From Childhood to Adult Inequality:
Parental Investments and Early Childhood Development

Diego Daruich*

New York University

[PRELIMINARY]

June 15, 2017

Abstract

Macroeconomic analysis of inequality and intergenerational mobility typically abstracts from the role of endogenous childhood development. This paper shows that this omission is not innocuous. It extends the standard general-equilibrium heterogeneous-agent life-cycle model with earnings risk and credit constraints to incorporate endogenous early childhood investments (parental time and money), family transfers, and college education. Agents cannot invest in their own childhood development or pay their parents to do so, which can lead to inefficiently low levels of investments. Estimating the model to match key moments of the data (e.g., average childhood investments and returns to skills), we find that a government program that funds early childhood investments would reduce inequality by 20% and improve mobility by 60%, as well as yield a three times larger welfare increase than a government transfer that uses the equivalent revenue. Introducing these investments in the model as a short-run small-scale policy—similar to an RCT—would underestimate the benefits by two-thirds. Long-run distribution changes are the main drivers of the underestimation in welfare and mobility outcomes, while general-equilibrium is important for the reduction of inequality. The paper also shows that the government is able to increase the welfare of every new generation by taking a temporary deficit to afford these investments during the transition.

JEL Classifications: J13, J24, J62.

Keywords: Inequality, Intergenerational mobility, Quantitative model.

*ddaruich@nyu.edu. I am grateful to my advisors, Raquel Fernandez and Gianluca Violante. For helpful comments, I thank Christopher Flinn, Sonia Gilbukh, Julian Kozlowski, and Pau Roldan. I also received valuable feedback from participants at seminars at New York University. I alone am responsible for any errors and interpretations.
1 Introduction

Macroeconomic analysis of inequality generally abstracts from the role of endogenous early childhood development, focusing instead on the wealth of the top 1% (e.g., Piketty and Saez, 2003; Diamond and Saez, 2011; Saez, 2016) or the role of skills and education (e.g., Katz and Murphy, 1992; Autor et al., 2008; Abbott et al., 2013). However, early childhood environment—defined by parents’ characteristics and their interactions with their children—has been shown to significantly impact adult outcomes (Gertler et al., 2013). This paper shows that these additional intergenerational linkages are quantitatively important and need to be considered. In particular, it shows that under-investment in children’s development is relevant for the macroeconomic analysis of inequality and social mobility, and can be improved by government policies that target childhood directly.

To facilitate the comparison with the literature, we use a standard macroeconomic Aiyagari-style lifecycle model but with new key intergenerational linkages. In particular, in addition to family transfers, we introduce parental time and monetary investments in the cognitive skills of children. Since agents cannot invest towards their own childhood development or pay their parents to do so, inefficiently low levels of investment may arise. Agents would want to sign a contract with their parents in which they choose the optimal level of investments and agree to pay it back once they are adults, but such a market does not exist. Government investments towards children reduce the problems generated by this missing market. We estimate the model to match evidence on the key novel elements (i.e., parental investments and returns to skills) and find that introducing a non-targeted government investments towards early childhood (e.g., mandatory pre-K schools) leads to reductions in inequality of 20%, intergenerational mobility improvements of 60%, and welfare gains of up to 24% (using a consumption equivalence measure)—a three times as large welfare gain than using the same resources for a government transfer. Taking into account the long-run distribution changes and general equilibrium effects is important for this. Introducing these investments in the model as a short-run small-scale policy—as in an RCT—would underestimate the benefits in welfare, mobility and inequality by approximately two-thirds. Long-run distribution changes are the main drivers of the differences in welfare and mobility outcomes, while general-equilibrium is important for the reduction of inequality. We also show that by taking a temporary deficit during the transition to the new steady state with the optimal level of investments, the

\footnote{García et al. (2017) estimate the benefits of an early childhood program to be 7 dollars for every dollar spent.}
government is able to increase the welfare of every new generation.

The model specified is a life-cycle economy in a dynastic framework with four main (multi-period) stages: childhood, college, labor, and retirement. When individuals are children, their parents make investments in their development of skills—which affect education and wages. Once individuals are of college age they begin making their own choices. Their initial states include their cognitive skills and assets, which are both endogenously determined by their parents. The individuals’ first choice is whether to go to college or start working. College is costly, but can be financed with parental resources or through working and borrowing. The third stage represents their labor market experience, in which agents choose their labor supply, savings, and consumption expenditures. Wages are uncertain and there is a borrowing limit. Children are (exogenously) born during this stage, and now agents also choose their (money and time) investments towards their own children’s skills. The model requires introducing a process for the returns to these investments. We adopt a childhood development function based on Cunha et al. (2010) but, differently from them, we model parental investments explicitly. Before their children are of college age, individuals decide the amount of monetary resources transferred to them. Agents would want to invest in their own childhood, or sign a contract with their parents in which they choose the optimal level of investments and agree to pay it back once they are adults, but such a market does not exist. This missing market will create an opportunity for the government to intervene and increase welfare. The last stage is retirement, in which agents have two sources of income: savings and retirement benefits. We study this economy in general equilibrium: An aggregate firm combines the different types of labor (by education) and capital to produce the final consumption good. The government levies taxes on consumption, labor, and capital in order to finance exogenous expenses as well as provide a lump-sum transfer and retirement benefits.

The model is successfully estimated to match evidence from the US in the 2000s. We estimate the returns to skills by education group using the National Longitudinal Survey of Youth (NLSY). Most parameters and functional forms are obtained from the literature, but we use Simulated Method of Moments for the remaining ones. Two target moments are informative about the two key novel elements of the model, i.e., parental money and time investments. First, the average expenditures on children’s development are based on child care, education, and health care expenditures per child estimated by the US Department of Agriculture. Second, the average time parents spend with children, which we estimate from the Child Development Supplement of the Panel Study of Income Dynamics (PSID). The model is successful at
matching these novel moments as well as standard ones.

The model also matches other relevant moments not used in the estimation. Even though the estimation only includes averages as parental investments’ moments, the model also fits evidence from the cross section, in particular the relationship between parental investments and parental characteristics (education and income). To further test the validity of the most novel part of the model, we simulate the introduction of a relevant small-scale government program that invests directly into early childhood development and find that the effects on education outcomes are in line with the evidence based on a randomized control trial (RCT).

As reflected in the model, children cannot invest in their own early childhood or compensate their parents to do so. Even if they were able to compensate them, borrowing constraints can lead to inefficient levels of childhood investment. 2 Our model allows us to quantitatively evaluate the importance of this inefficiency and what effect a simple government policy can have on welfare. We evaluate a childhood investments program and compare the welfare benefits obtained to the ones that could be achieved by using the same resources for a government transfer provided to agents when they become independent. Welfare gains are maximized (at 24% in consumption equivalence terms) when the government invests approximately $10,000 per child-year—which can be compared to the estimated costs of running a pre-K school of approximately $13,500 per child-year. This welfare gains are three times as large as if the same resources were used to fund the government transfer. In particular, the childhood investments program is associated with an income inequality reduction of 20%—enough to achieve Canada’s level of inequality—and an improvement in intergenerational mobility of 60%—the log-log income persistence coefficient would be reduced from 0.34 to 0.14, close to Nordic countries’ levels of mobility.

Comparing to the case in which the program is permanently implemented, we find that if such government investments were introduced for only one generation they would achieve only one-third of its benefits on welfare and mobility. We interpret this as evidence that a randomized control trial is likely to underestimate the long-run benefits of such a policy. General equilibrium forces are particularly important for inequality, as the increase in high-school graduates wages—relative to college graduates—explains two-thirds of the improvement of inequality. Even though benefits take time to accrue,

2These inefficiencies are explored theoretically in simpler models (e.g., Loury, 1981; Baland and Robinson, 2000). Other complexities in our model can also lead to inefficiently low levels of investment. For example, uncertain returns to those investments and incomplete markets can result in inefficiently low levels of investment. We highlight the other ones since they particularly depend on an intergenerational channel.
our transition dynamics evaluation suggests that if the policy was implemented permanently, half of the long-run mobility and welfare improvements would be achieved after 50 years—though this analysis depends on how the transition is implemented and financed. Later generations are the ones to benefit the most. Therefore, we find that if the government is able to temporarily increase its deficit in order to afford these extra investments, it would be able to increase the welfare of every new generation in the transition.

This paper also evaluates another popular policy regarding childhood development: parenting education. These programs focus on teaching parents techniques and games to solve discipline problems, and foster confidence and capability, among others. The key difference is that in this program, rather than investing towards children directly, it is parents who are trained on how to promote children’s development. We estimate the costs and returns of running a parenting education program based on RCT evidence, and find that such a program would have the potential of increasing welfare by 10% and substantially improve income inequality and mobility. Importantly, once again, we find that welfare and mobility improvements in the long-run general-equilibrium framework are larger than the ones we obtain by applying the policy as an RCT.

The rest of the paper is organized as follows. Section 2 discusses the literature. Section 3 presents our empirical findings on parental investments and returns to skills by education groups. Section 4 introduces the model, and Section 5 explains its estimation and conducts some validation exercises. The policy analysis exercises are presented in Section 6. Finally, Section 7 concludes. The Appendix contains additional details and exercises. There, we study the importance of endogenous childhood development on optimal taxation given that most of the models used for that analysis abstract from including endogenous intergenerational links like childhood development. We find that taking into account how childhood investments react to taxations implies that the optimal tax progressivity should be much lower than when childhood investments are not allowed to react.

