson and Carlsson, 2008).[47]

Even though PSID data includes some forms of parental expenditures, it typically misses expenditures on school fees, extracurricular activities, museums, nannies, and other related costs.[48] Thus, for our main estimates of parental monetary investments we use the CEX (survey years 1996 to 2000). We follow the steps suggested by the US Department of Agriculture and Lee and Seshadri (2019) and estimate

[47] See Appendix A.1 for details.

[48] One of the moments in the estimation is the correlation between parental time and money investments. For this moment, we are required to use PSID-CDS data for both forms of investment. We follow the same selection criteria as those used with the CEX.

the parental expenditures on child-care dropping parents who report no child-care expenses.[49] Table 3 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 second column focuses on families with two children in which the two parents live together while the child is under the age of 12. We prefer to use the estimates from the whole sample since the more restrictive sample is much smaller (4,024 vs. 1,546 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 2.6 hours/week. We interpret these findings as suggesting that the model estimation wouldn't change significantly if we made the sample selection more restrictive.

Table 3: Parental Investments: Summary

	All	Parents Together 2 Children
Sample Means		
Weekly Hours	18.0	20.6
	(0.3071)	(0.6721)
Yearly Expenditures	1,966	1,553
	(35.53)	(57.31)
Regression Coefficients		
Hours on College	3.734***	2.473**
	(0.518)	(1.179)
Log(Hours) on Log(Income)	0.123***	0.0481
	(0.0234)	(0.0760)
Expenditures on College	732.4***	665.7***
	(67.80)	(106.75)
Log(Expenditures) on Log(Income)	0.391***	0.634***
	(0.0285)	(0.0624)

*Source: PSID-CDS and CEX. Robust standard errors in parentheses. *, **, *** denote statistical significance at the 10, 5, and 1 percent, respectively. Weekly hours are based on "quality hours" as defined in the text, estimated using PSID-CDS data. Yearly expenditures includes child-care expenditures, estimated using CEX data. 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 time investments across children is reduced when we focus on more selective samples, but higher-educated and higher-income parents tend to spend more time and money on their children.

[49]Our estimates for average expenditures are very similar to those in the USDA report "Expenditures on Children by Families, 2015" (based on CEX data) and those used by Lee and Seshadri (2019) (based on PSID-CDS data).

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 termed "psychic costs of education" (e.g., Abbott et al., 2019; Krueger and Ludwig, 2016). We assume that school taste in utility terms follows the function $\kappa(\varepsilon, \theta) = exp\left(\alpha + \alpha_{\theta_c} log\left(\theta_c\right) + \alpha_{\theta_{nc}} log\left(\theta_{nc}\right) + \varepsilon\right)$. This function allows for higher skilled individuals to have (on average) lower levels of school taste if $\alpha_{\theta_c} < 0$ and $\alpha_{\theta_{nc}} < 0$. Then, ε is an idiosyncratic shock which is assumed to follow a normal distribution $N\left(\bar{\varepsilon}_{e_p} - \frac{\sigma_{\varepsilon^2}}{2}, \sigma_\varepsilon\right)$ 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_c} and $\alpha_{\theta_{nc}}$ are estimated to match the relation between college graduation and skills (measured by regressing a college graduation dummy on the log of cognitive and non-cognitive skills), σ_ε is estimated to match the variance in college graduation after controlling for skills (i.e., the variance of the residual in the previous regression), and $\bar{\varepsilon}_{e_p}$ is estimated to match the intergenerational persistence of education.

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 commonly used: 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 $\iota^s = 0.009$ (see Daruich and Kozlowski, 2019). In the model, the borrowing limit while in college is the set to match the cumulative borrowing limit on Stafford loans ($23,000).

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).[50] 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 δ 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 (until they are 26 years old) and obtain an estimate of total parental

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

transfers per child of \$48,381, which corresponds to 75% of average annual household income.[51] The benefit of using this data is that we can keep the estimation sample consistent with those 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 δ are in line with the literature.[52]

Aggregate production function We assume there is an aggregate firm with production function $Y = AK^{\alpha}H^{1-\alpha}$, 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 an 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 et al., 1998). (See Appendix C for details on the estimation.) A is chosen such that the price normalization mentioned above holds. The per-period capital depreciation rate is set such that the annualized depreciation rate is 6.5%.

4.2 Simulated Methods of Moments: Results

Thirteen parameters of the model are estimated using simulated method of moments. δ relates to the degree of altruism, while μ and ξ are the disutility of labor and time spent with children, respectively. α, α_{θ}, $\bar{\varepsilon}$, and σ_{ε} relate to the distribution of school taste and its relation to 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.

[51] We follow a procedure similar to that used by Johnson (2013). 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). In particular, we divide children into 2-year age groups between the ages of 17 and 26 and calculate the average parental transfer received by each of these groups. We then sum all these averages to estimate total transfers received. PSID transfer data include large transfers to buy houses or cars as well as in-kind transfers like college tuition. We also take into account housing costs if the child lives with the parents, assigning in such cases the average annual rent value of \$7,110 as a transfer. Our estimates are in line with, but slightly larger, than the ones from Johnson (2013) and Abbott et al. (2019)—20% and 30% larger, respectively—since we include a few more years of transfers and focus on children with one sibling (so that the family size for each parent is consistent with the model). Section 5.1.3 shows that the welfare gains from the policy studied here tend to be smaller when δ is higher (when the target moment of transfers is higher). Therefore, our results should be conservative if compared to alternative estimations based on Johnson (2013) and Abbott et al. (2019).

[52] An alternative procedure to estimate δ 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 approximately 0.03, as seen in Appendix C.4. Our robustness results in Section 5.1.3 show that δ 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 1 percentage point of our baseline estimates.

We implement a simulated method of moments procedure to estimate the model. We use a Sobol sequence in order to solve and simulate 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 these solutions 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 so many solutions are required to properly cover the hypercube. Having solved the model in all these points we then use a standard minimum distance estimator to find the set of parameters that best fit the moments of choice.[53] After bootstrapping the target moments in the data, we can also use the same model solutions to find standard deviations. Table 4 shows the estimated parameters and the corresponding moments in the simulated economy.[54]

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 that 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 log-skills in the economy is around its normalized value of zero. 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 5.1.3 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 lower in the model than in the data but we highlight that the estimated altruism factor $\delta = 0.475$ is in line with the literature (e.g., Manuelli and Seshadri, 2009; Abbott et al., 2019; Lee and Seshadri, 2019). Moreover, δ has a small standard deviation due to how much those transfers change in the model when the parameter δ is moved. A second parameter of interest is the substitutability between time and money investments given by $\gamma = -0.20$. This suggests that the elasticity of substitution is equal to 0.83.[55] Even though the standard deviation for this parameter is not as tight as others, Section 5.1.3

[53] Given that the distance distance between parameter sets may be large, we could also use these model solutions as starting points for a local estimator. We found this method to provide a good fit of the data without requiring the use of a local estimator.

[54] In addition to our target moments calculated with the full sample (call these M_0) in our data, we also calculate target moments M_n by bootstrap, for $n = 1, ..., N$. We obtain the set of parameters P_n that best fit each set of 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 $n = 1, ..., N$. A benefit of not using a local estimator as a second step in the estimation process is that we can use all the same original model solutions and simulations to calculate standard errors through bootstrapping without requiring the use of a local estimator for each set of target moments. See Appendix C.4 for more details.

[55] 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, heterogeneous 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 sets of parents who are more homogeneous (e.g., by education, income, number of children, or marital status), who are likely to be more similar in these other characteristics, the correlation

shows that our main results are almost unaffected when we move γ within one standard deviation.

Table 4: Estimation: parameters and moments

Parameter	Value	Std. Dev.	Description	Moment	Data	Model
Preferences						
μ	176.8	(9.12)	Mean labor disutility	Avg. hours worked	65.2	65.9
δ	0.475	(0.011)	Altruism	Parent-to-child transfer as share of avg. annual income	0.75	0.73
School Taste:						
α	5.38	(1.61)	Avg. taste for college	College share	33	30
α_{θ_c}	-0.55	(0.35)	College taste and cog. skills relation	College: cog skills slope	0.23	0.23
$\alpha_{\theta_{nc}}$	-1.15	(0.36)	College taste and non-cog. skills relation	College: non-cog skills slope	0.16	0.15
σ_ε	2.51	(0.46)	SD of college taste shock	College: residual variance	0.20	0.18
$\bar{\varepsilon}$	-1.55	(0.63)	Draw of school taste: mean by parent's education	Intergenerational persistence of education	0.70	0.75
Skill Formation Productivity:						
ξ	0.12	(0.03)	Parental time disutility of time with children	Avg. hours with children	18.0	17.2
$\bar{A}$	32.4	(1.30)	Returns to investments	Average log(skill)	0.0	0.0
α_m	0.91	(0.02)	Money productivity	Ratio of money to hours	218	183
γ	-0.20	(0.45)	Money-time substitutability	Money-time correlation	0.93	0.88
Interest rate						
ι $(\times 10^2)$	4.9	(1.22)	Borrow-save wedge	Share of borrowers	4.5	4.2
Government						
ω $(\times 10)$	2.05	(0.04)	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. For the money-time correlation, we group parents by deciles of time investments and compute average time and money investments for each decile. We then compute the correlation between the two averages. Share of borrowers is estimated from the 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 moment that is particularly informative for the corresponding parameter. See Appendix C.4 for more details.

4.3 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 Table 3, families from higher socioeconomic groups tend to invest more time and money in their children's development. The estimated model displays similar qualitative features. College-graduate parents in the

between time and money investments is generally high, leading to similar levels of complementarity in the estimation.

model spend on average \$752 more per child than high-school graduate parents. They also spend 4.5 more hours weekly with their children. These numbers are similar to our empirical estimates based on PSID-CDS and CEX data. The model tends to produce a larger regression coefficient of log-expenditures to log-income than the CEX data, which may be due to the fact that many (typically low-income) parents in the data report zero expenditures and they have to be dropped from this regression. Finally, the regression coefficient of log-hours to log-income is similar but slightly smaller in the model than in the data when using all types of families. These regression coefficients become even more similar if we use instead only families with two parents and two children.

Table 5: Validation: Non-targeted moments

Moment	Data		Model
Regression of parental investments to parents' characteristics (Table 3)			
	All Families	Homogeneous Families	
Hours on college ed. parent	3.7	2.5	4.5
Expenditures on college ed. parent	732	666	752
Log hours on log parent income	0.12	0.05	0.07
Log expenditures on log parent income	0.39	0.63	0.87
Intergenerational Mobility (Chetty et al, 2016 and PSID-CDS)			
Rank-Rank coefficient	0.26–0.29		0.29
Regression of college to log-parent income	0.24		0.18
Inequality (PSID)			
Gini	0.32		0.27
Top-Bottom	3.7		3.1
Savings (Inklaar and Timmer, 2013)			
Capital-Output Ratio (annualized)	≈ 3		2.8
Return to College (PSID and Heckman et al, 2006)			
Income Ratio: College – HS Graduate	1.6		1.7
Yearly return	≈ 10%		12%

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. In the model we use the average between the corresponding moments for children aged 0–3 and aged 4–7. All families refers to all children in the sample. Homogeneous families refers to children in households with two adults and two children.

Chetty et al. (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 0.29, within that range. Another measure of mobility is the correlation between college graduation and parents' income. Using PSID-CDS data, we regress a college graduation dummy to parents' log-income and obtain a coefficient of 0.24. Doing the same in the simulated model, we obtain a slightly lower but similar coefficient of 0.18. Labor income inequality is also well captured by the model: both the Gini and top-bottom coefficients are below but similar to the data.[56] The return to college in the model is similar to the empirical estimates in the data. The income ratio between college and high-school graduates in the model is 1.7 similar to the value found using PSID data (1.6). We also estimate the yearly return to attending college in the model (for the whole population and taking into account fees), which is another endogenous source of inequality, to be 12%, similar to the average estimates reported by Heckman et al. (2006).[57] Finally, regarding savings, the (annualized) capital-output ratio in the estimated model is 2.8 which is above but close to its typical empirical estimate of 3.

Using experimental evidence

We also use experimental evidence to test the validity of the most important novelty in our model: childhood development. Garcia et al. (2019) studied an RCT in which a small group of disadvantaged children were introduced into two high-quality early childhood development programs (ABC and CARE in North Carolina) that cost approximately \$13,500 per year.[58] The kids entered the program when they were around eight 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 into the early development of some children. This involves adding money directly into 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.[59] Although unexpected, parents are allowed to change their choices (including m and t) after g is introduced. Government spending g will crowd out parental money in-

[56]Top-bottom refers to the ratio of average incomes between the top 80th to 95th percentiles and those in the bottom 5th to 20th percentiles.

[57]We calculated the yearly return in the model to a college education by first evaluating the lifetime income change (in net present value) of becoming a college graduate for every individual in our simulated economy. We then subtract the college education cost, take the average across individuals and calculate the equivalent annual return.

[58]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. Given that our model period is 4 years long, we interpret the policy as providing g equal to \$54,000 (\$13,500 × 4 years) in the first period and \$13,500 (1 year) in the second period.

[59]If g was not a perfect substitute for m but provided an input different from m or τ, we would generally expect effects to be larger since in such a case the government would be able to provide something that parents are not able to provide (and hence, crowding out would be smaller). Effects would be smaller if g also crowds out τ, which in our formulation has to do with the estimated complementarity/substitutability between m and τ: The more substitutable, the more likely that g will not only crowd out m but also τ. In Section 5.1.3, however, we show that our main results are almost unaffected when we increase the elasticity of substitution by one standard deviation (from 0.83 to 1.33).

vestments 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.[60] Given the estimated complementarity between τ and m, parents in the model spend on average 5 more hours with their children after going through the intervention (the increase is even larger for lower-educated or lower-income parents).

The policy in the model is introduced with three specific characteristics in order to be comparable to that in 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 lower-educated and lower-income parents—Garcia et al. estimate that the average annual household earnings of the target population were approximately $5,000 (2000 USD). Therefore, in our model, we also want to compare the effect of such a policy on disadvantaged children. Finally, the children in 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 alone.

Garcia et al. observe the education and income of these children at two specific ages (the latest being age 30). Given that it is not trivial to choose an equivalent target population in our model, Figure 2 shows the effects in the model for different target populations. The three lines in each panel show the effect on all children, children of high-school-graduate parents, and children of high-school-graduate parents with low initial draws of skills (i.e., in the bottom third). Moreover, for each group, Figure 2 also shows the effects for children of parents with different levels of annual earnings. The dashed line shows the effects estimated by Garcia et al..[61]

Garcia et al. find that the policy led to an increase in the college graduation rate of approximately 15 percentage points. Our model, as shown by panel (a) of Figure 2, predicts large increases in this rate as well, particularly when focusing on low-income parents (recall that Garcia et al. estimate that the average annual household earnings of the target population were approximately $5,000). Panel (b) shows that income at age 30 increases by very similar amounts for intervened children in our model as does in Garcia et al.'s RCT. Garcia et al. 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 predicts similar returns, but possibly slightly smaller due to differences in the life-cycle-income profile.[62] Figure 2 also shows that more advantaged groups of

[60] 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 5), but in Section 5.1.4 we discuss an extension in which producing this early childhood input requires a college-graduate individual's time. Thus, a large-scale government intervention may affect the cost of early childhood development goods. We find, however, that this effect is counterbalanced by the increase in the supply of college graduates that the policy generates.