2 Related Literature

Macroeconomic analysis of inequality can be generally divided in two strands. One set of the literature focuses on the top 1%, with a particular interest on taxation (e.g., Piketty and Saez, 2003; Diamond
and Saez, 2011; Saez, 2016). The other one focuses on the bottom 99%, looking particularly at the role of skills and education (e.g., Katz and Murphy, 1992; Autor et al., 2008; Abbott et al., 2013). In addition to taxation, they are also interested in college-education policies. Among quantitative models 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 we introduce new intergenerational linkages that allow us to endogenize those initial conditions and evaluate policies that target children. The closest model to ours is probably Abbott et al. (2013), 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—as well as multiple college alternatives. The dynamic interactions between borrowing constraints and parental investments in children development may be important since limited assets and borrowing can limit the capacity of parents to invest money towards their children, which, due to complementarities, may also reduce their incentives to invest time towards them. This affects the income and wealth of the next generation, which once again shapes their capacity and incentives to invest in their own children.

Previous literature on childhood development estimates the production function of children’s skills (e.g., Todd and Wolpin, 2003; Cunha et al., 2010; Boca et al., 2013). In particular, we include estimations from Cunha et al. (2010) in our model. However, unlike them, we model explicitly how investments are chosen by parents. This is necessary to study how policies affect parental investment choices and the welfare in an equilibrium framework. Previous papers have modeled childhood development (e.g., Abbott, 2016; Boca et al., 2013), but these papers have abstracted from general equilibrium forces (and saving decisions in the case of Boca et al.) which has prevented them from evaluating large-scale policies. An exception might be Yum (2016). Differently from his work, our model explicitly takes into account the findings of the empirical literature which highlights the need for multiple periods of parental investments. Moreover, we also include monetary inputs in the formation of skills, flexible parental transfers and allow for borrowing as observed in the data, which are important to study policies that are affected by borrowing constraints.\(^3\) Importantly, our model is able to capture two properties highlighted by the literature (e.g., Cunha et al., 2010): *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). To the best of our knowledge, no such model has taken

---

\(^3\)The set of policies we study is also different.
these into account and connected them with inequality or social mobility in an environment suitable for policy analysis.

Including parental investments allows us to evaluate policies which directly focus on childhood development which, as suggested by the literature, has potential to reduce inequality and promote intergenerational mobility. Previous theoretical papers have highlighted that an environment with intergenerational investments in skills can lead to under-investment (e.g., Loury, 1981; Baland and Robinson, 2000). Using simpler models, they can show that borrowing and parental transfers constraints (i.e., children cannot transfer resources to the parents) can lead to inefficiencies, which the government can improve on by enforcing higher investments towards children. We contribute to this literature by providing a model that is suitable to quantitatively evaluate the effect of such government interventions. Moreover, our model takes into account uncertainty in the returns to investments, general equilibrium effects, and the distortionary role of changing taxes to finance policies.

3 Empirical Findings

Given that we are going to study inequality and mobility focusing on the effect of parental investments in the development of skills, we first show evidence on parental time and money investments towards children. This evidence is certainly subject to endogeneity concerns, so we highlight that this evidence will only be used to inform the model construction and to compare with the model’s simulations. Then, we show evidence that these skills are important for hourly wages.

3.1 Parental Investments

The Panel Study of Income Dynamics (PSID) survey follows a nationally representative sample of over 18,000 individuals living in the US. For a subsample, the Children Development Supplement (CDS) provides information on 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 capacities: Letter-Word (LW) and Applied Problem Solving (AP). For most of our analysis we focus on the LW test, but use the AP to check the robustness of the results. In general, results are robust to the choice of the
test. Moreover, the most interesting feature about the CDS is that it provides detailed time diaries for each child. We can observe hour by hour what activity is being performed (e.g., reading or playing), if the activity is being performed with someone (e.g., the father is reading with the child), and if someone is around while the child is doing the activity (e.g., the father is working while the child is reading next to him). The time diary is available for both a weekday and a weekend day. We obtain these time diaries and test scores for the years 2002 and 2007. For details on the sample selection, see Appendix A.1. We highlight three points from this data: (i) Children’s test scores are good predictors of college graduation; (ii) Children’s test scores are associated with parents’ characteristics; and, more importantly, (iii) Parental investments towards children’s development are associated with parents’ characteristics.

**Children’s test scores are good predictors of college graduation**  We first focus on children who were tested between the ages 11 and 13, and group children in quintiles according to their scores.4 We then look at college graduation rates for these children. Importantly, the main PSID survey includes the households of children of people once interviewed. Individuals that were once children (and hence included in the CDS) are then included in the main PSID survey. These allows us to connect young-age test scores from the CDS to college outcomes from the PSID. In order to reduce the concern that children may still be finishing their college education, we focus on children observed (at least once) in the main PSID sample after they are 24 years old. Figure 1 shows that young-age skills are positively correlated with later college-graduation rates. Children that were in the top quintile of the LW score distribution at age 12 are six times as likely to graduate college than children that were in the bottom quintile.

4Since age is an important determinant of test scores, we remove the age trend using a second-degree polynomial.
Figure 1: College graduation rate and young-age skills

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

Children’s test scores are associated with parents’ characteristics   Since these young-age skills are important for later outcomes, it is important to know where the skill difference comes from. Here we provide some evidence that skills are correlated with parents’ socioeconomic group. Once again we look at children’s test scores, but in addition to the LW score we include the AP score. Figure 2 reports average standardized scores (i.e., standard deviations below or above the mean) for children from parents with certain characteristics. Using the PSID data we are able to identify the education group of the parents as well as estimate their permanent income.\footnote{Permanent income is estimated using all income parents had between while the child was under 16 years old.} Test scores are positively correlated with parents’ income and education. This is robust to the choice of the test. Children of college-educated parents are on average 0.5 standard deviations above children of high-school graduates. Similarly, children of high-income parents are approximately 0.7 standard deviations above those of low income parents.
Parental investments towards children’s development are associated with parents’ characteristics. The main benefit of the CDS dataset is that we can observe detailed diaries on the time parents spend with their children. Using these time diaries, we define “time with parents” if the parent is doing the activity together with the child. If two parents are performing the activity, we interpret this as double the hours since time constraints must hold for the household. First, we add up all activities to estimate total active time with parents per week. Once again we remove the age trend, and approximate the average for each child around age 4. Figure 3 suggests that there is not much difference 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.

Figure 3: Total time with parents and parent’s characteristic

![Bar chart showing weekly hours and active hours for HS Graduate and College Grad.]

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

We group these detailed activities into six groups in order to facilitate the analysis. “Mental Work” refers to activities like reading, doing homework, or having conversations. “Active Leisure” groups activities like playing games or sports. “Meals” adds up the amount of time parents share meals with their children. “Household Chores” refers to basic activities like cooking, showering or dressing (in the case of children, it is most likely being showered or being dressed). “Passive Leisure” entails listening to the radio or laying around without doing a clear activity. Finally, “TV/Video Games” refers to times watching TV or playing video games. Figure 4 shows the distribution of these activities across parents from different socioeconomic groups. Even though lower-educated or lower-income parents spend similar total time with their children, a disproportionate larger amount of that time is spent watching TV or playing video games. And a smaller amount of time is allocated to active leisure or sharing meals.
It is clear that these activities are different in the amount of interaction they entail between the parent and the child. Watching TV requires much less interaction than reading, playing games, sharing dinner or having a conversation. Since interaction is believed to be one of the key sources of cognitive skills development, we exclude TV/video games, passive leisure, and household chores from total time with parents in order to define “quality time with parents.”

Figure 5 shows that once we focus on quality time, parents from high socioeconomic groups spend more time with their children. College graduates spend 3.7 (21\%) more hours per week than high-school graduates. Similarly, high-income parents spend 6 (38\%) more quality hours per week with their children than low-income parents.

---

7The differences that emerge are robust to just removing time watching TV or playing video games.
These differences in the amount of quality time may help us explain the differences in skills developed by young children—which have been shown to have significant impacts in adult outcomes. Given typical concerns of endogeneity, the model in Section 4 provides a framework in which parents develop children skills and these skills are important for education and labor outcomes later in life. The technology for skill development is based on the estimates from Cunha et al. (2010), where the authors focus on reducing the concerns of endogeneity. Our model allows us to connect this with intergenerational mobility and inequality, and analyze the dynamic interactions between parental interactions and these two. The evidence presented in this section will only be used to inform the construction of the model, compare its estimated moments, and then test its validity.

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

Figure 6: Spending toward children and parent’s characteristic

![Graph showing spending by parent's education and permanent income](image)

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

As mentioned above, one very relevant element missing is that in order to access good public education parents need to live in specific areas in the US. These areas are usually associated with higher house prices as well. Hence, it is the case that part of the cost of living includes the amenities that come with the area of the house, including the quality of schools available. In order to try to capture this, we calculate rent expenditures (rent is imputed for house owners). There is a concern that higher income individuals might choose to live in bigger houses. Since all that is needed for a child to access public school is one room in the appropriate neighborhood, we divide the annual rent value of the house by the number of rooms. Figure 7 shows that high-income parents annually spend $1,600 (133%) more than low-income parents on rent per room. College-educated parents spend $800 (57%) more than high-school-educated parents.
Figure 7: Rent cost per room and parent’s characteristic

![Graph](image)

Parent’s Education

Parent’s Permanent Income

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

Other samples  This aggregate evidence includes substantial heterogeneity in the parents’ characteristics. In particular, given that the model used in this paper will not have fertility or marriage choices, we highlight that the moments used for the estimation are almost unchanged when we focus on a more homogenous sample. The first column refers to the whole sample, as studied above. The other two columns focus on families with the two parents living together while the child is under the age of 12. Moreover, the second column focuses on families with between 1 and 3 children, while the last column restricts it to cases with only two children. Weekly hours with children is the moment we will 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 do, the differences for sample averages are only 2 hours. Yearly expenditures (including rent) will not be used in the estimation as another estimate that includes child care will be used. Nevertheless, since that estimate does not control for potential single parenthood it is good to note that differences are not too large either. We interpret these findings as suggesting that the model estimation wouldn’t change significantly if we made the sample selection more restrictive.
Table 1: Parental Investments: summary for alternative samples.

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

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

The bottom four rows summarize the cross-sectional evidence which will be used in the validation of the model. Part of the heterogeneity of investments across children is reduced when we focus on more selective samples. For example, in the full sample children of college-educated parents tend to spend 19% more time with their parents but if we focus on children of married parents this difference is reduced to between 13 and 15%. Particularly interesting is that the regression coefficient of log(hours) on log(income) is insignificant for the married samples. On the other hand, the coefficient of log(expenditures) on log(income) remains significant and stable around 0.5 for all samples.

3.2 Education and returns to skill

We evaluate the effect of skills on wages across different education groups. Previous studies have highlighted that skills increase hourly wages, and that this return is higher among college-educated individuals than among lower-educated groups. We contribute to this literature by highlighting that this pattern is robust to focusing on two-adults-households. For details on the sample selection, see Appendices A.2 and A.3. Importantly, the model in Section 4 will display similar properties once
We estimate the wage process for high-school and college graduates separately, allowing for differences across age and skills, which will provide an estimate for the returns to skill. The wage process will be an important element for the model, as it determines the career profile—including the amount of uncertainty. We propose that the wage process of individual $i$ with education $e$ at age $j$ is given by
\[ w_i^e \epsilon^e_j \varepsilon^e_{ij} \]
where $\epsilon^e_j$ is the age profile for the education group and $\varepsilon^e_{ij}$ is the idiosyncratic labor productivity. This last one is specified as:
\[
\log(\varepsilon^e_{ij}) = \lambda^e \log(\theta) + \eta^e_j
\]
\[
\eta^e_j = \rho^e \eta^e_{j-1} + \varepsilon^e_j, \quad \varepsilon^e_j \sim iid N(0, \sigma^e_\varepsilon).
\]
where $\theta$ is the level of cognitive skills and $\eta^e_j$ 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 $\sigma^e_{\eta_0}$. The heterogeneity of the impact of skills on wages $\lambda^e$ across education groups will be particularly relevant for the education choices of agents with different abilities.

First, we use data from PSID to estimate the age profile $\epsilon^e_j$ by education groups separately. We include year fixed effects to control for possible changes in average wages over time.\(^8\) We estimate the age profile as a second order polynomial. We use PSID (instead of the National Longitudinal Survey of Youth, hereafter NLSY) because it includes a representative cross-section every year, so it avoids having the average age of the sample change directly with the calendar year.\(^9\) Appendix Table A3 reports the results from this estimation.

Second, we move to NLSY to identify the effect ability has on wages. NLSY is useful for this because we observe the Armed Forces Qualification Test (AFQT) score for these individuals. For each household, we remove the appropriate age profile estimated from PSID, and estimate the returns to skill.\(^10\) Consistent with previous evidence, we find that returns to skill are higher for college graduates than high-school

\(^8\) We include individuals from the over sample in our regression in order to improve the estimation power. We introduce a dummy to identify these individuals in our regression but results are robust to excluding them.

\(^9\) 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.

\(^10\) 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.
graduates.

Table 2: **Returns to skill by education group.**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High School</td>
<td>College</td>
</tr>
<tr>
<td>log(AFQT)</td>
<td>0.533***</td>
<td>0.904***</td>
</tr>
<tr>
<td></td>
<td>(0.0216)</td>
<td>(0.0456)</td>
</tr>
<tr>
<td>Observations</td>
<td>16,869</td>
<td>7,437</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.104</td>
<td>0.147</td>
</tr>
<tr>
<td># of households</td>
<td>3574</td>
<td>1479</td>
</tr>
</tbody>
</table>

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

Finally, the residual after removing the age and skill component is used to estimate the process for the idiosyncratic shock $\eta_j^e$. We allow for measurement error and use a Minimum Distance Estimator, i.e. we use as moments the covariances of the wage residuals at different lags and age groups, separately for each education group. Since the model we are going to use has 4-years-long periods, we estimate this income process grouping observations over 4 years. The results, reported in Table 3, suggest that shocks are persistent, particularly for higher educated groups.

Table 3: **Income process estimation: idiosyncratic.**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High School</td>
<td>College</td>
</tr>
<tr>
<td>$\rho^e$</td>
<td>0.893</td>
<td>0.945</td>
</tr>
<tr>
<td>$\sigma_z^e$</td>
<td>0.023</td>
<td>0.009</td>
</tr>
<tr>
<td>$\sigma_{\eta_0}^e$</td>
<td>0.044</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Source: NLSY. Methodology is explained in the main text.
4 Model

We specify a life-cycle economy in a dynastic framework with four main stages (20 periods total): childhood, college, labor, and retirement. Each period in the model refers to 4 years. During childhood, the individual’s parents invest in her skills—which will increase her earnings. Individuals become independent (i.e., start making choices) after finishing high school with a level of skills that depends on their parents’ investments as well as an amount of assets also decided by their parents. Idiosyncratic uninsurable income risk makes individual earnings stochastic. In this stage, individuals choose whether to go to college or start working. Once agents exit the education phase, they enter the third stage, which represents their labor market experience. Throughout their lives, agents choose their labor supply, savings, and consumption expenditures. They can borrow only up to a limit, and save through a non-state-contingent asset. The framework is one of uncertainty both in earnings as well as child’s skill development: Individuals choose how much time and money to invest in their children’s development, but the final outcome is uncertain. Before the children are of college age, individuals decide the amount of monetary resources to transfer to them. The last stage is retirement. At this time, agents have two sources of income: savings and retirement benefits. We study this economy in general equilibrium. An aggregate firm combines the different types of labor (by education) and capital to produce the final consumption good. The government levies taxes on labor, consumption, and capital in order to provide a lump-sum transfer and retirement benefits as well as finance some exogenous expenses. We now describe the model in detail.

4.1 The individual problem

Figure 8 shows the life cycle of an agent, in which each period in the model refers to four years. Let \( j \) denote the age in periods (e.g., \( j = 1 \) refers to ages 0–3 while \( j = 2 \) to ages 4–7). From \( j = 1 \) until \( j = J_i \), the child lives with her parents. At age \( j = J_i \), the child becomes independent. Her initial states are assets, skills, and school taste (or psychic cost). Initial assets are money transfers from her parents. The skill level and school taste are related to parents’ characteristics and investments towards their children, as explained below.

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

**Education stage:** At $j = J_i$ the agent has the option to go to college (for one period, until $j = J_e$). The individual state variables are savings $a$, skill $\theta$, and psychic cost $\phi$. The agent chooses whether to go to college or start working. The education decision is irreversible. All agents become independent as high-school graduates ($e = 1$). If an agent chooses to go to college, her education changes to $e = 2$. The monetary cost of education is $p_e$, but, as is common in the literature (e.g., Heckman et al., 2006; Abbott et al., 2013), we also allow for psychic costs $\phi$ to affect the total cost of education. Modeling school taste is necessary because resources available to finance schooling and returns to education can only partially account for the observed education patterns. Particularly, we assume that the school taste enters as a separate term in the value function. We scale the school taste $\phi$ by a constant $\bar{\psi}_\epsilon$. After leaving school, the psychic cost is assumed not to affect any adult outcome. While working, stochastic labor efficiency $\eta$ evolves following $\Gamma_{e,j}(\eta)$. Notice that we allow for education- and age-dependent idiosyncratic shocks. In Section 5, we discuss the estimation of the costs of education. The income process is based on the estimation from Section 3.

College students face borrowing limits $a^s$ for subsidized loans. They have access to subsidized loans at rate $r^s = r + \iota^s$ where $\iota^s < \iota$. To simplify computation, we assume that college student debt is refinanced.
into a single bond that carries interest rate $r^-$. $\tilde{a}^s(a')$ is the function performing this transformation. When making this calculation we assume that fixed payments would have been made for 5 periods (i.e., 20 years) following graduation.$^{11}$ Borrowing limit $a^s$ and wedge $\iota^s$ will be based on federal college loans, to be explained in detail in Section 5. College students are allowed to work but we also allow for studying to take time. We set this time to be $\bar{t}$ for all students, which will reduce the number of hours worked by students in the model. This 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.