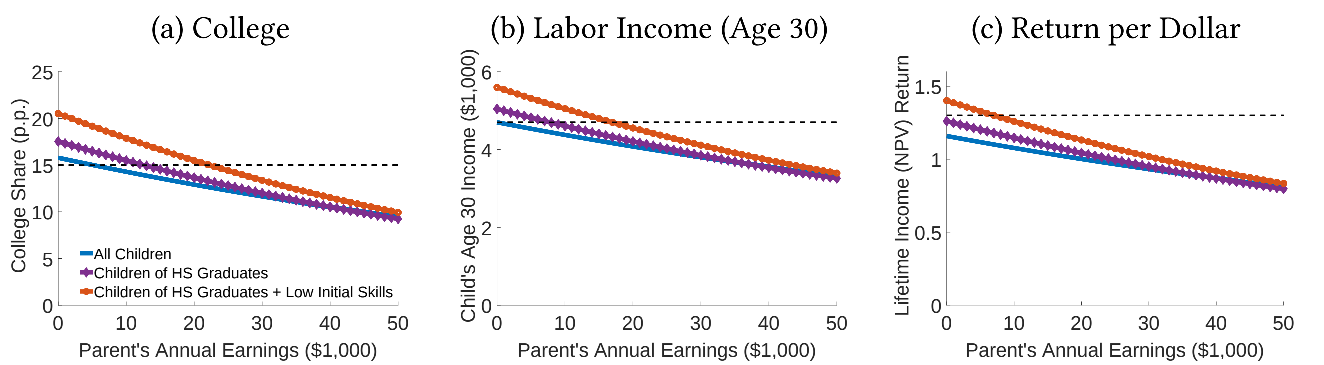
[61] Another way to check if our target population is in line with the one from Garcia et al., is to look at the outcomes of the control groups. If we focus on children with high-school-graduate parents whose annual earnings are approximately $5,000, college graduation rates are 7.5% and annual income at age 30 is approximately $30,000, close to the values reported in the RCT (8%–12% and $38,000, respectively).

[62] To compare our model predictions with those of a large-scale program (Head Start), we implemented a similar exper-

children (i.e., those with higher-educated or higher-income parents as well as those with higher initial levels of skills) tend to have smaller gains from the program. These mode-advantaged groups tend to have higher levels of parental investments so there is more room for crowding out when the policy is introduced. Even though Figure 2 shows that the lifetime labor income return can be smaller than 1, it is important to highlight that these are not the long-run gains, which, as we will show in Section 5, are much higher.

In Section 5 we study the effects of a similar (though large-scale) government investment policy, so the success of these validation exercises gives us confidence in the results we obtain in those counterfactuals.

Figure 2: Model Validation using RCT Evidence



Notes: 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, Heckman, Leaf, and Prados (2017). We simulate paying monetary expenses of a value of $13,500 per child-year in the first five years of childhood (unexpectedly, for only one cohort, and targeted to a small group of children). We evaluate alternative target samples regarding parents' characteristics and children's initial skills. Low initial skills refers to the bottom third at the time of birth. Policy effects are summarized using a second-degree polynomial. We then compare our findings regarding the effect on children's education and labor income to those of Garcia et al. (shown using the dashed line).

iment but using a lower amount of government investment g. Even though Head Start programs follow federal guidelines, they also vary across locations (Currie, 2001). Deming (2009) reports that costs per child-year vary between $5,800 and $7,400 (2000 USD). And even though most children only participate in Head Start for one year, some participate for two years. To capture a lower and upper bound, therefore, we studied the following two alternatives: (i) children get an investment of $g = \$5,800$ for one year during the first period of development, and (ii) children get an investment of $g = \$7,400$ for two years during the first period. To the best of our knowledge, empirical evidence on Head Start is less conclusive than from high-quality programs like the one studied by Garcia et al. (2019), with most papers finding positive results but with a relatively wide range of effects (e.g., Garces et al., 2002; Ludwig and Miller, 2007; Deming, 2009; Johnson and Jackson, 2019). In a recent paper, Johnson and Jackson (2019) finds that the average effects of Head Start on income and years of education for poor children to be approximately 10%. Our estimates (available upon request) suggest that our model predicts income increases of between 3% and 9%, suggesting that the estimated model is in line with or on the conservative side of these empirical estimates based on large-scale programs.

5 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; Pouliot, 2006). In particular, Loury (1981) provides a PE model in which government investments in childhood development 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 due to distortive taxation and GE 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 first evaluate alternative levels of government investment in early childhood development programs but in our analysis we focus in particular on the effects of introducing the same amount of investment g as in the case of the RCT studied by Garcia et al. (2019). We now study, however, a large-scale and long-run version of this policy so we need to take into account GE 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 driven 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.

We also study the effect of an alternative policy that focuses on developing skills through parenting education. The key difference is that here, rather than investing diretly in children's development, the program trains parents in how to promote their children's development. Estimating the returns and cost of this program is not trivial, but we confirm our main result from the early childhood investment program: welfare improvements in the long-run GE framework are larger than if we apply the policy as an RCT.

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 $V_{j=5}^{\tilde{P}}(a,\theta,\varepsilon,\lambda)$ be the welfare of agents with initial states (a,θ,ε) in the economy P if their consumption (and that of their descendants) were multiplied by $(1+\lambda)$:

$$V_{j=5}^{\tilde{P}}(a,\theta,\varepsilon,\lambda) = E^{P}\left\{\sum_{j=5}^{j=J_d}\beta^{(j-5)}u\left(c_j^P(1+\lambda),h_j^P\right)+\beta^{J_c}bV_{j'=5}^{\tilde{P}}(\varphi,\theta_k,\varepsilon',\lambda)\right\},$$

where for the sake of clarity we have abstracted in the presentation from including the school taste and child's consumption as well as from 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

$$\bar{V^P}(\lambda) = \int_{a,\theta,\phi} V_{j=5}^{\tilde{P}}(a,\theta,\varepsilon,\lambda)\,\mu_P(a,\theta,\varepsilon),$$

where μ_P refers to the distribution of initial states $\{a,\theta,\varepsilon\}$ 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 $V_{j=5}^{\tilde{P}}(a,\theta,\varepsilon,0)$, and (ii) changes in the probabilities of each state $\mu_P(a,\theta,\varepsilon)$. Having explained how we measure welfare, we now move forward with the policy evaluation.

5.1 Government Investments in 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 Figure 2, directly tested the mechanisms involved by introducing government investments as an RCT and using experimental evidence to compare the results. Its success gives us confidence in the policy evaluations we now perform. Differently from that validation, we now introduce government investments in a permanent and universal manner, i.e., investments are introduced for all children and forever. 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 GE 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 ages 0–3), evaluating this policy for different amounts of resources available. In Appendix G.1 we evaluate alternatives that use part of those resources to also invest in older children. We find that there are small additional gains from allocating part of the resources to invest in older children, but these gains are relatively small. Earlier investments lead to larger gains because the child's skills production function implies that skills are more malleable at younger ages.

5.1.1 Results

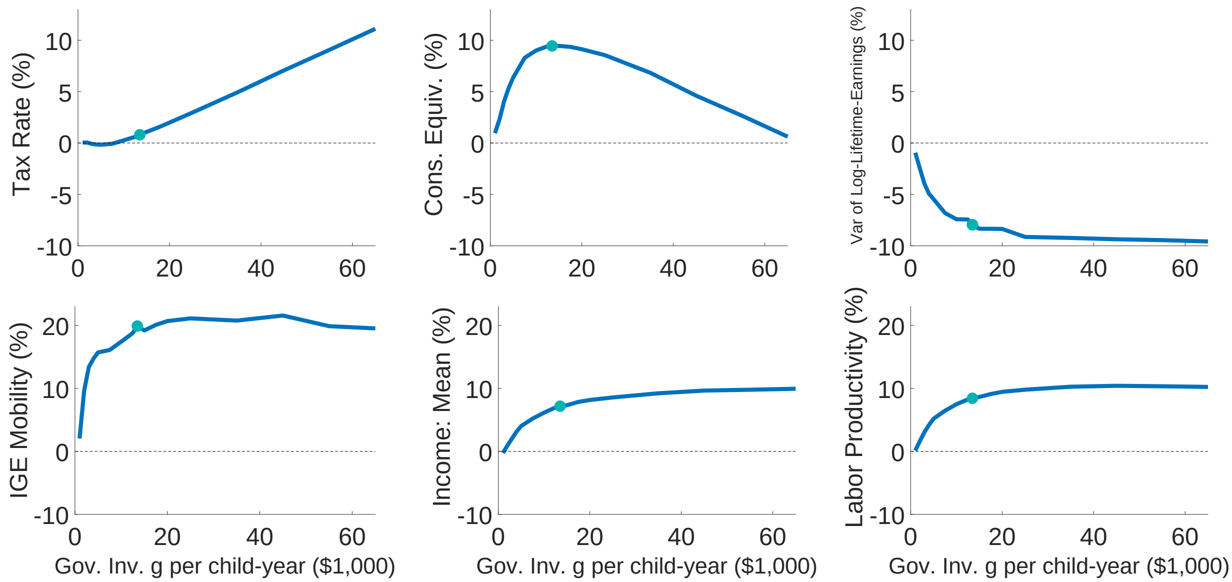
Figure 3 shows the long-run results of the policy, with the horizontal axis referring to the amount of government early childhood investments g per child-year. Introducing government childhood investments leads to substantial welfare gains of 9.4% as measured by the consumption equivalence for a newborn 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 Garcia et al. (2019) was estimated to cost approximately $13,500 per child-year (in 2000 USD). Given that this is a level of investment 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. Although we focus here on this policy, Appendix G.1 shows the results for other alternatives.

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 5.4% long-run welfare gain, approximately half of that 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 own childhood.[63]

[63] 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 used are more than $120,000 per child (i.e., $30,000 per child-year if using them only in the first period), returns on those investments are small relative to the large cost of raising taxes to afford them.

Figure 3: 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 7.2%, all of which is explained by the increase in labor productivity of 8.4%—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 et al. (2014)—would increase by 19.9%. This implies that the rank-rank coefficient is reduced by almost 0.06. Moreover, inequality—as measured by the variance of log lifetime earnings—would be reduced by 7.9%. These last two changes are approximately half as large as the difference between US and Canadian levels of immobility and inequality (Krueger et al., 2010; Chetty et al., 2014).[64]

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

[64] As an alternative benchmark, we can compute how much inequality and mobility would change if parental investments or skills were more equally distributed. If all parents invested in early childhood development as much as the average college-graduate parent, inequality would be reduced by 4.3%, while intergenerational mobility would increase by 11.1%. If, instead, all agents had the same average level of skills, inequality would be reduced by 15% and mobility would increase by 14%.

creates better skills for that child but also creates a better parental background for the following generation.

Table 6 decomposes the welfare gains by simulating the same policy applied in three alternative ways. The bolded bottom row refers to the benchmark results (i.e., in the long run taking into account tax changes and GE effects). In the first row, we introduce government investments for only one generation, without balancing the government's budget or taking into account GE effects. Effects are evaluated on the generation that receives the intervention so that this implementation can be compared to that expected from an RCT, which is typically of small scale and applied to only one generation. Unlike 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 3.9%, less than half of our benchmark. Mobility increases by only one-half of the benchmark increase, while inequality actually increases instead of decreases.[65]

Table 6: Results decomposition

Alternative Exercises			**Change from Baseline (%)**				
Long Run	**General Equilibrium**	**Budget Balanced**	**Consumption Equivalence**	**Average Income**	**Labor Returns**	**Inequality**	**Mobility**
No	No	No	3.9	8.0	8.4	5.3	12.6
Yes	No	No	9.1	11.7	13.4	5.6	25.4
Yes	No	Yes	9.5	11.6	13.4	5.2	24.7
Yes	Yes	No	10.2	7.2	8.6	-7.7	20.2
Yes	Yes	Yes	**9.4**	**7.2**	**8.4**	**-7.9**	**19.9**

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 the variance of log-lifetime-earnings, 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.

In the second row, we allow for long-run effects to take place: the policy is implemented permanently, taking into account the fact that by improving one generation's level of skills we are also improving the productivity of future investments. By permanently introducing this policy, the chances of children being born into a low-skilled family are reduced. Welfare gains increase by 5.2 p.p. (from 3.9% to 9.1%), suggesting that long-run intergenerational dynamics generate over one-half of the baseline welfare gains of 9.4%. Similarly, intergenerational mobility increases by over one-half of the baseline mobility increase (from 12.6% to 25.4%). Inequality, however, still is not reduced.

[65]In this scenario, there is a shift towards more college graduates (from 30% to 37%) which, holding average income by education group relatively constant, increases income dispersion.

The next two rows show the effects of GE forces and taxation. The third row shows that in PE (i.e., holding prices constant), the policy is self-financing in the long run because of the large increase in labor income, and even leads to tax reductions. This leads to an increase in welfare gains from 9.1% to 9.5%, while other elements are almost unaffected. The fourth row shows that it is GE forces that generate most of the reduction in inequality: increasing skills augments the share of college graduates, which reduces the wages of college graduates relative to high school graduates. Even though reducing the wages of college graduates reduces the labor productivity gains associated with this policy, GE forces reduce inequality and increase welfare gains by over one-tenth of the baseline value if we abstract from the fact that taxes need to be increased once GE forces are taken into account (i.e., going from the second to the fourth row). Finally, taking into account that, in GE, taxes need to be increased to finance this policy reduces welfare gains by almost one-tenth.[66]

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 one-half of the welfare gains and the increase in intergenerational mobility. This suggests that these gains may take a long time to accrue, but the transition dynamics studied below formally evaluate this concern. By reducing the wages of college graduates relative to high school graduates, GE effects reduce labor productivity but also reduce inequality, which increases welfare gains by over one-tenth. Finally, even though the policy can be self-financing in PE, once the reduction in college-graduate wages is taken into account the policy requires an increase in taxes to balance the government budget, which reduces welfare gains by almost one-tenth.

Sources of welfare gains

Recall that, by definition, welfare gains emerge from two sources: (i) changes in the value of an agent at each state $V_{j=5}(a, \theta, \varepsilon)$, and (ii) changes in the distribution over those states $\mu(a, \theta, \varepsilon)$. 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, ε). 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. Assume that an agent knows her parent's skills level and education group, and this is unchanged between the baseline economy and the one in which the policy is in place.[67] Children of high-school-educated parents benefit the

[66] We find that most long-run welfare gains are driven by the level effect of the reform on average consumption. As discussed by Bénabou (2002) and elaborated for non-homothetic utility functions in Abbott et al. (2019), 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 more than eight-tenths of the total welfare gains are driven by the level effect.

[67] 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

most, with a welfare gain of up to 8%. On the other hand, agents that know they are going to be born to college-graduate and high-skilled parents benefit less, with gains of approximately 5%. The welfare gains for each of these groups is 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. This reduction leads to changes in the distribution $\mu(a, \theta, \varepsilon)$ towards states associated with higher utility $V_{j=5}(a, \theta, \varepsilon)$. 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 $V_{j=5}$; and (ii) fixing the values $V_{j=5}$ and only considering the changes in the distribution μ. If we fix the distribution μ to the original steady state, welfare gains are 2.5%, i.e., one-fourth of the total gains. On the contrary, if we fix the values $V_{j=5}$ to the original steady-state, we find welfare gains of 7.3%, 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 utility at each state.