Formally, let $V^s_j$ and $V^w_j$ be the value of an agent of age $j$ in school and working, respectively. $V^s_j (a, \theta, e, \phi)$ shows the value of an agent in college and with assets $a$, skills $\theta$, and school taste $\phi$. It is defined by

$$V^s_j (a, \theta, e, \phi) = \max_{c, a', l} u(c, l + \bar{t}) - \epsilon(\phi, \theta) + \beta E_{\eta|e} V^w_{j+1} (\tilde{a}^s(a'), \theta, e, \eta)$$

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

$$y = \text{w}^e_j(\theta) l, \quad a' \geq a^s, \quad 0 \leq l \leq 1 - \bar{t}, \quad \eta \sim \Gamma_e,0.$$

The agent is risk averse and her preferences are represented by $u(c, l)$ which is increasing and concave in consumption $c$ and decreasing and concave in labor $l$. Notice that we already assume that an agent cannot start with negative assets. She can borrow up to the limit $a^s$, and the return on positive savings is $1 + r$. However, if the agent decides to borrow she will pay interest rates $r^s > r$. Future is discounted by $\beta$. We denote as $w^e$ the wage for an agent who is currently in school at level $e$. We allow college students to provide high-school level labor.

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

$$V^w_j (a, \theta, e, \eta) = \max_{c, a', l} u(c, l) + \beta E V^w_{j+1} (a', \theta, e, \eta'),$$

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

$^{11}$Given the fixed payment nature of student loans and the assumption that they are repaid in 5 periods, we can transform college loans into regular bonds using the following formula: $\tilde{a}^s(a') = a' \times \frac{r^s}{1 - (1 + r^-)^{-5}} \times \frac{1 - (1 + r^-)^{-5}}{r^-}.$
\[ y = w^e e_j^e(\theta) l, \quad a' \geq a_{e,j}, \quad 0 \leq l \leq 1, \quad \eta' \sim \Gamma_{e,j}(\eta). \]

The agent can borrow up to the limit \( a_{e,j} \), and the return on positive savings is \( 1 + r \). However, if the agent is borrowing she pays interest rates \( r^- > r \). The return from working is the wage \( w^e \) scaled by \( e_j^e(\theta) \)—a function of the worker’s age, ability, and education. The disutility of working is in \( u \). Then, let \( V_{j}^{sw} \) be the value of an agent who can choose between working (as a high-school graduate) and going to college,

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

**Working stage:** From \( j = J_e \) until \( j = J_r \), the agent works and her individual problem is equivalent to (2). However, the problem will change when the children are born at the exogenously given fertility period \( j = J_f \). We assume that each agent has one children—or, alternatively, each household has one household offspring. For two periods the agent will have to choose the number of hours \( \zeta \) and amount of money \( m \) to invest in child’s development of skills. Moreover, once the children become independent (at \( j = J_k \)), the agent chooses the transfer \( \varphi \) to her offspring.

**With young children:** This is where the key novelties of the model are present. For two periods from \( j = J_f \), the agent works and also invests directly in her child’s development of skills \( \theta_k \). Hence, \( \theta_k \) is added as a state of variable at this stage. We normalize the initial distribution of \( \theta_k \) such all children are born with \( \theta_k = 1 \).\(^{12}\) We model altruism à la Barro and Becker (1989), in which parents care about the utility of their children. The skill development function consists of two nested CES functions. The outer CES is based on Cunha et al. (2010) but, differently from them, we model parental investments explicitly in the inner CES. The outer CES function states that next-period child’s skill \( \theta_k' \) depends current child’s skill \( \theta_k \), parents’ skills \( \theta \), and parental investments \( I \)—as well as an idiosyncratic shock \( \nu \).\(^{13}\) The inner CES function shapes parental investments using both time \( \tau \) and expenditures \( m \) towards children.\(^{14}\)

\(^{12}\)This assumption simplifies the problem as it reduces the state space by one variable. Importantly, it is also in line with the epigenetic literature which, as summarized by Cunha and Heckman (2010), suggests that gene and environmental influences cannot be meaningfully separated.

\(^{13}\)We abstract from non-cognitive skills due to computation reasons. Given that evidence suggests that cognitive skills have higher returns on wages we focus on these. Non-cognitive skills may be important for the development of children’s skills but we leave that for future research.

\(^{14}\)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.
\[
V_j(a, \theta, e, \eta, \theta_k) = \max_{c, a', l, \tau, m} u(c, l) - v(\tau) + \beta \mathbb{E} V_{j+1}(a', \theta, e, \eta', \theta_k'),
\]

(3)

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

\[
y = w^e \varepsilon_j^e(\theta) l, \quad a' \geq a_{e,j}, \quad 0 \leq l + \tau \leq 1, \quad \eta' \sim \Gamma_{e,j}(\eta)
\]

\[
m \in \{m_1, m_2, \ldots\}, \quad \tau \in \{\tau_1, \tau_2, \ldots\}
\]

\[
\theta_k' = [\alpha_j \theta_k \phi_j + \alpha_j \theta^{\phi_j} + \alpha_j I^{\phi_j}]^{1/\phi_j} e^\nu, \quad \nu \sim N(0, \sigma_{j,\nu})
\]

\[
I = \overline{A} [\alpha_m m^\gamma + (1 - \alpha_m) \tau^\gamma]^{1/\gamma}.
\]

After two periods, child’s skills are fixed, so the problem is equivalent to (2) but with the extra state variable of \(\theta_k\).\(^{15}\) Then, at \(j = J_k\) (i.e., when the child is of age \(j = J_i\)) the agent needs to decide the size of monetary transfers \(\varphi\) 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 \(V_{\text{Transfer}}\). The period before this happens is a standard working period but with the additional state for the child’s skills (which are assumed to be fixed at this stage).

\[
V_{j_{k-1}}(a, \theta, e, \eta, \theta_k) = \max_{c, a', l} u(c, l) + \beta \mathbb{E} V_{\text{Transfer}}(a', \theta, e, \eta', \theta_k'),
\]

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

\[
y = w^e \varepsilon_j^e(\theta) l, \quad a' \geq a_{e,j}, \quad 0 \leq l \leq 1, \quad \eta' \sim \Gamma_{e,j}(\eta).
\]

Then the agent needs to decide how to distribute her assets between herself and her child. We assume that this transfer to the child needs to be non-negative. When making this choice, the parent already

\(^{15}\)This assumption simplifies the solution but is also in line with the evidence from the CDS data that shows that while children under the age of 8 tend to transition between skill’s ranks, this is not much the case for children above that age. In particular, we assign children to terciles of the skills distribution and look at the transition matrix between those terciles over time. We find that the trace of the transition matrix grows from under 1 for children between 2 and 8 years old to over 1.6 for children between 8 and 12 years old.
knows the realization of his income shock $\eta$, but is not aware of the school taste draw $\phi_k$ of her child.

$$V_{\text{Transfer}} (a, \theta, e, \eta, \theta_k) = \max_{\varphi} V_{J_k} (a - \varphi, \theta, e, \eta') + b \mathbb{E} V_{J} (\varphi, \theta_k, \phi_k),$$

$$\varphi \geq 0, \quad \phi_k \sim N(\bar{\phi}_e, \sigma_\phi) \quad (4)$$

Notice that unlike (3), the value function at this stage now includes the continuation value of the children $V_{J_i}$. This is the last period in which parents’ choices affect their descendants. As the problem is written recursively, this implies that at every period in which parents’ choices affect children’s outcomes, the value function of their descendants will be taken into account. This embeds the parental altruism motives. The psychic costs of the children are stochastic but correlated with the parents’ level of education. Moreover, recall that the effective school distaste from equation (1) may also depend on the skill level of the agent. The functional form of the altruism, as well at the stochastic processes of skills and psychic costs are specified in Section 5. After the agent’s children become independent, the individual problem is equal to (2).

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

$$V_j (a, \theta, e) = \max_{c, a'} u (c, 0) + \beta V_{j+1}^w (a', \theta, e),$$

$$c + a' = \theta + \pi (\theta, e) + a (1 + r),$$

$$a' \geq 0,$$

where $\pi$ are the retirement benefits, which depend on the education and skill level.\footnote{We use education together with the skill level, as a proxy to approximate average lifetime income with which are the retirement benefits determined. See Section 5 for details.}
4.2 Definition of Stationary Equilibrium

The model is solved numerically to characterize the stationary equilibrium allocation. The model includes $J_d$ overlapping generations. Stationarity implies that we study an equilibrium in which the cross-sectional distribution for any given cohort of age $j$ is invariant over time periods. Particularly important is that the distribution of initial states is determined by the choices of the older generations. The equilibrium allocation requires that households choose education, consumption, labor supply, parental time and money investments, and parental transfers such that they maximize their expected utility, firms maximize profits, prices clear markets.$^{17}$ See Appendix for details.