5.1.2 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 possibility by looking at the transition dynamics.[68]

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 from the new steady state, and are known 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.[69] Prices (wages and the real interest rate) adjust every period to clear the markets.[70] Figure 4 shows the effects on all new cohorts, with cohort 0 being the first cohort to be intervened.

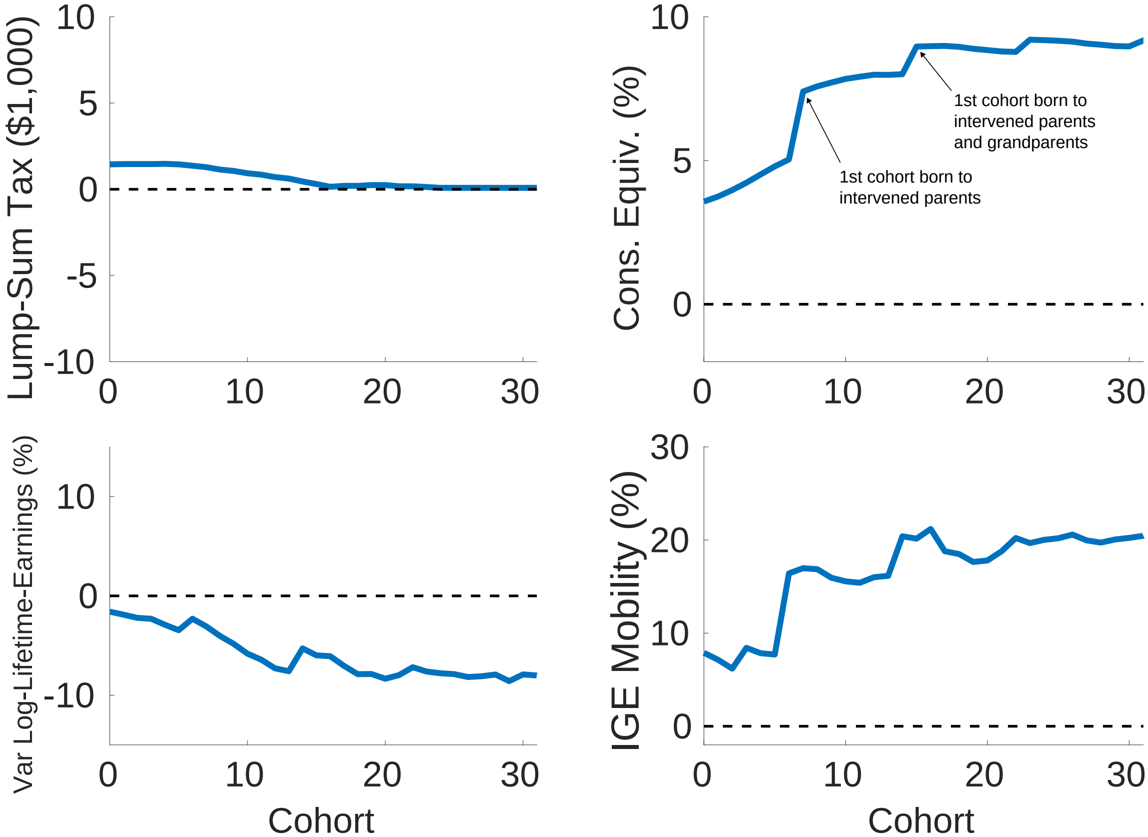
that the parent's distribution over assets and idiosyncratic labor productivity (conditional on education and skills) differs between those two economies.

[68]Studying transition dynamics, however, is complicated because there are many ways the policy change can be implemented (e.g., determining 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 et al. (2015) study 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.

[69]We have also examined a policy in which the labor income tax, instead of an additional lump-sum tax, is adjusted every period to balance the budget. Results are relatively similar, but we focus on the lump-sum tax case in the transition analysis for computational reasons. See Appendix D.1.1.

[70]The equilibrium along the path is similar to the one in the steady state (as defined in Appendix B) but adds time subscripts to all decision rules, value functions, aggregate inputs, prices, tax policy, and distributions.

Figure 4: 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 each cohort and their parents. All values are relative to the initial steady state.

The first cohort to receive the government investments, as shown by the top-right panel of Figure 4, 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 more than 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 for the first cohort born to intervened parents and intervened grandparents. They obtain 9% of welfare gains. 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.[71]

The policy is associated with a steady decrease in income inequality that is paralleled by the change

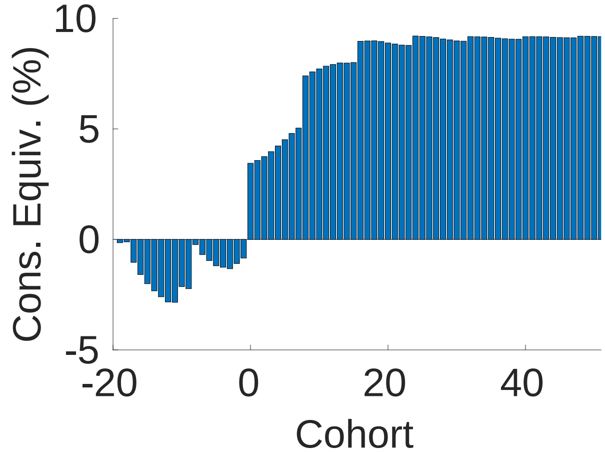
[71] The small changes in gains for cohorts born between the jumps are due to the slow adjustment in prices and lump-sum taxes.

in wages. Appendix Figure D1 shows that wages of college graduate steadily decrease by up to 5%, 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 $1,400 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 80 years it is almost eliminated. Appendix Figure D1 also shows that as the policy is introduced, interest rates increase because agents optimally reduce savings because future generations are more likely to be better off—due to both higher skills and lower lump-sum taxes.

Older generations alive at the time the policy is introduced lose. Figure 5 shows that the welfare losses for each cohort alive at that time are between 0 and 3%. Individuals aged 44 and above (i.e., cohorts -10 and under) 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 less than 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 older individuals alive at the time the policy is introduced. This way of financing the transition highlights, once again, the main missing market in the economy: the government borrows to invest in early childhood development and finances these investments by raising the taxes these intervened children need to pay as adults.

Figure 5: 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 accordingly computed for agents of age 40 at the time.

5.1.3 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 in early childhood just as in the previous section and calculate the welfare gains. Table 7 reports the gains in the short-run (i.e., for the first generation) PE case when taxes are not adjusted (similar to an RCT applied to a small representative family) as well as for the long-run GE case when labor income taxes are adjusted such that the government's budget is rebalanced. 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 achievable—by moving each parameter between plus/minus one standard deviation—is at most one-sixth of the total gains. More important, our main result that gains in the long-run GE case are substantially larger than those in the short-run PE case seems robust to these changes.

It is instructive, however, to analyze the effect of some parameters. First, the larger the parental altruism δ, 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 school taste. Larger values of α, related to the average disutility from attending college, 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, 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 7: Welfare gains robustness to estimated parameters

	Cons. Equiv. Change from Baseline					
	Short-Run PE			**Long-Run GE**		
	Down	Up	**Total**	Down	Up	**Total**
δ	0.06	-0.02	0.09	0.34	-0.19	0.53
μ	-0.01	0.00	0.01	-0.13	-0.06	0.07
α	0.06	-0.22	0.28	-0.66	0.81	1.47
α_{θ_c}	0.09	-0.12	0.21	0.00	-0.56	0.56
$\alpha_{\theta_{nc}}$	0.01	-0.02	0.03	-0.13	-0.14	0.01
$\bar{\varepsilon}$	-0.01	-0.02	0.01	-0.21	-0.20	0.02
σ_ε	-0.16	0.03	0.19	0.70	-0.78	1.48
$\bar{A}$	0.01	-0.02	0.02	-0.11	-0.23	0.11
α_m	-0.05	0.04	0.10	-0.38	-0.02	0.36
γ	-0.00	-0.04	0.03	-0.21	-0.20	0.01
ξ	-0.00	-0.00	0.00	-0.19	-0.21	0.02
ι	-0.00	0.00	0.00	-0.07	-0.19	0.12
ω	0.02	-0.02	0.04	-0.09	-0.27	0.17
Baseline	**3.9**			**9.4**		

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 each parameter may affect the results. We solve the model for each given parameter set, introduce the same policy from Section 5.1.1, and report the consumption equivalence welfare gains. Short-Run PE refers to the short-run PE gains without adjusting taxes and Long-Run GE refers to the long-run GE gains adjusting labor income taxes such that the government's budget is balanced.

Given the importance of the child's skills production function in the model, we also study how sensitive our results are to changes in those parameters. Recall that the values used here are from Cunha et al. (2010), so we move parameters according to their reported standard deviations. We move sets of 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.[72] Table 8 reports the change in welfare gains from

[72]We group two sets of parameters that have similar interpretations: α_1 and α_2 relate to the persistence of children's skills, while α_3 and α_4 relate to the importance of parents' skills. It is important to note that this exercise, in the case of the α components, is also changing the sum of the α_s to be lower or higher than one. An alternative would be to compensate the change of the increase/decrease of one component by adjusting some or all of the other ones. However, this would make the exercise harder to implement and even to interpret since we would need to subjectively choose which components to adjust. We note that even though our implementation does seem to change the total productivity as well as the relative productivity of each factor, we estimate each case such that the average level of log-skills is normalized to zero.

introducing the same government investment policy.

Table 8: Robustness to child's skills production function

	Change from Baseline					
	Cons. Equiv. SR-PE			**Cons. Equiv. LR-GE**		
	Down	Up	**Total**	Down	Up	**Total**
α_1,α_2	0.51	-0.56	1.07	1.64	-2.70	4.34
α_3,α_4	0.48	-0.44	0.92	0.98	-1.48	2.46
α_5	0.11	-0.20	0.31	0.03	-0.89	0.92
ρ	-0.32	0.39	0.71	-1.26	0.96	2.21
σ_ν	0.18	-0.08	0.26	0.07	-0.66	0.73
$Var\left(\theta_{k_0}\right)$	-0.06	-0.07	0.01	-0.66	-0.67	0.01
$Corr\left(\theta,\theta_{k_0}\right)$	-0.06	-0.07	0.00	-0.69	-0.44	0.25
Baseline		**3.9**			**9.4**	

We move each parameter by one standard deviation as reported by Cunha et al. (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 each parameter may affect the results. We re-estimate the model for each given set of parameters, introduce the same policy from Section 5.1.1, 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 those internally estimated as reported in Table 7. 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-fourth, i.e., keeping long-run welfare gains always above 6.7%. As suggested by Cunha et al., 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 $\alpha_1$and α_2 are (the more persistent children's skills are), the harder it is for investments to be effective. Similarly, the larger α_3 and α_4 are (the more important parental characteristics are), the more ineffective government investments are.[73] An 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 skills

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

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 et al.'s estimates may be biased. In particular, they suggest that estimates of ρ may be biased towards zero and estimates of α_1 and α_2 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 7, 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 7 shows results change by at most one-half for a two-standard-deviation change in any single parameter, so only very large biases are likely to significantly affect our main results. More importantly, our main result that long-run GE welfare gains are larger than those estimated in the short-run PE case seems robust to such large changes as well.

5.1.4 Extension: With Early Childhood Education Market

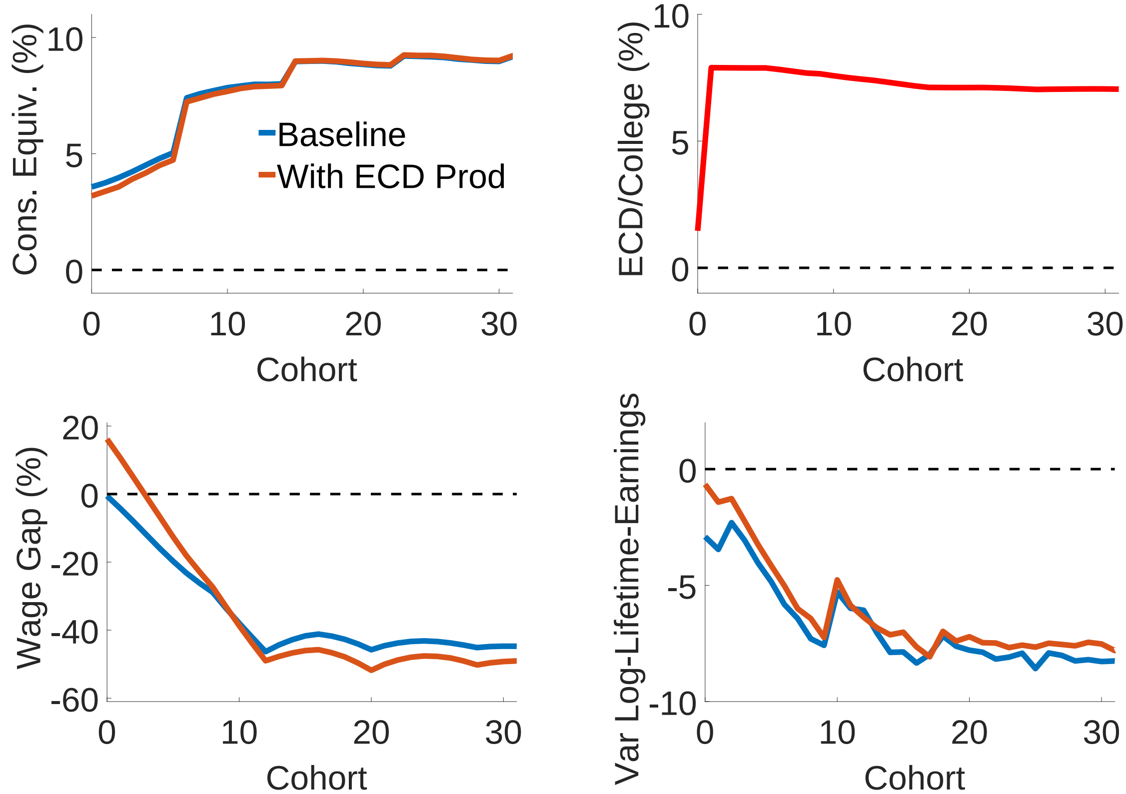
In the baseline model we assume that the early childhood money input is equal to the good produced in the economy. This constant returns to scale assumption misses that the elements required to produce this early childhood development input may be scarce, particularly when the policy is scaled up. In this section we provide a simple extension to the model in which this early childhood input is actually hours spent with a college-graduate individual—which is in line with the costs reported for the RCT program on which we base our main analysis (Garcia et al., 2019).

The price of early childhood education is now given by the wage of college graduates, 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 6 shows the effects of this policy, highlighting the main differences relative to the baseline case.

Figure 6: 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 the gap between the equilibrium wages for college-graduate and high-school graduate 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-graduate labor is a scarce input. On the other hand, over time the policy itself will increase the share of college graduates, 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-graduate 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).[74]

[74] In the long run, the crucial question 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.

5.2 Parenting Education Program

We now 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 in 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 RCT evidence from Gertler et al. (2014) and Attanasio et al. (2016). A relevant caveat is that this evidence is from developing countries. To the best of our knowledge, this is the best evidence available on the costs and returns of such programs so we are limited by this fact, but we try to adjust for this limitation by showing how results change for alternative estimates of returns. We implement this program as a government policy (both in a PE framework similar to an RCT and in GE) as well as a new education good that parents can acquire in a private market. Importantly, once again we find that welfare improvements in the long-run GE 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.

5.2.1 Introducing and Estimating Parenting Education in the Model

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 $\theta_{PE} = \left\{\theta_{PE,c}, \theta_{PE,nc}\right\}$ in an extended version of children's development function (6),

$$\begin{aligned} \theta'_{q,k} &= \left[\alpha_{1,q,j}\theta_{c,k}^{\rho_{q,j}} + \alpha_{2,q,j}\theta_{nc,k}^{\rho_{q,j}} + \alpha_{3,q,j}\max\left\{\theta_c, \theta_{PE,c}\right\}^{\rho_{q,j}} + \alpha_{4,q,j}\max\left\{\theta_{nc}, \theta_{PE,nc}\right\}^{\rho_{q,j}} + \alpha_{5,q,j}I^{\rho_{q,j}}\right]^{1/\rho_{q,j}} e^{\nu_q} \\ \nu_q &\sim N(0, \sigma_{q,j,\nu}), \quad q \in \{c, nc\} \qquad (8) \\ I &= \bar{A}\left[\alpha_m m^{\gamma} + (1-\alpha_m)\tau^{\gamma}\right]^{1/\gamma}, \end{aligned}$$

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 $\theta_{PE,q}$ entered as a perfect substitute for θ_q), we would expect general welfare benefits to be even larger. Therefore, our results here may be thought of as a lower bound in that respect.[75]

[75] An alternative compelling interpretation of the effect of parenting education would be that it increases $\bar{A}$. Effectively, both alternatives increase the derivative of $\theta'_{q,k}$ to τ or m (i.e., the productivity of investments). The key element is how to benchmark the increase in 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 the cost of and returns to parenting education (in terms of θ_{PE}) is not easy, so we take the approach below in order to estimate a lower bound on the benefits of such a policy. Although we would like to estimate cost and returns from parenting education programs in the US, 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 from two programs with similar curricula. Based on the upper-bound available for an RCT in Colombia (Attanasio et al., 2016), we estimate the cost of running such policy in the US to be $11,400 per family in the first period with children.[76] 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 et al. (2014). Parents of growth-stunted children were randomly selected to participate in the program when their children were between 0 and 2. By the time these children were approximately 22 years old, Gertler et al. (2014) estimate that parenting education program had led to a 12% increase in the children's income.[77] As shown in Table 9, 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.[78] This is obtained by θ_{PE} that is one standard deviation above the average θ (for both cognitive and non-cognitive skills). We refer to this value as the benchmark return to parenting education, but we report results for various levels of θ_{PE}.

estimating a lower bound on the potential impact of parenting education.

[76]Running a similar policy in Colombia was estimated to cost between $450 and $750 per child (Attanasio et al., 2016). This program actively employed a group of women ("Madres Lideres") with average education equivalent to a high school degree. If we assume that running the program in the US would require similar inputs, 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 a cost, we assume that in the US they would employ college graduates instead. In Colombia, $450 represented approximately the average monthly salary of a high school graduate. Assuming this would require a college graduate in the US (whose average salary 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.

[77]We focus on the estimates for earnings on the current job that exclude individuals who migrated to other countries (Table S.14 in the Appendix of Gertler et al. (2014)). This estimate is smaller than the average finding for all individuals of 25%. Thus, consistent with our other choices, we are likely to obtain a lower bound on the gains of such programs.

[78]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 high school graduates whose income is in the bottom 5%.

Table 9: Parenting Education: benchmarking θ_{PE}

θ_{PE} relative to Avg. θ	**θ_{PE} relative to Policy benchmark**	**Change from Baseline (%) Income Bottom**
-1.6 SD	-2.6 SD	0.00
-1.4 SD	-2.4 SD	0.11
-1.2 SD	-2.2 SD	1.11
-1.0 SD	-2.0 SD	2.13
-0.8 SD	-1.8 SD	3.07
-0.6 SD	-1.6 SD	4.06
-0.4 SD	-1.4 SD	5.22
-0.2 SD	-1.2 SD	6.33
0.0 SD	-1.0 SD	7.22
+0.2 SD	-0.8 SD	8.40
+0.4 SD	-0.6 SD	9.48
+0.6 SD	-0.4 SD	10.63
+0.8 SD	-0.2 SD	11.48
+1.0 SD	**0.0 SD**	**12.31**
+1.2 SD	+0.2 SD	13.10

We use the estimated model (starting from steady state) to simulate experimental evidence from a parenting education program in the spirit of Gertler et al. (2014). We simulate low-income, low-skilled parents with low-skilled children going through a 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.)

We now evaluate parenting education in three steps. The first two steps estimate what the return would be if the government were to run a program of parenting education, in a short-run PE as well as a long-run GE scenario. Here the government enrolls (and pays for) every agent to obtain parenting education, independent of whether or not it is ineffective for them—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.

5.2.2 Parenting Education as a Government Program

We introduce parenting education in the previous steady state and evaluate the effects on the relevant cohort receiving those benefits: children born to the generation receiving the parenting education. This environment is useful to understand the first-order, 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 10: Parenting Education Program

θ_{PE} relative to benchmark	Change from Baseline (%) Cons. Equiv.	Avg. Income	Inequality	Mobility	College	Tax Revenue	Tax Rate
	Short-Run Partial Equilibrium						
-1.8 SD	0.85	1.86	1.55	4.45	5.37	2.42	0.00
-1.6 SD	1.13	2.43	2.04	5.79	6.95	3.18	0.00
-1.4 SD	1.38	3.02	2.59	7.09	8.46	3.91	0.00
-1.2 SD	1.63	3.62	3.23	8.55	9.95	4.64	0.00
-1.0 SD	1.86	4.17	3.79	9.77	11.33	5.33	0.00
-0.8 SD	2.38	5.22	4.89	11.56	14.00	6.83	0.00
-0.6 SD	2.84	6.18	5.88	12.91	16.39	8.20	0.00
-0.4 SD	3.29	7.09	6.79	14.61	18.65	9.50	0.00
-0.2 SD	3.69	7.92	7.54	15.99	20.70	10.69	0.00
Benchmark	**4.06**	**8.66**	**8.21**	**16.98**	**22.57**	**11.78**	**0.00**
0.2 SD	4.40	9.34	8.79	17.83	24.28	12.78	0.00
	Long Run General Equilibrium						
-1.8 SD	1.77	1.43	-2.72	6.79	1.49	2.22	-0.08
-1.6 SD	2.32	1.98	-2.70	8.26	2.31	2.49	-0.18
-1.4 SD	2.87	2.29	-3.12	9.29	2.61	2.60	-0.28
-1.2 SD	3.27	2.59	-3.61	10.16	2.90	2.72	-0.35
-1.0 SD	3.79	2.85	-4.29	11.03	3.32	2.93	-0.44
-0.8 SD	4.92	3.60	-4.46	11.97	4.55	3.08	-0.69
-0.6 SD	5.48	4.36	-4.79	13.85	5.00	3.39	-0.76
-0.4 SD	6.29	4.86	-5.30	14.30	5.55	3.43	-0.94
-0.2 SD	6.95	5.39	-4.98	15.32	6.30	3.64	-1.05
Benchmark	**7.65**	**5.68**	**-5.14**	**15.47**	**6.40**	**3.95**	**-1.16**
0.2 SD	8.19	6.05	-5.35	16.70	6.87	4.06	-1.26

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. Long-run general equilibrium 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. Short-run partial equilibrium focuses on the effects on the children of the first (and only) cohort of parents 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 10 shows the results in the short-run PE case. Effects are reported as percentage 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_c and θ_{nc}) 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 an $800 increase in the tax revenue per household each year. Thus, parenting education, assuming the policy is at least as effective as 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).

Next, we evaluate the parenting education program in a long-run GE environment: it is introduced permanently and we consider the economy in the new steady state. We adjust government labor income 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 10 shows the results.

Parenting education remains highly beneficial. Once again, even for the cases in which the policy is 1.8 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 PE 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 approximately 0.05 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 PE. Parenting education also proves to be a policy that would increase tax revenue in the long run. By increasing the share of high-skilled individuals, the growth in income, consumption, and savings even allows the government to reduce labor tax rates.

5.2.3 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 child is born they have one more choice to make: whether or not to acquire parenting education at the price of $11,400 (same as the government estimate). Table 11 shows the results for such an exercise, for different levels of productivity θ_{PE} relative to the estimated benchmark. At the benchmark the consumption equivalence is 7%, which is associated with a 100% take-up among the low-skilled parents. Intergenerational mobility would increase by 13%. Similarly, average income would increase by 5.4%, while inequality would be reduced by 4.8%. Relative to the benchmark, the productivity of the program would need to be far below the benchmark for the take-up to be 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 11: Parenting Education Market: Long-Run General Equilibrium Effects

	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	0.46	0.69	-0.66	1.39	0.13	0.25	-0.09	45.82	0.00	0.00
-1.6 SD	1.27	1.15	-1.90	4.06	0.66	0.32	-0.28	79.12	0.00	0.00
-1.4 SD	1.61	1.66	-2.08	5.63	1.47	0.45	-0.35	82.54	0.00	0.00
-1.2 SD	2.13	2.13	-2.51	6.18	1.87	0.78	-0.43	87.22	0.00	0.00
-1.0 SD	3.15	2.75	-2.72	6.54	2.49	0.78	-0.68	93.93	0.00	0.00
-0.8 SD	4.19	3.43	-3.71	9.09	3.96	1.12	-0.89	98.84	0.00	0.00
-0.6 SD	4.87	3.87	-4.20	10.42	4.23	1.47	-0.98	100.00	0.00	0.00
-0.4 SD	5.51	4.31	-4.74	11.36	4.91	1.60	-1.12	100.00	4.09	0.00
-0.2 SD	6.28	4.82	-5.29	11.90	5.58	1.72	-1.28	100.00	19.80	0.00
Benchmark	**7.02**	**5.43**	**-4.85**	**13.40**	**6.45**	**1.82**	**-1.44**	**100.00**	**33.41**	**0.00**
0.2 SD	7.64	5.95	-5.16	13.17	6.91	2.10	-1.54	100.00	50.17	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-lifetime-earnings 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 a market good that households can choose 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 those in the 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 GE) 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 paid for by the government.

6 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 GE 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 early childhood development directly. Introducing universal government investments in early childhood development (e.g., mandatory schools for children age 3 and younger) leads to a long-run reduction in lifetime-earnings inequality of 8% and an increase in intergenerational mobility of 20%. These changes are approximately half as large as the difference between US and Canadian levels of inequality and mobility.[79] This policy yields long-run welfare gains (in consumption equivalence terms) of 9.4%.

These welfare gains are twice those obtained by introducing the same early childhood program as a short-run PE policy—similar to those studied in RCTs. The long-run change in the distribution of parental characteristics generates most of the total welfare gains. 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., gender, single-parent families, or endogenous 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 effects would remain true in models with richer family heterogeneity (as long as parents remain important for early childhood development). It would also be interesting to link early childhood development with college major choices, since Arcidiacono et al. (2016) show that college majors are associated with pre-college skills. Finally, choosing optimal early childhood policies that take the transition fully into account would be particularly interesting given that welfare gains are heterogeneous by cohorts. We have explored some alternatives including government borrowing and slow introduction of early childhood investments, but we believe more research is necessary.

[79] As an alternative benchmark, if all agents had the same average level of skills, inequality would be reduced by 15% and mobility would increase by 14% in our estimated model.

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 (2019). 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.

Agostinelli, F. and M. Wiswall (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.

Attanasio, O. P., C. Fernández, E. O. Fitzsimons, S. M. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2014). Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in colombia: cluster randomized controlled trial. *Bmj 349*, g5785.

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.

Bakış, 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.

Barr, A. and C. R. Gibbs (2019). Breaking the cycle?: Intergenerational effects of an anti-poverty program in early childhood. Working paper.

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. J. Lochner (2019). Early and late human capital investments, borrowing constraints, and the family. *Journal of Political Economy*.

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.

Cunha, F. (2013). Investments in children when markets are incomplete. Working paper.

Cunha, F., I. Elo, and J. Culhane (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Technical report, National Bureau of Economic Research.

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

Currie, J. (2001). Early childhood education programs. *Journal of Economic perspectives 15*(2), 213–238.

Daruich, D. and J. Kozlowski (2019). Explaining intergenerational mobility: The role of fertility and family transfers. *Review of Economic Dynamics*.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *The Review of Economic Studies 81*(1), 137–185.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics 1*(3), 111–34.

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*, Volume 2 of *Handbook of Macroeconomics*, pp. 1789–1891. Elsevier.

Elango, S., J. L. García, J. J. Heckman, and A. Hojman (2015). Early childhood education. In *Economics of Means-Tested Transfer Programs in the United States, Volume 2*, pp. 235–297. University of Chicago Press.

Fernandez, R. and R. Rogerson (2003). Equity and resources: An analysis of education finance systems. *Journal of Political Economy 111*(4), 858–897.

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.

Garces, E., D. Thomas, and J. Currie (2002). Longer-term effects of head start. *American economic review 92*(4), 999–1012.

Garcia, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2019). Quantifying the life-cycle benefits of an influential early childhood program. *Journal of Political Economy*.

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., F. Perri, and G. L. Violante (2010). Unequal we stand: An empirical analysis of economic inequality in the united states, 1967–2006. *Review of Economic dynamics 13*(1), 15–51.

Heckman, J. J. and G. Karapakula (2019). Intergenerational and intragenerational externalities of the perry preschool project. Technical report, National Bureau of Economic Research.

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.* , Volume 1 of *Handbook of the Economics of Education*, Chapter 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*, 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.

Johnson, M. T. (2013). Borrowing constraints, college enrollment, and delayed entry. *Journal of Labor Economics 31*(4), 669–725.

Johnson, R. C. and C. K. Jackson (2019). Reducing inequality through dynamic complementarity: Evidence from head start and public school spending. *American Economic Journal: Economic Policy 11*(4), 310–49.

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.

Krueger, D., F. Perri, L. Pistaferri, and G. L. Violante (2010). Cross-sectional facts for macroeconomists. *Review of Economic dynamics 13*(1), 1–14.

Lee, S. Y. and A. Seshadri (2019). On the Intergenerational Transmission of Economic Status. *Journal of Political Economy*.

Lobman, C., S. Ryan, and J. McLaughlin (2005). Reconstructing teacher education to prepare qualified preschool teachers: Lessons from new jersey. *Early Childhood Research & Practice 7*(2), n2.

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

Ludwig, J. and D. L. Miller (2007). Does head start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly journal of economics 122*(1), 159–208.

Manuelli, R. E. and A. Seshadri (2009, May). 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. Technical report.

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

Richter, L. M., B. Daelmans, J. Lombardi, J. Heymann, F. L. Boo, J. R. Behrman, C. Lu, J. E. Lucas, R. Perez-Escamilla, T. Dua, et al. (2017). Investing in the foundation of sustainable development: pathways to scale up for early childhood development. *The Lancet 389*(10064), 103–118.

Saez, E. (2017). Taxing the rich more: Preliminary evidence from the 2013 tax increase. *Tax Policy and the Economy 31*(1), 71–120.