4.3 Role for government

As reflected in the model, children cannot invest in their own early childhood or pay their parents to do so. This missing market can lead to inefficient early childhood investments when parents are not perfectly altruistic (i.e., $b < 1$), parental transfers constraints bind, or borrowing constraints bind. Baland and Robinson (2000) use a simpler model of parental investments (though to study child labor) and highlight the theoretical role of each of these channels. Their discussions are useful to illuminate the intuitions here since theoretical conclusions are harder to obtain with our more complex model. We may think that perfectly altruistic parents will always invest efficiently towards their children. However, even if parents were perfectly altruistic, they may want to transfer resources from their children—particularly if they expect their children to be better off than themselves. Binding transfer constraints may not allow that, leading to under-investment as a way to smooth intergenerational consumption. If children could borrow when they were young, they could transfer resources to their parents and compensate them for the increased investments. Alternatively, even if children could compensate their parents by transferring resources to them, borrowing constraints can also lead to inefficiently low investments. In this case, borrowing constrained parents may use reduced parental investments as a way of transferring resources from future periods.

$^{17}$The government is allowed to have other expenses which are wasteful since the estimation will be designed such that the income redistribution, rather than the government’s budget, is matched. However, whenever we introduce a policy in Section 6 this is done such that this extra government expenses remain fixed—such that the government’s budget remains constant.
This missing market creates a role for the government. By providing or introducing mandatory childhood investments, the government can help increase childhood investments. Loury (1981) provides a simpler partial-equilibrium model in which he can show that government investments towards childhood are welfare improving even when (lump-sum) taxes are needed pay for them. In Section 6 we will quantitatively evaluate such a policy in our more developed and estimated model, which also takes into account general equilibrium effects.

5 Estimation

The model is estimated to match household level data. Therefore, an agent in the model corresponds to a household with two adults in the data. Consequently, every household in the model has one household as offspring. We use three primary data sources: (i) Panel Study of Income Dynamics (PSID); (ii) Child Development Supplement (CDS) to PSID; and (iii) 1979 cohort of the National Longitudinal Survey of Youth (NLSY79). We select a population for which our model can be taken as a reasonable approximation to household behavior and impose two main selection criteria on the data. First, as is standard in the literature Huggett et al. (2011), we drop household observations with income below a certain threshold. We choose this threshold as the one that corresponds to one person working 20 hours a week for the minimum wage (approximately $8,000 total annual household income). Second, there is no decision regarding fertility in our model, so for the moments on parental investments we focus on households with two children. There is no marriage decision in our model, so to avoid differences in income and time availability due to single parenthood, we keep only married households. Details about sample selection are reported in Appendix. We numerically solve the steady state of this economy. We then compute the ergodic distribution of the economy to match moments from the US in 2000. All prices mentioned are in 2000 dollars.

We describe below how we parametrize the model economy. Some of the parameters can be estimated “externally,” while others must be estimated “internally” from the simulation of the model. Table 4 summarizes all the parameters in the model.
5.1 Preliminaries

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

Prices: Prices are normalized such that the average income of a high school graduate at age 48 is equal to one in the model. In the (PSID) data, this income is equal to $58,723. 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 $\bar{a}(e)$ for high-school and college graduates separately. The limits are set to $\{-24,000, -34,000\}$ for high school and college graduates, respectively. The yearly price of college is from the Delta Cost Project, where we get $6,588.\textsuperscript{18}

Replacement benefits and taxes: The pension replacement rate is based on the Old Age, Survivors, and Disability Insurance federal program. We use education and skill level to estimate the average lifetime income, on which the replacement benefit is based. See Appendix C.1 for details. The tax function is assumed to be $T(y,a,c) = \tau_y y + \tau_k a r I_{a\geq0} + \tau_c c - \omega$. Based on McDaniel (2007), we set $\tau_y = 0.27$, $\tau_k = 0.4$, and $\tau_c = 0.08$. $\omega$, the government’s lump sum transfer to households, will be estimated such that the government’s budget clears after paying for retirement benefits.

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—only 13% took out non-federal loans. Moreover, the average loan value was similar for both federal and non-federal cases. Since average values were similar but federal loans were significantly more common, we focus on federal loans for our model estimation. Among federal loans, the Stafford loan program was the most common: 96% of the undergraduates who borrowed took out Stafford loans. The second most common loans were the Perkins loans, but they were much smaller: only 11% of borrowers used

\textsuperscript{18}We take into account grants and scholarships, such that only private tuition costs are considered.
Perkins loans and average amounts were one quarter of average Stafford amounts. Therefore, we focus particularly on Stafford loans. Stafford offers multiple types of loans so we use the weighted average interest rate to set $\iota^s = 0.009$. The borrowing limit while in college in the model is the set to match the cumulative borrowing limit on Stafford loans ($23,000$).

**School Taste:** In this class of models, it is difficult to match the education distribution. Previous studies introduced non-pecuniary (psychic) costs of education (e.g., Abbott et al., 2013; Krueger and Ludwig, 2015). We assume that school (dis)taste is utility terms follows the function $\epsilon(\phi, \theta) = \exp(\alpha + \alpha_\theta \log(\theta) + \phi)$. The function allows for higher skilled individuals to have (on average) lower levels of school distaste. Then, $\phi$ is an idiosyncratic shock which is assumed to follow a normal distribution $N(\bar{\phi}_{ep} - \frac{\sigma_{\phi}^2}{2}, \sigma_{\phi})$ 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 $\alpha$ will be estimated to match the college graduation shares, $\alpha_\theta$ will be estimated to match the relation between college graduation and skills, $\bar{\phi}_{ep}$ will be estimated to match the intergenerational persistence of education, and $\sigma_{\phi}$ will be estimated to match the variance in college graduation after controlling for skills.

**Wages:** We assume that wages follow the process estimated in Section 3. Recall that this estimation suggested that skills have higher returns for college graduates.

**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 the unobserved factor in their estimation. We assume that the child development function is of the nested CES form. The outer CES is based Cunha et al. but we parametrize the investment factor as the inner CES, with parental time and monetary as inputs. Hence, the functional forms are

$$\theta'_k = \left[\alpha_{j,\theta_k} \theta_k^{\phi_j} + \alpha_{j,\theta} \theta^{\phi_j} + \alpha_{j,I} I^{\phi_j}\right]^{1/\phi_j} e^\nu, \quad \nu \sim N(0, \sigma_{\nu})$$

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

We allow parameters in the outer CES to vary depending on the age of the children. Cunha et al. provide various estimates for these parameters but we choose the one based on the model closest to ours, which leads to $\alpha_{j,\theta_k} = \{0.30, 0.45\}$, $\alpha_{j,\theta} = \{0.38, 0.45\}$, $\alpha_{j,I} = \{0.32, 0.10\}$, $\sigma_{\nu} = \{0.44, 0.22\}$, and
\( \phi_j = \{-0.18, -0.78\} \). Then, we estimate \( \alpha_m \) to match the average ratio of money to time investments, \( \phi_I \) to match the correlation between the two investments, and \( \bar{A} \) to match the efficiency of these investments such that the average level of skills in the estimated economy is one. Recall that the average number of hours with children is also a moment of the estimation. Then, since we estimated the income process under the normalization that the average level of skills is one, the estimated economy should also display this property.

**Preferences:** We specify the period utility over consumption and labor as

\[
u(c) = \frac{c^{1-\gamma_c}}{1-\gamma_c} - \mu \frac{\ln 1 + \gamma_l}{1 + \gamma_l}.
\]

We follow the literature and assume that \( \gamma_c = 2 \) and \( \gamma_l = 2 \). \( \mu \) is estimated to match average hours of labor. When parents choose their time with children \( \tau \), the disutility is assumed to be linear, i.e.,

\[
v(\tau) = \xi \tau.
\]

\( \xi \) is estimated to match average hours with children. Finally, the altruism factor \( b \) is estimated to match the intergenerational persistence of income.\(^{19}\)

**Aggregate Production Function:** We assume there is an aggregate firm with production function

\[
Y = K^\alpha H^{1-\alpha},
\]

where \( H \) is a CES aggregator of labor supply of the two education groups.

\[
H = [sL_1^\rho + (1 - s) L_2^\rho]^{1/\rho}.
\]

We set \( \alpha = \frac{1}{3} \) and estimate the CES aggregator. We estimate \( \rho = 0.3 \). This leads to an elasticity of substitution between high-school and college graduates of \( \frac{1}{1-\rho} = 1.4 \) which is in line with previous estimates (e.g., Katz and Murphy, 1992; Heckman et al., 1998). We also estimate both \( s \) to be approximately \( \frac{1}{2} \). See the appendix for details on the estimation.

Twelve parameters of the model are estimated using Simulated Method of Moments. \( b \) relates to the degree of altruism, while \( \mu \) and \( \xi \) are the disutility of labor and time with children. \( \alpha, \alpha_g, \bar{\phi}, \) and \( \sigma_\phi \) relate to the distribution of school taste and its relation with skills and parental education. \( \bar{A}, \alpha_m, \) and \( \gamma \) relate to the effect of parental time and money investments in building skills. \( \omega \) relates to the government’s redistribution of income. Finally, \( \iota \) is the wedge in the interest rate between saving and

---

\(^{19}\)This parameter could also be estimated to match average monetary transfers from parents to children. However, this moment is hard to estimate in the data. On the other hand, recent availability of administrative data has improved the estimation of intergenerational mobility of income so we prefer to use this moment.
borrowing.