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 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). 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. Using this data, we estimate parental time investments towards children's development (to be used in the model estimation) and show evidence that time with children towards their development are associated with parents' education and income (to be used in the model validation).

Sample Selection: The results presented in Section 4 are for two samples of children. 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., 2,556 families and 4,024 children. This is the first sample we use, i.e., 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. If we restrict the sample to families with two children, this reduces the number of families and children in the sample to 1,020 and 1,546.[80] Similarly, restricting the sample to children whose parents remain married between their time of birth and age 12, reduces the sample to 500 families and 796 children. 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 the initial sample that includes all families.

[80]Note that this does not require that we have information on both children, it only requires that the family has two children.

Table A1: CDS Summary Statistics: Whole Sample

Age Group	**3-7**	**8-12**	**13-18**
Letter-Word Score	16.7	41.1	48.0
	(11.0)	(7.6)	(5.8)
Applied-Problems Score	16.6	33.9	41.0
	(7.7)	(6.2)	(6.8)
Child's Age	5.3	10.2	15.1
Mother's Age	32.1	37.3	42.3
Father's Age	35.2	40.2	45.3
Mother's Education (years)	14.2	14.1	14.1
Father's Educarion (years)	13.9	13.9	13.8
Mother Works	79.0	83.3	84.1
Father Works	66.0	61.8	61.1
Mother's Work Hours (weekly)	24.4	27.0	28.8
Father's Work Hours (weekly)	28.8	26.9	26.5
Mother's Hourly Wage	17.5	18.5	17.7
Father's Hourly Wage	23.7	27.1	27.6
Family's Total Income	1,262.1	1,377.8	1,599.8
Number of Children	1,721	2,421	2,061

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

Table A2: CDS Summary Statistics: Selected Sample

Age Group	**3-7**	**8-12**	**13-18**
Letter-Word Score	18.1	43.3	50.5
	(11.4)	(6.6)	(3.8)
Applied-Problems Score	18.0	36.0	44.0
	(7.2)	(6.0)	(6.0)
Child's Age	5.3	10.2	15.1
Mother's Age	34.7	39.4	44.3
Father's Age	36.8	41.4	46.2
Mother's Education (years)	15.0	14.8	14.7
Father's Educarion (years)	14.6	14.5	14.5
Mother Works	76.1	83.9	89.4
Father Works	97.2	94.7	87.2
Mother's Work Hours (weekly)	22.9	26.2	29.5
Father's Work Hours (weekly)	42.8	41.1	38.2
Mother's Hourly Wage	22.8	21.1	20.7
Father's Hourly Wage	27.4	32.1	31.5
Family's Total Income	1,871.9	2,180.1	2,521.0
Number of Children	348	535	475

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

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 more details on the time investments estimated using this data.

Parental time 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.[81] 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. We find that that there are small differences in total time with children between 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. Even though lower-educated or lower-income parents spend similar total amounts of time with their children as college-graduate or higher-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 et al., 2004; Swing et al., 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).[82] Table 3 shows that college graduates spend 3.7 more hours per week than high-school graduates. Similarly, parents in the top income quintile spend 6 more quality hours per week with their children than parents in the bottom quintile.

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 et al. (2010), where the authors focus on identification and reducing 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.

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

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

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 representative sample equal to 13,055. The distribution across education groups within this starting point is:

- High-school dropouts: 1,482 Individuals, with a total number of observations of 12,052.
- High-school graduates: 7,288 Individuals, with a total number of observations of 61,515.
- College graduates: 4,135 Individuals, with a total number of observations of 38,217.

Restricting to households with two adults, the number of individuals is reduced to 10,851. We keep individuals with at least 12 observations of income and we further restrict observations to those with positive hours of labor in the household (but lower than 10,000 annually), leading to 2,533 individuals. We also drop individuals who at least once report hourly wages under $1 or above $400 and those who report extreme changes of income (i.e., annual growth above 400%, or reduction by 66%. This leads to a final number of individuals of 2,501. These are distributed as follows:

- High-school dropouts: 276 Individuals, with a total number of observations of 1,655.
- High-school graduates: 1,359 Individuals, with a total number of observations of 9,511.
- College graduates: 865 Individuals, with a total number of observations of 6,510.

A.3 Income Profile: NLSY Sample Selection

We start with 6,111 individuals from the representative sample, with a total of 158,886 observations. We exclude observations in the army, and restrict to those between the ages of 25 and 63. This reduces the number of observations to 136,955. Restricting to households with two adults, the number of individuals is reduced to 5,079. 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 4,399 (51,812). We further restrict observations to those with positive hours of labor in the household (but lower than 10,000 annually). We keep individuals with at least 12 observations of income. This reduces the number of individuals to 2,219 (37,815 observations). After grouping observations in 4 year periods (like the model), we also drop individuals who at least once report hourly wages under half the minimum wage or above $400 and who report extreme changes of income (i.e., annual growth above 400%, or reduction by 66%). This leads to a final number of individuals of 1,575. These are distributed as follows:

- High-school dropouts: 100 Individuals, with a total number of observations of 661.
- High-school graduates: 988 Individuals, with a total number of observations of 7,072.
- College graduates: 487 Individuals, with a total number of observations of 3,394.

A.4 Additional Figures and Tables

Table A3: Age profile of wages by education group

VARIABLES	(1) HS Grad	(2) College
Age	0.0312***	0.0557***
	(0.00387)	(0.00577)
Age^2	-0.000271***	-0.000530***
	(4.65e-05)	(6.89e-05)
Constant	2.084***	1.927***
	(0.0779)	(0.118)
Observations	9,130	6,015
R-squared	0.051	0.093
# of households	1357	864

*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 $\{c_j(s_j), h_j(s_j), a'_j(s_j)\}$, parental time and money investments $\{d^{\zeta}_j(s_j), d^m_j(s_j)\}$, and parental transfers $\{\varphi(s_j)\}$; value functions $\{V_j(s_j), V^s_j(s_j), V^{sw}(s_j)\}$; (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 μ such that:

1. Given prices, decision rules solve the respective household problems and $\{V_j(s_j), V^s_j(s_j), V^{sw}(s_j)\}$ 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:

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

   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 Children's Skills Production Function

Cunha et al. (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. Cognitive and non-cognitive skills are included separately in their estimations. Cunha et al. highlight that abstracting from the two types of skills leads to estimates that suggest that investments in the early childhood development of low-skilled children are much less productive. Similarly, the production function of each type of skills is estimated separately assuming a CES function. We use Cunha et al.'s preferred estimation (on a representative sample) for the outer CES, which requires allowing parameters to vary with the age of the children (which is why parameters are allowed to vary with j). Table C1 shows the parameter values and standard deviations.

Table C1: Children's Skills Production Function: estimates from Cunha et al. (2010)

	Cognitive Skills ($q = c$)		**Non-Cognitive Skills** ($q = nc$)	
	1st Stage ($j = 1$)	**2nd Stage** ($j = 2$)	**1st Stage** ($j = 1$)	**2nd Stage** ($j = 2$)
Current Cognitive Skills $\left(\hat{\alpha}_{1,q,j}\right)$	0.479	0.831	0.000	0.000
	(0.026)	(0.011)	(0.026)	(0.010)
Current Non-Cognitive Skills $\left(\hat{\alpha}_{2,q,j}\right)$	0.070	0.001	0.585	0.816
	(0.024)	(0.005)	(0.032)	(0.013)
Parent's Cognitive Skills $\left(\hat{\alpha}_{3,q,j}\right)$	0.031	0.073	0.017	0.000
	(0.013)	(0.008)	(0.013)	(0.008)
Parent's Non-Cognitive Skills $\left(\hat{\alpha}_{4,q,j}\right)$	0.258	0.051	0.333	0.133
	(0.029)	(0.014)	(0.034)	(0.017)
Investments $\left(\hat{\alpha}_{5,q,j}\right)$	0.161	0.044	0.065	0.051
	(0.015)	(0.006)	(0.021)	(0.006)
Complementarity parameter $\left(\rho_{q,j}\right)$	0.313	-1.243	-0.610	-0.551
	(0.134)	(0.125)	(0.215)	(0.169)
Variance of Shocks $\left(\hat{\sigma}^2_{q,j}\right)$	0.176	0.087	0.222	0.101
	(0.007)	(0.003)	(0.013)	(0.004)

Standard errors in parentheses.

We are able to use their estimates because they also report summary statistics for θ, particularly the estimated covariances. For our analysis, we use AFQT and Rotter's locus of control scores as measures of cognitive and non-cognitive skills, respectively, in NLSY. The assumption necessary is that these measures are good proxies for the measures Cunha et al. estimate as cognitive and non-cognitive skills. Given that they also use AFQT and Rotter's locus of control scores in the measurement equation of their dynamic factor model estimation, we believe our assumption is reasonable. Then, for the dispersion of our measures to be in line with theirs, we standardize our measures of cognitive and non-cognitive skills to have the same standard deviation that they report in their analysis. Thus, a one-standard deviation change in their paper can be interpreted in the same way as a one-standard deviation change in our empirical analysis of the effect of skills on wages or education as well as in our model. Regarding the scale of skills in the model, the productivity of investments $\bar{A}$ is estimated such that the average level of log-cognitive-skills in the estimated economy is zero, since this normalization approach is used by Cunha et al. and we estimated the income process under this normalization. Cunha et al. re-normalize skills every period, while our estimation procedure only targets that the normalization holds at the end of the skills development process. Nevertheless, average skills (in logs) are close to zero (0.1) at the end of the first period of development and results are robust to making $\bar{A}$age dependent so that skills are normalized every period.[83] However, as a robustness check, in Section 5.1.3 we also evaluate how our results would change if parameters in the skills production function are moved in either direction.

We also need to transform their parameters from 2-year periods to 4-year periods. Let $\theta_{c,k,t}$ and $\theta_{nc,k,t}$ be the cognitive and non-cognitive skills of the child in period t. To go from 2-year periods to 4-year periods, we simply iterate in the production function and assume that the shock ν only takes place in the last iteration.[84] We further assume that the cross-effect of skills (i.e., of cognitive on non-cognitive and of non-cognitive on cognitive) is only updated every two periods. Removing this assumption does not change results significantly since the weights corresponding to these elements are very small or even zero in the estimation (in Table C1, see rows 2 under columns 1 and 2, as well as row 1 under columns 3 and 4), but it destroys the CES functional form if $\rho_{c,j} \neq \rho_{nc,j}$. Under these assumptions, the production function of cognitive skills over a 4-year period is given by:

$$\theta_{c,k,t+2} = \left[\hat{\alpha}_{1,c,j}^2 \theta_{c,k,t}^{\rho_{c,j}} + \left(1 + \hat{\alpha}_{1,c,j}\right) \hat{\alpha}_{2,c,j} \theta_{nc,k.t}^{\rho_{c,j}} + \left(1 + \hat{\alpha}_{1,c,j}\right) \hat{\alpha}_{3,c,j} \theta_{c}^{\rho_{c,j}} + \dots \right.$$
$$\left. + \left(1 + \hat{\alpha}_{1,c,j}\right) \hat{\alpha}_{4,c,j} \theta_{nc}^{\rho_{c,j}} + \left(1 + \hat{\alpha}_{1,c,j}\right) \hat{\alpha}_{5,c,j} I^{\rho_{c,j}} \right]^{1/\rho_{c,j}} e^{\nu_{q,t+2}}$$

Thus, the persistence parameter needs to be squared (i.e., $\alpha_{1,c,j} = \hat{\alpha}_{1,c,j}^2$), while other parameters inside the CES function need to be multiplied by 1 plus the persistence parameter (e.g., $\alpha_{2,c,j} = \left(1 + \hat{\alpha}_{1,c,j}\right) \hat{\alpha}_{2,c,j}$). For the shock, in the baseline exercise we assume that the variance of the shock in the 4-year model is twice the one in the 2-year model (i.e., $\sigma_{q,j}^2 = \hat{\sigma}_{q,j}^2$). In Section 5.1.3 we evaluate how our results would change if the variance of the shock, among others, was different.

[83] Results are available upon request.

[84] Given that investments I are chosen only once within periods in the model this assumption seems reasonable.

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 skills level to estimate the average lifetime income, on which the replacement benefit is based. We estimate the average life time income to be $\widehat{y}_j(\theta, e) = w_e E_{e,j}(\theta, \overline{\eta}) \times \overline{h}$ 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 (9) to obtain the replacement benefits.

The pension formula is given by

$$\pi(\theta, e) = \begin{cases} 0.9\widehat{y}(\theta, e) & \text{if } \widehat{y}(\theta, e) \leq 0.3\bar{y} \\ 0.9(0.3\bar{y}) + 0.32(\widehat{y}(\theta, e) - 0.3\bar{y}) & \text{if } 0.3\bar{y} \leq \widehat{y}(\theta, e) \leq 2\bar{y} \\ 0.9(0.3\bar{y}) + 0.32(2 - 0.3)\bar{y} + 0.15(\widehat{y}(\theta, e) - 2\bar{y}) & \text{if } 2\bar{y} \leq \widehat{y}(\theta, e) \leq 4.1\bar{y} \\ 0.9(0.3\bar{y}) + 0.32(2 - 0.3)\bar{y} + 0.15(4.1 - 2)\bar{y} & \text{if } 4.1\bar{y} \leq \widehat{y}(\theta, e) \end{cases} \tag{9}$$

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

C.3 Aggregate Production Function

Wages from PSID

In order to estimate the aggregate production function (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:

$$Y_t = AK_t^{\alpha} H_t^{1-\alpha}$$
$$H_t = \left[sH_{1t}^{\rho} + (1-s)H_{2t}^{\rho}\right]^{1/\rho}$$

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

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

Estimating this equation in first differences and using lags of $\Delta log\left(\frac{H_{2t}}{H_{1t}^{1}}\right)$ as instruments, in particular, leads to estimates of 0.43 for ρ and 0.53 for s—in line with the estimates from Katz and Murphy (1992) and Heckman et al. (1998).[85]

C.4 Simulated Method of Moments: Moments' Selection

In this section we show a method that is useful to show which parameters are closely related to each moment, which helps justify our selection of moments. 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 a hypercube of the parameter space, we draw 150,000 candidate parameter vectors from uniform Sobol (quasi-random) points, solve and simulate the model, and compute the implied moments in steady state. Second, for each parameter, we associate a relevant target moment. Third, for each parameter, we divide the vector of this particular parameter into 50 quantiles and compute the 25th, 50th, and 75th percentiles of the associated moment in each quantile.[86] Finally, we show these percentiles of the moment along with the value of that moment 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 that parameter is 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 moment's 25th and 75th percentiles are further apart).

[85]We can also estimate this using first differences without instruments and results are similar (i.e., $\rho = 0.497$ and $s = 0.521$).

[86]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.

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, respectively, as shown by the first row of Figure C1. More precisely, there is a positive relation between the level of altruism (δ) and transfers to children. As parents value their children more (higher δ), 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 correlation's 25th and 75th percentiles. This is due to the fact that, for example, when ξ approaches zero all parents invest the same amount of time (i.e., the maximum), or alternatively, when the share of money α_m approaches 1, time with children is reduced to almost zero 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 5.1.3, results were almost unchanged by moving γ within two standard deviations.

Figure C1: Identification

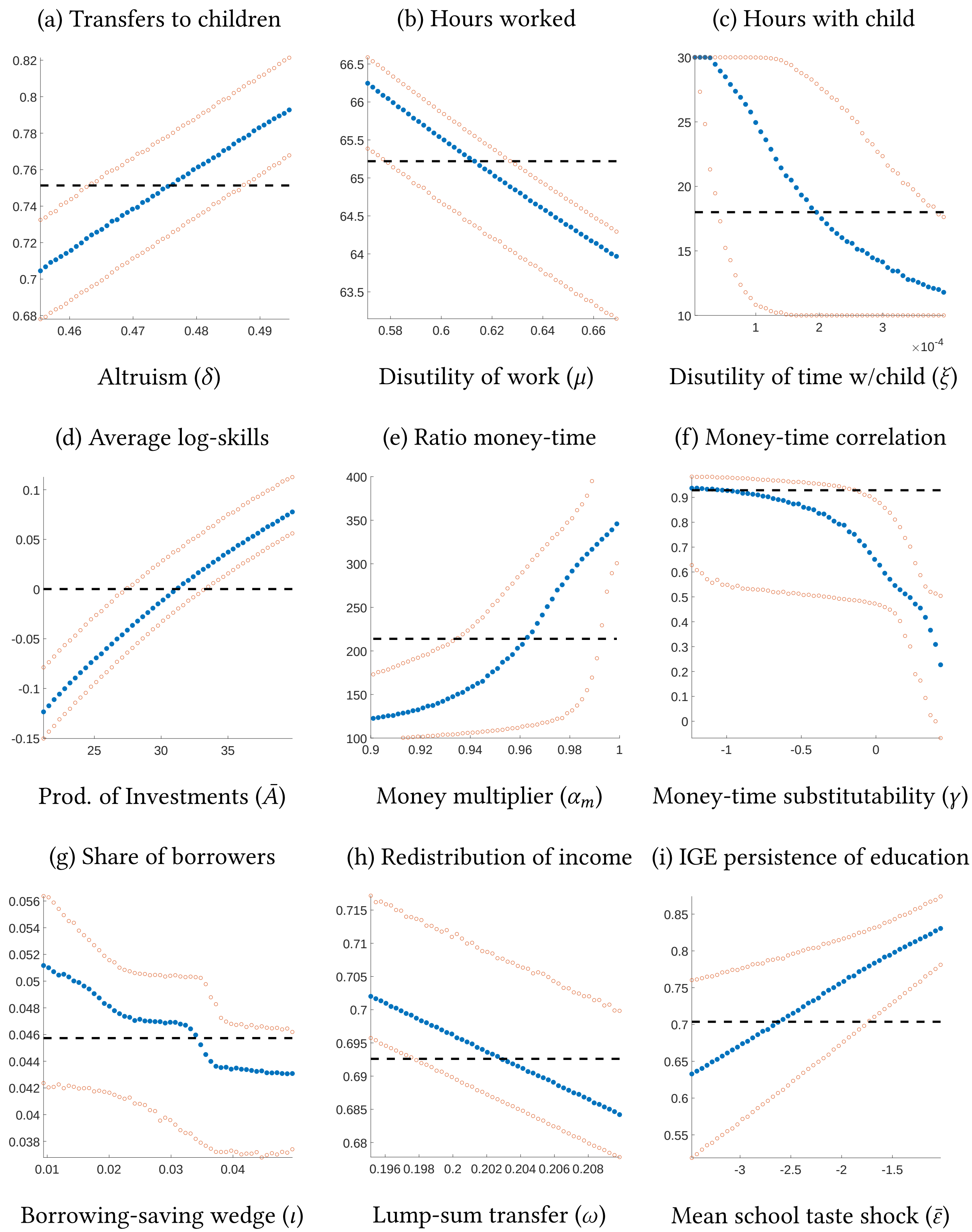
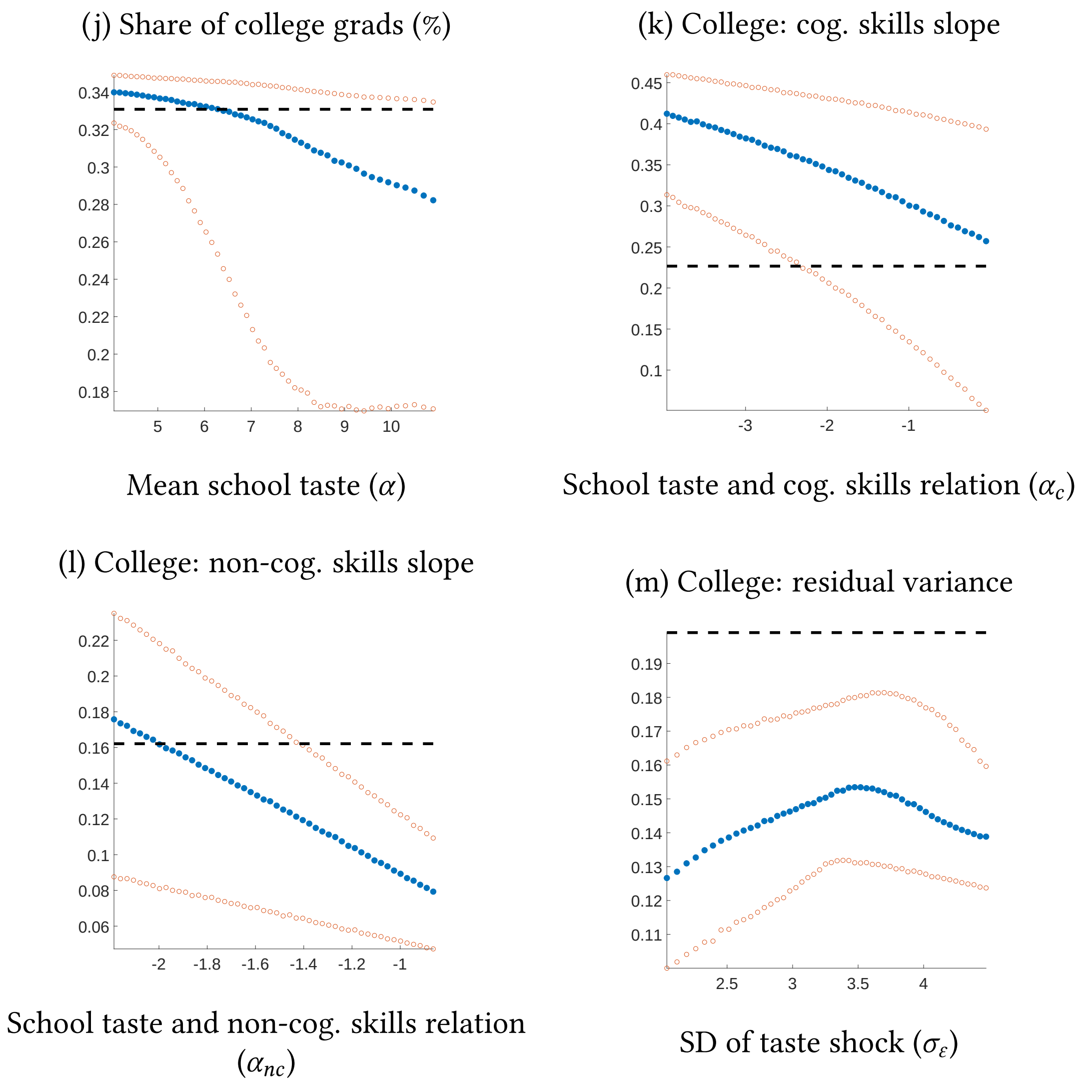


Figure C1 (cont.): Identification



Note: For each parameter's quantile, the blue dots shows the median, while the 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.

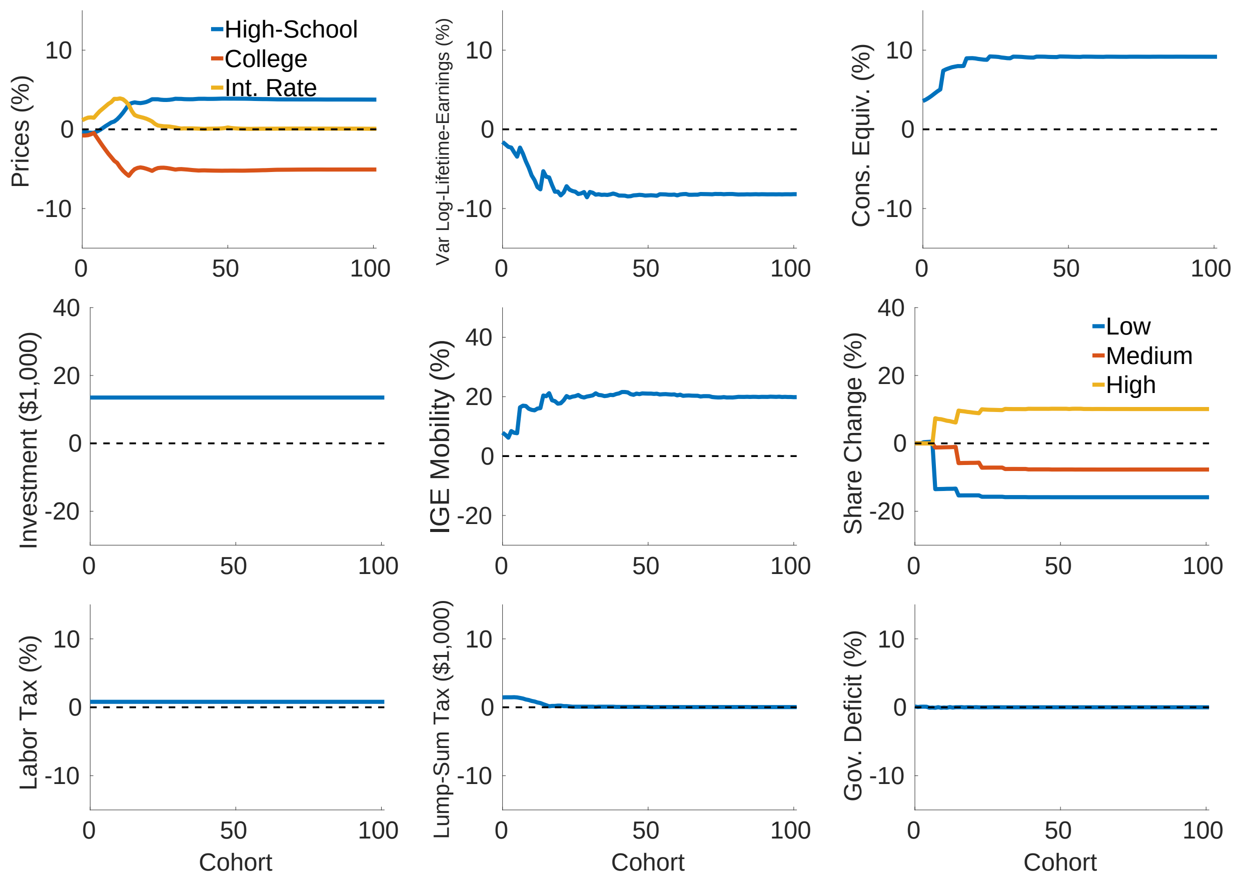
D Transition Details

Section 5 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 a transition and 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 4 showed the main results regarding welfare, inequality and intergenerational 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 D1 expands that analysis by also including information on price and tax changes and on welfare changes for children born to different socioeconomic groups.

Figure D1: Transition dynamics: more details of balanced budget case

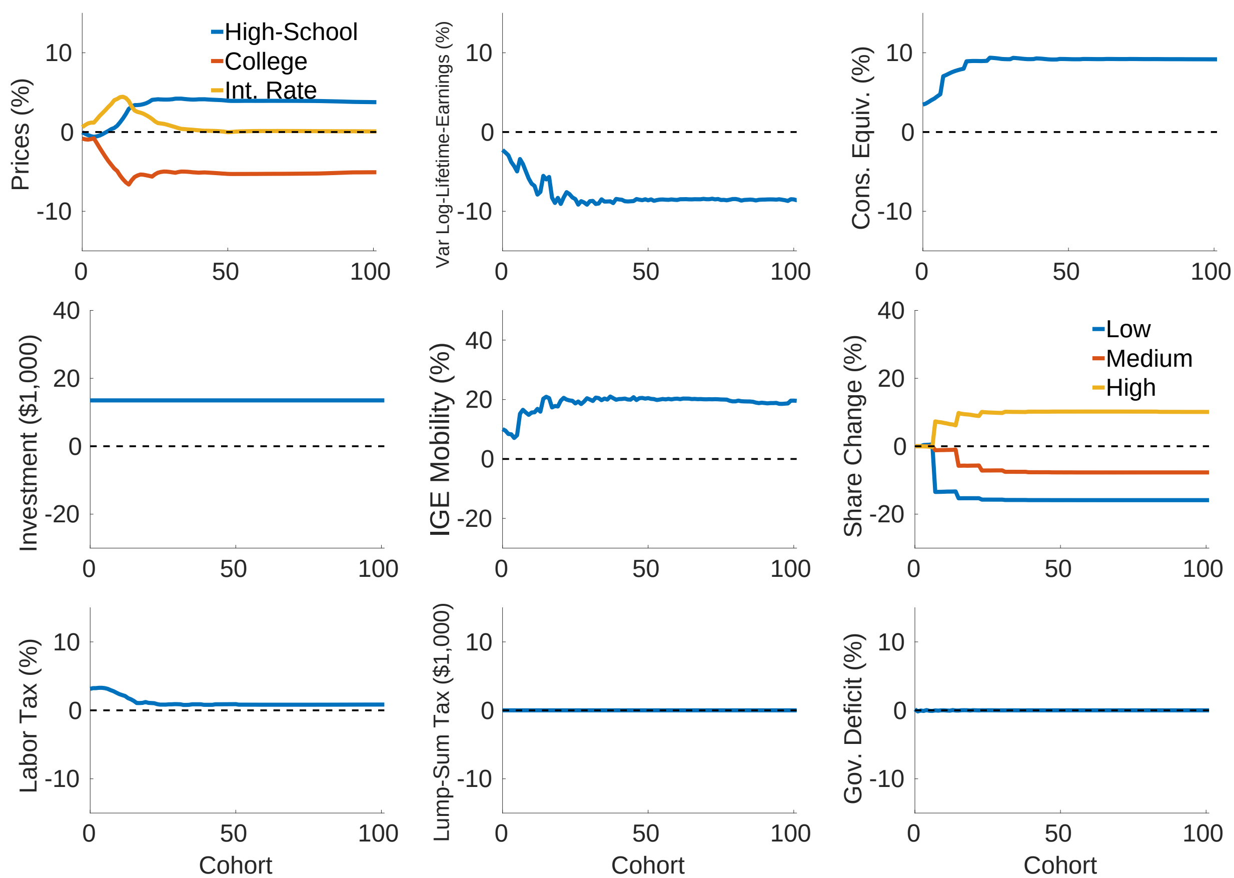


Notes: The policy (including the investments and labor income tax change) is introduced unexpectedly. We compute the transition introducing a lump-sum tax such that the government's budget balances every period. Consumption equivalence is shown for a newborn from the cohort defined by the horizontal 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 each 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.1.1 Balanced every period with labor income tax

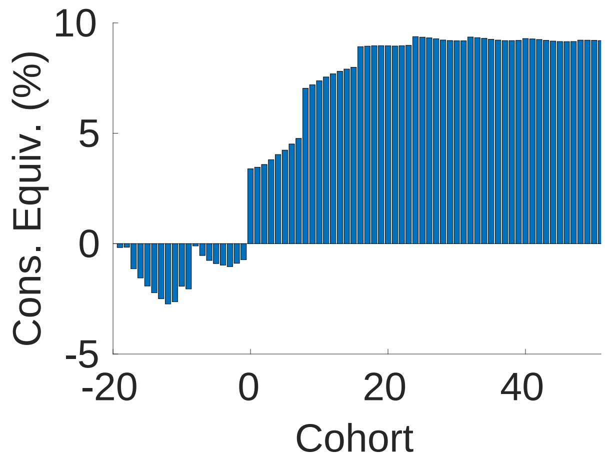
Figure 5, in the main text, shows the transition dynamics when the government keeps its budget balanced every period in the transition using a lump-sum tax. Here we show the same outcomes when the government uses the distortive labor income tax τ_y instead of the lump-sum tax. Figures D2 and D3 show that the main results are almost unchanged. Welfare gains during the transition are smaller when the transition is financed with progressive taxes, but differences are very small (less than 0.3 percentage points), most likely because the policy is almost self-financing in the long run and hence does not require large increases in taxes. The labor income tax rate increases by 3.3 percentage points early in the transition, but it is then reduced at a similar pace as the lump-sum tax in the other case.