5.2 Simulated Methods of Moments: Results

Table 4 shows the estimated parameters and moments in the simulated economy. The model provides a good fit of the data. Average parent-to-child transfer are well fitted by the model. Education distributions, its relation with skills, and its intergenerational persistence are close to their data counterparts. Average time working and with children are successfully matched. The relation between money and time investments is well captured in the model. Finally, the share of borrowers in the simulated model is similar to the one found in the Survey of Consumer Finances. We also remark that the average productivity of parental investments is selected such that the average level of skills in the economy is around its normalized value of one. The income redistribution in the model, as measured by the ratio of the variances of log disposable-income and log pre-government income, is close to its empirical estimate.
### Table 4: Estimation: parameters and moments.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
<th>Moment</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preferences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>0.26</td>
<td>Altruism</td>
<td>Parent-to-Child Transfer (k)</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td>$\mu$</td>
<td>358</td>
<td>Mean Labor Disutility</td>
<td>Avg. hours worked</td>
<td>64.5</td>
<td>64.3</td>
</tr>
<tr>
<td><strong>School Taste:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>9.0</td>
<td>Avg. taste for College</td>
<td>College share</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td>$\alpha_\phi$</td>
<td>-2.4</td>
<td>College taste-Skills relation</td>
<td>College-Skills slope</td>
<td>0.30</td>
<td>0.31</td>
</tr>
<tr>
<td>$\sigma_\phi$</td>
<td>2.8</td>
<td>SD of College taste shock</td>
<td>College: residual variance</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>$\phi$</td>
<td>-1.6</td>
<td>Draw of school taste:</td>
<td>IGE persistence of education</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean by parent’s education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Skill Formation Productivity:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.3</td>
<td>Parental time disutility</td>
<td>Avg. hours with children</td>
<td>19.1</td>
<td>18.4</td>
</tr>
<tr>
<td>$\bar{A}$</td>
<td>191</td>
<td>Returns to investments</td>
<td>Average skill ratio</td>
<td>1.0</td>
<td>1.01</td>
</tr>
<tr>
<td>$\alpha_m$</td>
<td>0.74</td>
<td>Money productivity</td>
<td>Ratio of money to hours</td>
<td>264</td>
<td>269</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.49</td>
<td>Money-Time substitutability</td>
<td>Money-Time correlation</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td><strong>Interest rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\iota (\times 10^2)$</td>
<td>2.1</td>
<td>Borrow-save wedge</td>
<td>Share of borrowers</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td><strong>Government</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega (\times 10^2)$</td>
<td>2.5</td>
<td>Tax rebate</td>
<td>Income Var Ratio:</td>
<td>0.70</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Disposable to Pre-Gov</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Mobility is based on Chetty et al. (2016), correcting for income as in our PSID sample. Hours worked and skill formation moments are estimated from PSID-CDS data. Share of borrowers is estimated from Survey of Consumer Finances. College share is based on Census (2000). STEM major share (within college), college-skills slope, and college residual variance are estimated using NLSY. Intergenerational persistence of education is based on Checchi et al. (1999).

### 5.3 Validation Exercises

We can test the validity of the estimated model by looking at moments that are not directly targeted. Table 5 summarizes this validation results. As shown in Section 3, families from higher socioeconomic groups tend to invest more time and money towards their children. The estimated model displays similar qualitative features. The elasticities of log-hours to college and income are similar in the data and the model. Regarding the elasticities of log-expenditures, the model tends to produce larger elasticities than the data. This may be due to the fact that expenditures typically associated with higher-income groups (i.e., child care and school fees) are not available in the data used to calculate these elasticities.
Chetty et al. (2014) estimates intergenerational mobility, as measured by the log-income persistence across generations, to be 0.34. In the model this persistence is 0.33. Income inequality is also well captured by the model: both the Gini coefficient and the top-bottom are similar to the data.\textsuperscript{20} Regarding savings, the (annualized) capital-output ratio in the estimated model is 3.1 which is above but close to its typical empirical estimate of 3.

Table 5: \textbf{Validation:} Not targeted moments.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Data</th>
<th>Model</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investments towards children:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hours to College</td>
<td>0.19</td>
<td>0.18–0.21</td>
<td>CDS</td>
</tr>
<tr>
<td>Expenditures to College</td>
<td>0.45</td>
<td>0.77–0.82</td>
<td>CDS</td>
</tr>
<tr>
<td>Hours Elasticity to Income</td>
<td>0.18</td>
<td>-0.12–0.27</td>
<td>CDS</td>
</tr>
<tr>
<td>Expenditures Elasticity to Income</td>
<td>0.50</td>
<td>0.98–1.48</td>
<td>CDS</td>
</tr>
<tr>
<td>Intergenerational Mobility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-Income</td>
<td>0.34</td>
<td>0.33</td>
<td>Chetty</td>
</tr>
<tr>
<td>Labor Income Inequality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gini</td>
<td>0.32</td>
<td>0.32</td>
<td>PSID</td>
</tr>
<tr>
<td>Top-Bottom</td>
<td>3.9</td>
<td>3.7</td>
<td>PSID</td>
</tr>
<tr>
<td>Savings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital-Output Ratio (annualized)</td>
<td>$\approx 3$</td>
<td>3.1</td>
<td></td>
</tr>
</tbody>
</table>

\textit{All moments are computed using the estimated parameters. Chetty refers to Chetty et al (2016).}

Finally, we use experimental evidence to test the validity of the most important novelty in our model: childhood development. From the steady state, we simulate a policy in which the government unexpectedly introduces money directly in the early development of some children. This involves adding money directly to $m$ in the children’s development function in equation 6—unexpectedly, when they are 0–3 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 $I = \bar{A} [\alpha_m (m + g)\gamma + (1 - \alpha_m)\tau\gamma]^{1/\gamma}$. This experiment is in line with the one studied by García et al. (2016), in which children were randomized into early childhood development programs that costed approximately $13,500 per year.\textsuperscript{21} Table 6 shows that our results

\textsuperscript{20}Top-bottom refers to the ratio of average incomes between the top 80–95 percentiles and the bottom 5–20 percentiles.

\textsuperscript{21}At the time of the study the cost was $18,000 but adjusting prices by inflation (to obtain prices in 2000 dollars) brings the cost down to $13,500.
regarding education, both on years of education and college graduation, are in line with their findings. The success of these validation exercises gives us confidence in the policy evaluations we perform next.

Table 6: Validation: Experimental evidence.

<table>
<thead>
<tr>
<th></th>
<th>RCT</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of Education</td>
<td>0.35</td>
<td>0.41</td>
</tr>
<tr>
<td>Share College Graduates</td>
<td>13%</td>
<td>11%</td>
</tr>
</tbody>
</table>

All moments are computed using the estimated parameters. We simulate experimental evidence on early childhood development in the spirit of the study of Garcia, Heckman et al. (2016)—GHLP in the table. We simulate paying monetary expenses of a value of $18,000 per child in the first period of childhood development (once again, unexpectedly and for only one cohort). We then compare our findings regarding education to those of Garcia, Heckman et al for male individuals.

6 Policy Evaluation

6.1 Defining Welfare Measure: Consumption Equivalence

In this section we will evaluate a government intervention that directly increases childhood investments. When evaluating policies we will particularly interested in inequality, intergenerational mobility, and average income. As a summarizing measure that allows us to compare policies we will also look at consumption equivalence under the veil-of-ignorance, in the 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 baseline economy. We refer to consumption equivalence as the percentage change in consumption \( \lambda \) in the baseline economy that makes agents indifferent between being born in the baseline economy \( (P = 0) \) and the one in which the policy \( P \) is in place. In particular, let \( \tilde{V}_{J_i}^P (a, \theta, \phi, \lambda) \) be the welfare of agents with initial states \( (a, \theta, \phi) \) in the economy \( P \) if their consumption (and that of their descendants) were multiplied by \( (1 + \lambda) \):

\[
\tilde{V}_{J_i}^P (a, \theta, \phi, \lambda) = \sum_{j=J_i}^{J_d} \beta^{(j-J_i)} u \left( c_j^P (1 + \lambda), h_j^P \right) + \beta^{J-x} bE \tilde{V}_{J_i}^P (\varphi, \theta_k, \phi', \lambda)
\]

where, for the sake of clarity, I have abstracted from including the school distaste and highlighting that policy functions depend on the states. Notice that these policy functions are assumed to be unchanged.
relative the economy with policy $P$ when evaluating this. For example, consumption $c^P$ refers to the consumption chosen by individuals in economy $P$ and is unchanged by $\lambda$. Then, for any $\lambda$ we can obtain a measure of average welfare by averaging this function $\bar{V}^P(\lambda) = \int_{\mu_P} \bar{V}^P_{ij}(a, \theta, \phi, \lambda)$, where $\mu_P$ refers to the distribution of initial states $\{a, \theta, \phi\}$ in the economy $P$. Then, we define the consumption equivalence $\lambda^P$ 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.,

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

Having explained how we will measure welfare, we are now able to move forward with the policy evaluation.

### 6.2 Government Investments towards Childhood Development

As reflected in the model, children cannot invest in their own early childhood or pay their parents to do so. Imperfect altruism, borrowing constraints, and parental transfers constraints can all lead to inefficient levels of childhood investment—as explained in Section 4.3 and explored theoretically in simpler models (e.g., Loury, 1981; Baland and Robinson, 2000). In particular, Loury (1981) provides a simpler partial-equilibrium model and shows that government investments towards childhood are welfare improving even when (lump-sum) taxes are needed pay for them. Here we will quantitatively evaluate such a policy in our more developed model, which is more appropriate for quantitative evaluations. In particular, our model features multiple periods and forms of parental investments, estimated returns to those investments, distortive taxation and general equilibrium effects.