Figure D2: Transition dynamics: balanced every period with labor income tax



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, starting at $1,000 per child-year for the first cohort and growing 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 each 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 D3: 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 the 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 accordingly for agents of age 40 at the time.

D.2 Other Alternatives

Figure 5, 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 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.[87]

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

[87] 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.

gains for all new cohorts and most individuals alive at the time of the introduction.

D.2.1 Only Intervened Cohorts Pay

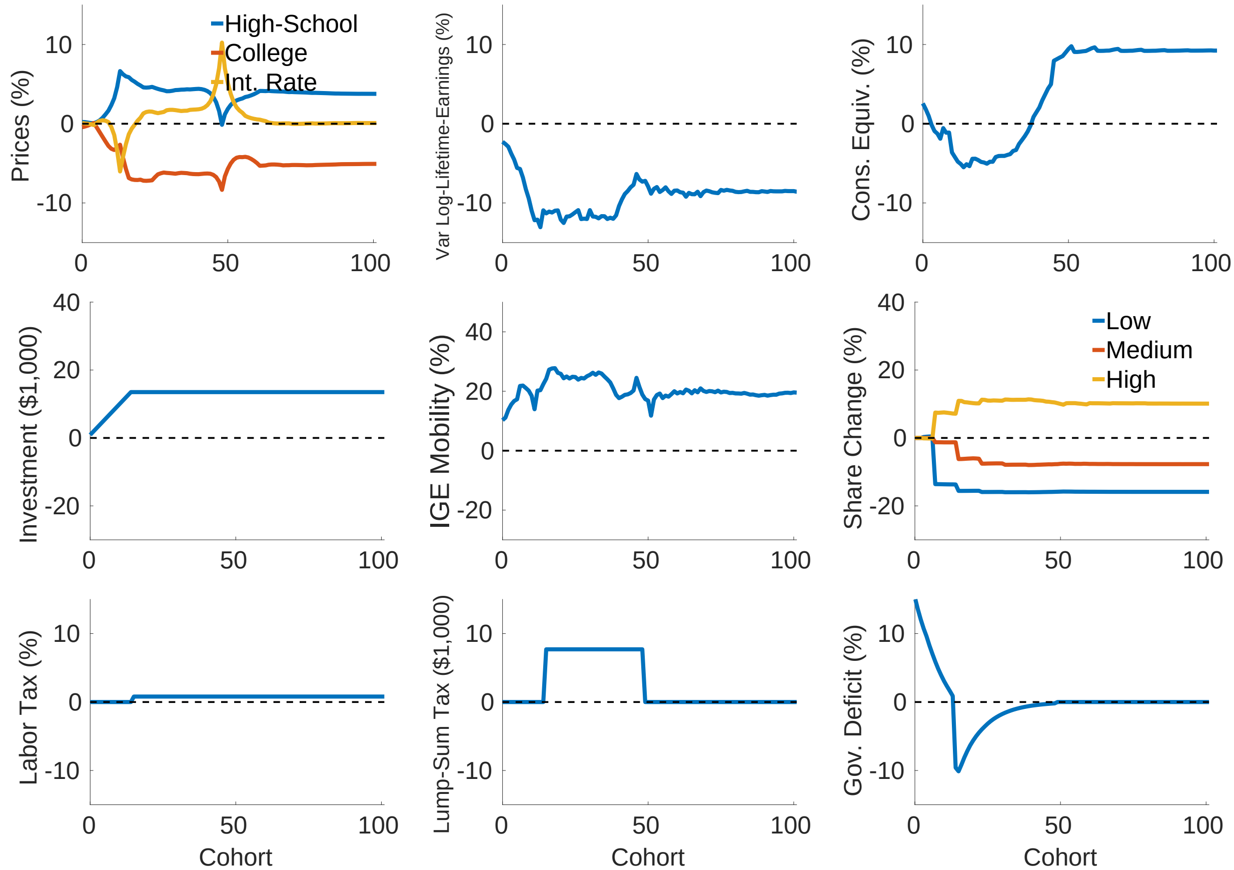
Figure D4 shows the transition dynamics to the baseline policy in which the government invests $13,500 per child-year, when only intervened cohorts have to pay higher taxes. To compensate for the smaller early increase in taxes, the government is allowed to borrow at an interest of 3%. But it has to use the later higher taxes to pay off its debt by the time cohort 50 is born.[88] 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 D5 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 debt. Only after 40 cohorts are born do we observe welfare gains once again.[89]

[88]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.

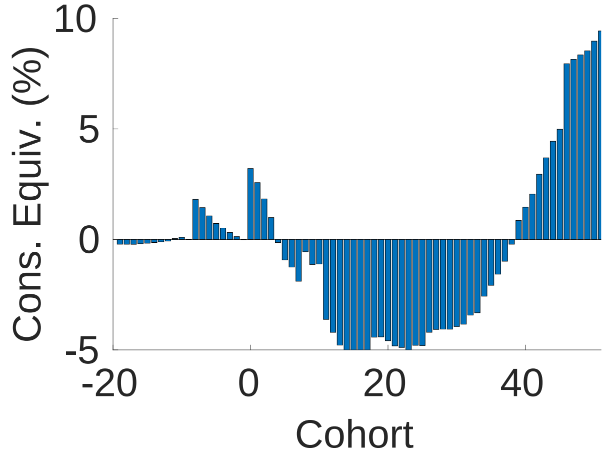
[89]These results clearly depend on the assumptions of times of repayments and interest rates. However, they do show that government borrowing may not be enough.

Figure D4: 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 each 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 D5: Welfare gains including older cohorts



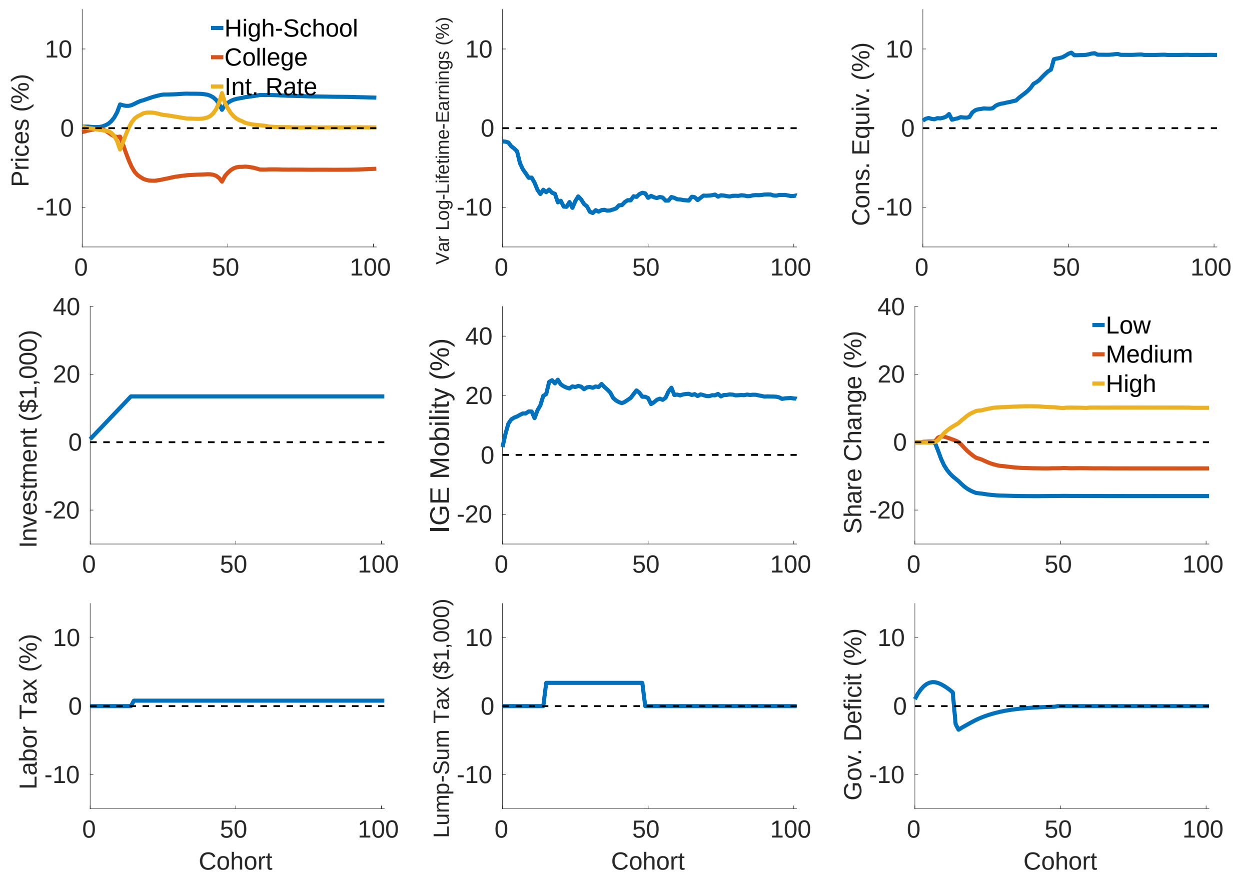
Notes: Welfare gains are reported for cohorts born after the policy is introduced (i.e., cohorts from 0 on) as well as for cohorts already alive at the time (i.e., cohorts less than 0). For the first group, welfare gains are computed for newborns. For the cohorts already alive at the time the policy 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 accordingly for 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%.

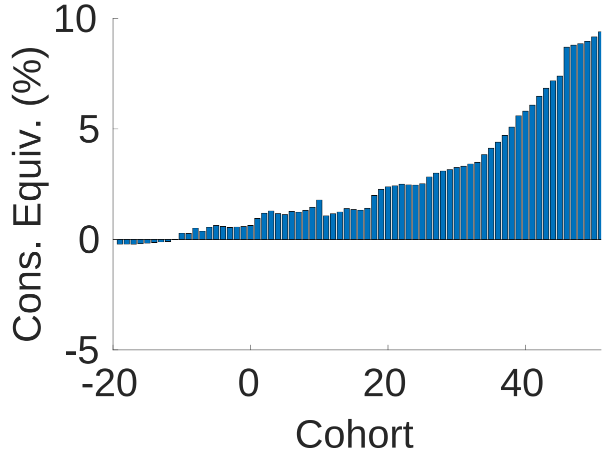
Figures D6 and D7 show that this combination of slow introduction and higher taxes only for intervened cohorts is able to achieve welfare gains for most cohorts. All new cohorts accrue welfare gains, as do most cohorts alive at the time of the introduction. Only cohorts whose children have already grown 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 D6: 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, starting at $1,000 per child-year for the first cohort and growing by $1,000 for every new cohort until they reach the target of $13,500. We compute the transition by 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 D7: 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 the 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 accordingly for 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 5.1.4, is re-estimated to match the same set of moments from the full model. The estimated parameters and moments are shown in Table E1.

Table E1: Extension estimation: with early childhood education production function

Parameter	Value	Description	Moment	Data	Model
Preferences					
μ	187	Mean labor disutility	Avg. hours worked	65.2	65.2
δ	0.47	Altruism	Parent-to-child transfer as share of income	0.75	0.73
School Taste:					
α	9.74	Avg. taste for college	College share	33	32
α_{θ_c}	-0.49	College taste and cog. skills relation	College: cog skills slope	0.23	0.24
$\alpha_{\theta_{nc}}$	-1.80	College taste and noncog. skills relation	College: noncog skills slope	0.16	0.16
σ_ε	3.64	SD of college taste shock	College: residual variance	0.20	0.18
$\bar{\varepsilon}$	-3.87	Draw of school taste: mean by parent's education	Intergenerational persistence of education	0.70	0.60
Skill Formation Productivity:					
ξ	0.06	Parental time disutility of time with children	Avg. hours with children	18.0	15.5
$\bar{A}$	86.8	Returns to investments	Average skill ratio	0.0	0.0
α_m	0.95	Money productivity	Ratio of money to hours	218	208
γ	-0.28	Money-time substitutability	Money-time correlation	0.93	0.93
Interest rate					
ι ($\times 10^2$)	3.35	Borrow-save wedge	Share of borrowers	4.5	5.0
Government					
ω ($\times 10$)	2.01	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. For the money-time correlation, we group parents by deciles of time investments and compute average time and money investments for each decile. We then compute the correlation between the two averages. Share of borrowers is estimated from the Survey of Consumer Finances. College share, college-skills slope and college residual variance are estimated using NLSY. All moments matter for all parameters, but each line highlights the moment that is particularly informative for the corresponding parameter. See Appendix C.4 for more details.

F Extension: With Three-Periods of Skills Development

The baseline model in the paper only included two periods ($j = 8$ and 9) during which skills θ were allowed to develop. This limitation might create the concern that the model would overstate the effect of the early childhood investments policy considered. To overcome this concern, we extend the model to have one additional period of child development during $j = 10$ (ages 8–11). The model is re-estimated to match the same set of moments from the baseline model, but we take into account that the government is already investing a sizable amount of resources in children's education during this stage. Estimates for the amount of federal and state government expenses in education during this stage vary, but we choose a low end in order to fall on the conservative side of our results.[90]The estimated parameters and

[90]According to the "Census of Governments: Finance - Survey of School System Finances," the average expense is approximately $5,000 (2000 USD) per student. We choose $3,500 (close to the 20th percentile) so that it is more likely that the effect of the early childhood policy will fade out as children grow older, hence falling on the conservative side of our main results. We abstract from local funding of public schools since this would imply more redistribution than is actually taking place, as it is parents in the same small school district who are paying for these local funds through local taxes (Fernandez

moments are shown in Table F1.