We simulate a policy in which the government invests money directly in the development of children. This involves adding money directly to $m$ 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

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

Parents can now alter their parental investments. In particular, we would expect that this would lead to crowding out of expenses $m$. It is possible, however, that due to complementarities time investments $\tau$ may increase. Notice that our validation exercise, as shown in Table 6, directly tested this mechanism by introducing government investments as a randomized controlled trial and using experimental evidence to compare the results. Its success gives us confidence in the policy evaluations we perform now. Once again we introduce government investments but in a
permanent manner, and we are going to finance it by altering taxes such that the government budget is unchanged. In particular, we adjust government (labor) taxes $\tau$. Moreover, we will look at the long-run effect of such policy, taking into account general equilibrium effects.

We will consider two alternative policies for government investments towards children. First, government invests directly only during the children’s first period (i.e., between the ages 0–3). Second, government invests directly on both periods equally (i.e., between ages 0–3 and 4–7). In order to compare the welfare increase these achieve with the one that children could have obtained by themselves using the same funds, we also introduce a third policy which simply provides agents with a government transfer when they become independent. We evaluate these three policies for different amounts of resources available. Figure 9 shows the results. The horizontal axis refers to the resources available for each child-year, when they are between the ages 0 and 7. In other words, $1,000 implies $1,000 = $8,000 per child.

Introducing government childhood investments leads to substantial welfare gains, of up to 24% as measured by the consumption equivalence. The top-right figure shows that the maximum welfare gain is achieved by investing $10,000 per child-year, distributed equally across the two periods. As a reference point, the early childhood program evaluated by García et al. (2016) was estimated to cost approximately $13,500 per child-year (in 2000 dollar terms). The childhood investment policies here are associated with only small changes in the tax-rate since the higher investments lead to higher levels of productivity, which imply higher tax revenues. This is not true for the case of the transfer since this policy has smaller direct effects on income. In particular, childhood investment policies raise the income of individuals in the bottom decile by almost twice as much than at the mean, which is associated with a reduction of income inequality of up to 20%. This reduction in inequality would be enough for the US to achieve Canadian levels of inequality. By reducing the number of children from poor families who receive low childhood-investments, this government policy is able to increase intergenerational mobility by 60%, i.e. reducing the persistence of log-income from 0.34 to 0.14. This level of mobility is currently observed only at the countries that display the highest levels of intergenerational mobility. In general, agents prefer these resources to be used to for childhood investments rather than to fund a government transfer given to them as they become independent. This happens because the government can do something that these agents cannot do by themselves, i.e., invest in their childhood.
Government investments towards childhood achieve most of its benefits on welfare and mobility through the long-run cumulative change in the initial distribution. We focus on the policy that achieved the largest welfare gains: investing $10,000 per child-year, equally distributed across the two periods. Table 7 decomposes the welfare gains by simulating the same policy applied in three alternative ways. First, we introduce government investments for only one-generation, without balancing the government’s budget or taking into account general equilibrium effects. Effects are evaluated on the generation that received the intervention. This implementation can be compared to the one to be expected by an RCT which is typically small-scale and individuals do not expect future generations to also be included in the intervention. We find that in this case welfare and mobility effects are only one-third of the ones obtained in our benchmark. Inequality is reduced by only one-tenth of the benchmark reduction.
Next, we allow for long-run effects to take place: the policy is implemented permanently and the fact that by improving one-generation’s level of skills we are also affecting the productivity of future investments is also taken into account. Moreover, by permanently introducing this policy, the chances of children becoming low-skilled are reduced and so are the chances of being born into a low-skilled family. Now we obtain welfare and mobility effects close to those obtained in the benchmark. However, inequality improvement is still only one-third of the one obtained in the benchmark. As we force the government’s budget to be balanced we find that welfare effects actually increase. This is due to the fact that the income tax is actually reduced, since this policy increases net revenues. Finally, general equilibrium forces reduce the wages of college-graduates relatively to high-school graduates. This reduces welfare, particularly through the effect on children of college-educated high-skilled parents who are most likely to be college graduates themselves. On the other hand, inequality is particularly improved: general-equilibrium forces account for two-thirds of the inequality improvement obtained.

Children born to low socioeconomic groups are the ones benefited the most by this policy. Figure 10 shows the policy’s welfare once the veil-of-ignorance is partially revealed. Here, we look at the consumption equivalence for an agent who knows her parent’s skill level and education group, assuming this is unchanged between the baseline economy and the one in which the policy is in place. Children of high-school educated and low-skilled parents benefit the most, with a consumption equivalence of 20%. On the other hand, agents that know are going to be born to college-educated and high-skilled parents benefit much less, with a consumption equivalence of 2%. On its right, Figure 10 also shows the share of children being born to each of these groups in both economies. It is clear that once the policy is introduced the distribution of skills moves upward, reducing significantly the share of individuals born to low-skilled parents.

Table 7: Decomposition:

<table>
<thead>
<tr>
<th>Long-Run</th>
<th>Budget Balanced</th>
<th>General Equilibrium</th>
<th>Change from Baseline (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>No</td>
<td>Cons. Equiv.</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

36
6.3 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 and possibly go through periods in which welfare is reduced. We evaluate this by looking at the transition dynamics. However, transition dynamics provide a lot of flexibility in its implementation (e.g., how is policy introduced and/or financed in the transition). To the best of our knowledge, no paper has studied the optimal transition with such degree of flexibility. Only Bakış et al. (2015) has looked into 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 much more complicated model but is out of the scope of this study. However, here we highlight that it is possible for this policy to be welfare improving for every generation in the transition, at least if we assume government has the capacity to borrow and save during the transition, and only need to be balanced in net present value once the whole transition is taken into account.

We assume that the government investments are introduced at the optimal level, unexpectedly, and are known to remain in place for ever. Whether the policy is welfare improving for every generations depends on how the transition is financed. The agents being benefited the most are the ones born in the future, as they receive both the government investment and higher skilled parents. Earlier generations
also receive the government investment, but are not as likely to be born to high-skilled parents as that change takes time to accrue. If the increase in government spending is to be financed by the earlier generations it is less likely that the policy will be welfare improving for them. Figure 11 shows an example in which the policy is being financed by a temporary reduction in government transfers, which goes back to its baseline level after 100 years. This transition path is welfare improving for all the generations.

Figure 11: Transition Dynamics

Note: Consumption equivalence, intergenerational mobility, and income are shown for the generations born in the period defined by the horizontal axis.
6.4 Other Policies

In Appendix D we evaluate another popular policy regarding childhood development: parenting education. These programs focus on teaching parents techniques and games to solve discipline problems, foster confidence and capability, among others. The key difference is that here, rather than investing towards children directly, parents are trained on how to promote children’s development. We estimate the costs and returns of running a parenting education program based on the randomized control trials evidence from Gertler et al. (2013) and Attanasio et al. (2016). A relevant caveat is that this evidence is from developing countries, but we try to control for that as explained in the Appendix. In particular, we implement this program as a government policy (both in a partial-equilibrium framework—similar to an RCT— and in a general-equilibrium) as well as a new education private market that parents can acquire. Our results suggest that such a program has the potential of increasing welfare by 10% and substantially improve income inequality and mobility. Importantly, once again we find that welfare and mobility improvements in the long-run general-equilibrium framework are larger than if we apply the policy as an RCT. The long-run cumulative change in the distribution is important to obtain all the benefits, since a higher-skilled distribution of parents will provide better conditions for the children. Transition dynamics results suggest that one-third (two-thirds) of the benefits are obtained after one (two) generation has gone through the program.

We also study the importance of endogenous childhood development on optimal taxation—within our constrained tax function. Most macroeconomic analysis of inequality focuses on progressive taxation. However, most of the models used for that analysis abstract from including endogenous intergenerational links like childhood development. In addition to the traditional trade-off between equality and efficiency of labor, endogenous childhood skills leads to a new trade-off. On the one hand, higher progressivity may let poor parents increase investments towards child’s skills. On the other hand, such progressivity would increase labor distortions, reducing the incentive to invest towards children’s skills. In general, the question that emerges is: Do tax policy evaluations change once we include endogenous parental investments? The model introduced here is useful to answer this question as it adds endogenous childhood development and parental transfers to a standard life-cycle macroeconomic model. In order to evaluate the importance of endogenous intergenerational links, we compare the effects in our model with endogenous links to the effects obtained by the same model but with exogenously fixed links. The model with exogenous childhood development is equivalent to the original model, but where the
intergenerational transition matrix of skills is fixed to be equal to one obtained (endogenously) in the original steady state. This matrix defines that the distribution of children’s skills depends (exclusively) on parents’ skills and education group. Hence, the tax system cannot affect the development of skills directly, though it may affect it through education choices. The model with exogenous childhood skills predicts small welfare increases from increasing tax progressivity, in line with the literature that uses this kind of models to evaluate tax progressivity (e.g., Floden and Lindé, 2001; Conesa and Krueger, 2006; Heathcote et al., 2014). On the other hand, the full model developed here predicts large welfare gains from substantially reducing tax progressivity. By increasing the incentives to invest towards children, a less progressive taxation can increase welfare in the long run. However, a more flexible tax function may allow for conditional transfers that increase progressivity, allowing poor parents to invest without harming incentives. We leave this for future research.