Table F1: Extension estimation: with three-periods of skills development

Parameter	Value	Description	Moment	Data	Model
Preferences					
μ	140	Mean labor disutility	Avg. hours worked	65.2	68.9
δ	0.53	Altruism	Parent-to-child transfer as share of income	0.75	0.80
School Taste:					
α	6.33	Avg. taste for college	College share	33	29
α_{θ_c}	-1.09	College taste and cog. skills relation	College: cog skills slope	0.23	0.22
$\alpha_{\theta_{nc}}$	-1.41	College taste and noncog. skills relation	College: noncog skills slope	0.16	0.17
σ_ε	2.63	SD of college taste shock	College: residual variance	0.20	0.18
$\bar{\varepsilon}$	-1.46	Draw of school taste: mean by parent's education	Intergenerational persistence of education	0.70	0.76
Skill Formation Productivity:					
ξ	0.05	Parental time disutility of time with children	Avg. hours with children	18.0	15.1
$\bar{A}$	38.6	Returns to investments	Average skill ratio	0.0	0.0
α_m	0.96	Money productivity	Ratio of money to hours	218	181
γ	-0.34	Money-time substitutability	Money-time correlation	0.93	0.98
Interest rate					
ι ($\times 10^2$)	4.06	Borrow-save wedge	Share of borrowers	4.5	5.0
Government					
ω ($\times 10$)	2.28	Lump-sum transfer	Income variance ratio: Disposable to pre-gov	0.69	0.69

Notes: Parent-to-child transfers, hours worked, skill formation moments and intergenerational persistence of education are estimated from PSID-CDS data. For the money-time correlation, we group parents by deciles of time investments and compute average time and money investments for each decile. We then compute the correlation between the two averages. Share of borrowers is estimated from the Survey of Consumer Finances. College share, college-skills slope and college residual variance are estimated using NLSY. All moments matter for all parameters, but each line highlights the moment that is particularly informative for the corresponding parameter. See Appendix C.4 for more details.

Table F1 presents the main results in this extended model, comparing them to the baseline ones from Section 5.1.1. Even though long-run welfare gains and other policy effects are slightly smaller, they are still sizable, with welfare gains of 7.2% (vs. 9.4%), for example. More important for our main result, however, this model leads to an even larger gap between long-run and short-run welfare gains. While the ratio of long-run to short-run welfare gains in the baseline model was equal to 2.4, in the extended model with three periods of development it is equal to almost 2.8. This is because in a model with three periods of development, parents are even more important since the gains of the early childhood program have more periods during which effects can fade out. In the long run, however, the policy combines the additional investments together with better parents, making the fading out of the results less prominent than in the short run.

and Rogerson, 2003).

Table F2: Comparison of results with 3-periods extension

	Short-Run PE	Long-Run GE		
	Cons. Equiv.	Cons. Equiv.	Inequality	Mobility
Three-Periods Extension	2.6	7.2	-7.1	12.8
Baseline Model	3.9	9.4	-7.7	20.2

Notes: We simulate introducing the same level of investments as in Garcia et al (2017) in the baseline model with two periods of skill development and in the extended model with three periods of skill development. Early childhood investments of $13,500 per child-year when children are between 0 and 3 years old are introduced. Long-run GE refers to looking at outcomes in the new long-run steady state, adjusting wages and interest rates to clear the market and adjusting the labor income tax to keep the government's budget unchanged. In the short-run PE case, we calculate the effect of a one-generation policy and evaluate the effect on that generation without adjusting taxes or prices. Consumption equivalence is determined by newborns under the veil of ignorance. Inequality refers to the variance of log-lifetime-earnings while intergenerational mobility refers to minus the regression coefficient between children's and parents' income ranks.

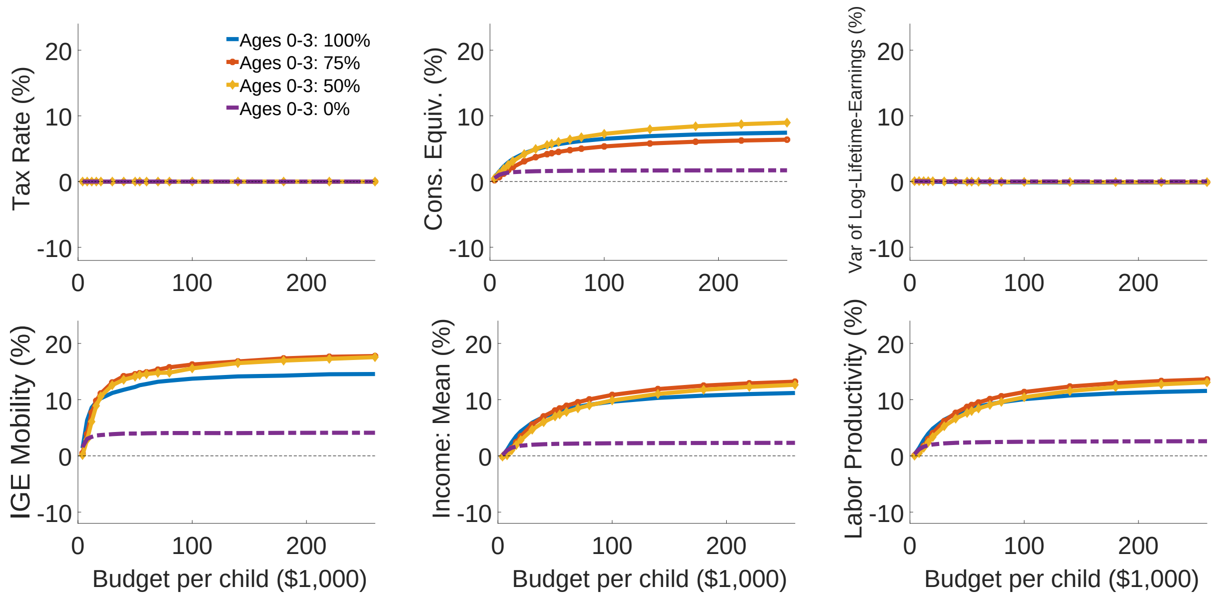
G Investments in Older Children and Targeting

We evaluate additional alternatives to the policy evaluated in Section 5. Instead of using all resources to invest in all children ages 0–3, we explore using some of the resources to invest in older children as well. We also explore targeting particular groups of children who may be considered disadvantaged.

G.1 Early Childhood Investments: More Alternatives

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 (ages 4–7). Figure G1 shows the results of these policies in the short-run PE case. As expected, welfare gains increase with the amount of resources since none of the potential costs are being considered in this case. Regarding the optimal allocation of such resources across different ages, these results suggest that if the budget available is less than $40,000 per child, it is better to allocate all the resources in the the first period (i.e., 0-3 years old). When the available budget is larger, there are small additional gains from allocating half of the resources towards investing in older children (since investing everything in one period is not as productive).

Figure G1: Childhood investments: Short-Run Partial-Equilibrium



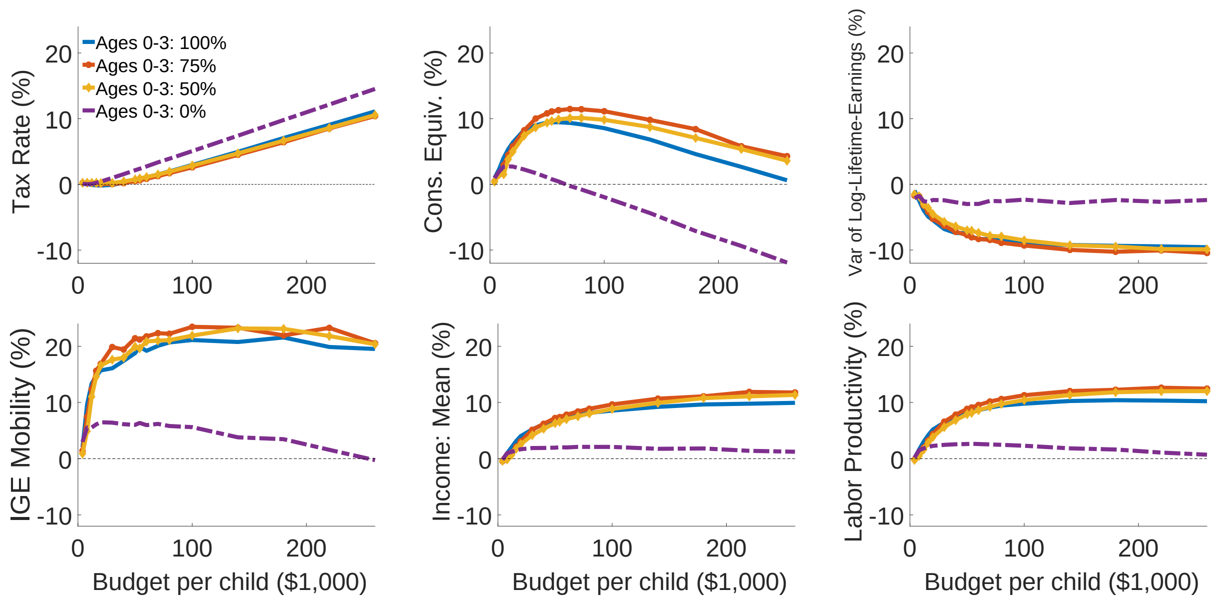
Notes: We simulate policies in which the government uses different amount of resources either to invest directly in the development of children or to fund a government transfer to all individuals (at the age of 16). Among childhood investment policies, we also evaluate different alternatives: (i) use all resources in children ages 0–3 (as in the main text); (ii) use 80% of the resources for children ages 0–3 and 20% for children ages 4–7; and (iii) use 50% of the resources for children ages 0–3 and 50% for children ages 4–7. We evaluate these policies for different amounts of resources available in a short-run partial-equilibrium framework. 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 ages 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.

Does this intuition also hold true when we consider long-run GE forces as well as the fact that taxes may need to increase to finance these policies? Figure G2 shows the results of these policies in the new steady-state, taking into account GE effects as well as adjusting the labor income tax such that the government's budget remains balanced. Once again, we find that if the budget available is less than $40,000 per child it is better to allocate all the resources in the the first period (i.e., ages 0-3 years). Similar to the short-run PE case, when the available budget is larger, there are small additional gains from allocating part of the resources towards investing in older children. In this case, however, using 25% (instead of 50%) of the resources to invest in children ages 4–7, and 75% for children ages 0–3, leads to the largest welfare gains.[91] However, the differences in welfare gains are small (less than 2 percentage points) relative to the case of investing all resources in children ages 0–3, and so are the differences in other outcomes (inequality, mobility, and average income). Gains are smaller if a larger share of resources is used in children ages 4–7 since earlier investments generally lead to larger gains because the child's skills production function as estimated by Cunha et al. (2010) implies that skills are

[91]Differences between the two cases may stem from the fact that in the long-run case, parents' characteristics are able to adjust. In particular, parents tend to be more skilled, more educated, and have higher savings. Therefore, in the long run parents are more likely to be able to use their own resources to invest in the second period of development after the government invests in the first period.

more malleable at younger ages.

Figure G2: Childhood investments: Long-Run General-Equilibrium



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 ages 0–3 (as in the main text); (ii) use 80% of the resources for children ages 0–3 and 20% for children ages 4–7; and (iii) use 50% of the resources for children ages 0–3 and 50% for children ages 4–7. We evaluate these policies for different amounts of resources available in a long-run general-equilibrium framework. 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 ages 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.

G.2 Early Childhood Investments: Targeting

Most of the early childhood development empirical literature is based on evidence regarding disadvantaged children. Gains are expected to be larger among children of lower-income and/or lower-educated parents. This is true in our model as well as highlighted in Section 5, but we now explore the possibility of targeting particular groups of children who may be considered disadvantaged. Although we would like to explore targeting the policy according to parents' income, this is particularly difficult in our model with endogenous labor supply since such a policy would distort the labor choice, making the labor supply decision rule not easy to solve—which given the size of the model is a major computational problem. Therefore, we limit our analysis to targeting by parents' education and skills.

Table G1 shows the welfare gains under the veil of ignorance for all children who are able to obtain the government investment, when children know they are going to be born to a parent with a given education and skill group in both economies.[92] In contrast, Table G2 shows the welfare gains under the veil of ignorance for all children, irrespective of whether they receive the policy or not. These gains are particularly important since the largest part of the gains highlighted in this paper stem from the fact that the policy creates better parents. If we were to focus on the gains of, for example, only children born to low-skilled high-school graduate parents, we would miss this source of long-run gains. Both tables show the gains in the short-run PE case and the long-run GE case (i.e., after the distribution of parents' characteristics, in particular, have adjusted).

We find no evidence that targeting a large level of government early childhood investments may lead to higher gains in our model.[93] On the one hand, targeting the investments towards children with lower resources implies that crowding out of parental investments should be smaller, hence probably leading to larger gains. On the other hand, families with larger resources may also be able to continue the government's early-age investments g with further investments in future stages of development (e.g., ages 4–7 or helping finance college). Due to the dynamic complementarities of the production function of skills (Cunha et al., 2010) and the complementarity between skills and education (based on Table 1), this means that gains may be larger for children from families with higher resources. While a targeted policy may also be less costly and, thus, be associated with lower tax increases, it can also reduce incentives to invest in skills and education if only lower-skilled, lower-educated (or lower-income) individuals are able to obtain the government investments.

Table G1: Early Childhood Development Targeting: Welfare Gains in the Targeted Groups

	Parent's Skills		
Parent's Education	**Low**	**Low + Medium**	**All**
High School			
SR-PE	3.2	3.6	3.7
LR-GE	3.8	6.6	8.1
High School + College			
SR-PE	3.2	3.8	3.9
LR-GE	3.8	7.1	9.4

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 evaluate targeting the policy according to parents' characteristics. Low and medium skills refer to the bottom one-third and two-thirds of the distribution of skills, respectively. Welfare gains are evaluated according to the consumption equivalence of newborns under the veil of ignorance for children who qualify for the government intervention (see main text for details). SR-PE refers the short-run one-generation policy in partial-equilibrium (without tax adjustments) evaluating the effect on that generation. LR-GE looks at gains in the new long-run steady state, adjusting the labor income tax to keep the government's budget unchanged and adjusting wages and interest rates to clear the markets.

[92]There is still heterogeneity within each of these groups since parents' assets and idiosyncratic labor productivity can vary. Children know they are going to be born a parent with a given education and skill level in both economies, but they also know that the parent's distribution over assets and idiosyncratic labor productivity (conditional on education and skills) is different between those two economies.

[93]We have also explored targeting for lower amounts of g. We find that targeting may be optimal for lower amounts of government investments since high-income parents are more likely to fully crowd out these investments in such a case. In this case, however, we also find that our main result that long-run gains are larger than short-run ones remains true. These results are available upon request.

Table G2: Early Childhood Development Targeting: Welfare Gains for all Newborn

	Parent's Skills		
Parent's Education	**Low**	**Low + Medium**	**All**
High School			
SR-PE	0.6	2.2	2.5
LR-GE	1.0	4.7	6.2
High School + College			
SR-PE	0.6	3.0	3.9
LR-GE	1.0	6.0	9.4

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 evaluate targeting the policy according to parents' characteristics. Low and medium skills refer to the bottom one-third and two-thirds of the distribution of skills, respectively. Welfare gains are evaluated according to the consumption equivalence of newborns under the veil of ignorance for all children (see main text for details). SR-PE refers to the short-run one-generation policy in partial-equilibrium (without tax adjustments) evaluating the effect on that generation. LR-GE looks at gains in the new long-run steady state, adjusting the labor income tax to keep the government's budget unchanged and adjusting wages and interest rates to clear the markets.