7 Conclusion

Doepke and Tertilt (2016) argue that there is a potentially large role for family economics within macroeconomics. This paper moves in that direction by combining a macroeconomic model that is appropriate for policy analysis on income inequality and intergenerational mobility, with the findings on childhood development. We use a standard macroeconomic Aiyagari-style life-cycle general-equilibrium model and introduce parental investment in the cognitive skills of children—which is later associated to 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. College can be financed either with parental transfers (which are endogenous) or through working and (subsidized) borrowing. The model also includes endogenous labor supply which is important for the financing of the policies to be evaluated.

Agents cannot invest towards their own childhood development or pay their parents to do so, which can lead to inefficiently low levels of investment. This paper shows that under-investment in children’s development is relevant for the macroeconomic analysis of inequality and social mobility, and can be improved by government policies that target childhood directly. Agents would want to sign a contract with their parents in which they choose the optimal level of investments and agree to pay it back once
they are adults, but such a market does not exist. Government investments towards children reduce the problems generated by this missing market. Introducing a non-targeted government investments towards early childhood (e.g., mandatory pre-K schools) leads to reductions in income inequality of 20%, intergenerational mobility improvements of 60%, and welfare gains (in consumption equivalence terms) of 24%—larger than using the same resources for a government transfer.
References


A Empirical Findings: Details

A.1 Child Development Supplement: To be updated

The results presented in Section 3 are for the whole sample of children born to at least high-school educated mothers for which we have data on their time diaries. The summary statistics for these children are presented in Table A1.

Table A1: CDS Summary Statistics: Whole Sample.

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

Table shows summary statistics for sample used to study parental investments in Section 3.

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

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

The summary statistics of this sample are shown in Table A2. It is seen that this sample is rather high-income and highly educated relative to a sample that includes families with more children and are not married. All fathers work for at least some period when children are young. Interestingly, while early on all fathers are working, this share goes down as the share of mothers working goes up. Even though this sample is more representative of a stable high socioeconomic group, it is shown in Section 3 that differences on parental investments are still sizable. Differences in parental investments including single parent households would also be interesting, but are out of the scope of the model used here.

Table A2: CDS Summary Statistics: Selected Sample.

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

*Table shows summary statistics for (most selective) sample used to study parental investments in Section 3.*
A.2 Income Profile: PSID Sample Selection

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

- High-school dropouts: 3,649 Individuals (40% from SRC), with a total number of observations of 32,495.
- High-school graduates: 13,091 Individuals (53% from SRC), with a total number of observations of 104,595.
- College graduates: 5,001 Individuals (75% from SRC), with a total number of observations of 44,704.

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

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

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

- High-school dropouts: 554 Individuals, with a total number of observations of 2,350.
- High-school graduates: 3,574 Individuals, with a total number of observations of 16,960.
- College graduates: 1,479 Individuals, with a total number of observations of 7,552.
### A.4 Additional Figures and Tables

Table A3: Age profile of wages by education group.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High School</td>
<td>College non-STEM</td>
<td>College STEM</td>
</tr>
<tr>
<td>Age</td>
<td>0.0320***</td>
<td>0.0782***</td>
<td>0.0592***</td>
</tr>
<tr>
<td></td>
<td>(0.00282)</td>
<td>(0.00531)</td>
<td>(0.00724)</td>
</tr>
<tr>
<td>Age²</td>
<td>-0.000321***</td>
<td>-0.000823***</td>
<td>-0.000588***</td>
</tr>
<tr>
<td></td>
<td>(3.43e-05)</td>
<td>(6.38e-05)</td>
<td>(8.75e-05)</td>
</tr>
<tr>
<td>Over Sample</td>
<td>-0.0360***</td>
<td>-0.0319</td>
<td>-0.0502</td>
</tr>
<tr>
<td></td>
<td>(0.00996)</td>
<td>(0.0231)</td>
<td>(0.0322)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.270***</td>
<td>1.620***</td>
<td>2.127***</td>
</tr>
<tr>
<td></td>
<td>(0.0595)</td>
<td>(0.112)</td>
<td>(0.152)</td>
</tr>
<tr>
<td>Observations</td>
<td>17,206</td>
<td>6,344</td>
<td>2,543</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.077</td>
<td>0.136</td>
<td>0.123</td>
</tr>
<tr>
<td># of households</td>
<td>3057</td>
<td>1042</td>
<td>395</td>
</tr>
</tbody>
</table>

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 4. 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 $\{d^e(s_j)\}$, consumption, labor supply, and assets holdings $\{c_j(s_j), l_j(s_j), a'_j(s_j)\}$, parental time and money investments $\{d^p_j(s_j), d^m_j(s_j)\}$, 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 $\{K, L_1, L_2\}$; (iv) prices $\{r, w^1, w^2\}$; (v) tax policy $\{\tau_c, \tau_y, \tau_k, \omega\}$; and (vi) a vector of measures $\mu$ 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:

\[ L_1 = \sum_{j=J_e}^{J_e} \int_{S_j} \varepsilon_j^1 (\theta) l(1, s_j) d\mu_j + \sum_{e=2}^{3} \sum_{j=J_e}^{J_e} \int_{S_j} \varepsilon_j^e (\theta) l(e, s_j) d\mu_j \]

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

For college level:

\[ L_2 = \sum_{j=J_e+1}^{J_e} \int_{S_j} \varepsilon_j^2 (\theta) l(2, s_j) d\mu_j. \]

4. Asset market clears

\[ K = \sum_{j=J_e}^{J_d} \int_{S_j} a_j(s_j) d\mu_j. \]

5. Good market clears:

\[ \sum_{j=J_e}^{J_d} \int_{S_j} c_j(s_j) d\mu_j + \delta K + \sum_{j=J_e}^{J_d} \int_{S_j} p_{e1} \{ d^e(s_{j_e}) \geq 2 \} + \sum_{j=J_d}^{J_f+1} \int_{S_j} d_j^{m} (s_j) \mu_j = F(K, H) \]

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

6. Government budget holds with equality

\[ \sum_{j=J_e}^{J_d} \int_{S_j} \pi (\theta, e) d\mu_j + G = \sum_{j=J_e}^{J_f} \int_{S_j} T(y(s_j), k(s_j), c(s_j)) d\mu_j. \]

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

7. Individual and aggregate behaviors are consistent: measures \( \mu \) 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 and Results: Details

C.1 Replacement benefits: US Social Security System

The pension replacement rate is obtained from the Old Age Insurance of the US Social Security System. We use education level as well as the skill level to estimate the average lifetime income, on which the replacement benefit is based. We estimate the average life time income to be \( \hat{y}_j(\theta, e) = w^e \varepsilon^e_j(\theta, \bar{\eta}) \times \bar{h} \) with \( \bar{\eta} \) and \( \bar{h} \) referring to the average efficiency and hours worked. Then averaging over \( j \), mean income \( \hat{y} \) is calculated and used in (7) to obtain the replacement benefits.

The pension formula is given by
\[
\pi(\theta, e) = \begin{cases} 
0.9\hat{y}(\theta, e) & \text{if } \hat{y}(\theta, e) \leq 0.3\bar{y} \\
0.9(0.3\bar{y}) + 0.32(\hat{y}(\theta, e) - 0.3\bar{y}) & \text{if } 0.3\bar{y} \leq \hat{y}(\theta, e) \leq 2\bar{y} \\
0.9(0.3\bar{y}) + 0.32(2 - 0.3)\bar{y} + 0.15(\hat{y}(\theta, e) - 2\bar{y}) & \text{if } 2\bar{y} \leq \hat{y}(\theta, e) \leq 4.1\bar{y} \\
0.9(0.3\bar{y}) + 0.32(2 - 0.3)\bar{y} + 0.15(4.1 - 2)\bar{y} & \text{if } 4.1\bar{y} \leq \hat{y}(\theta, e) 
\end{cases} 
\]  

(7)

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

C.2 Aggregate Production Function

Wages from PSID

In order to estimate the APF we need to first estimate the wage for each year and education group. For this, we return to the PSID data and remove the 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 $\omega$ 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:

$$Y = K^\alpha H^{1-\alpha}$$

$$H = [sL_1^\rho + (1-s)L_2^\rho]^{1/\rho}$$

We can then estimate the parameters $s$ and $\rho$ using the following equation:

$$\log \left( \frac{\omega_2^t}{\omega_1^t} \right) = \log \left( \frac{1-s}{s} \right) + \rho \log \left( \frac{L_2^t}{L_1^t} \right)$$

We can estimate this using OLS or First Differences. Moreover, we also do IV using lags as instruments. This approach leads to estimates around 0.3 for $\rho$, and 0.5 for $s$—in line with the estimates from Katz and Murphy (1992) and Heckman et al. (1998).
D Other Policies

D.1 Parenting Education

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

$$
\theta'_k = \left[ \alpha_{j,\theta_k} \theta_k^{\phi_j} + \alpha_{j,\theta} \max \{ \theta, \theta_{PE} \}^{\phi_j} + \alpha_{j,I} \theta_k^{\phi_j} \right]^{1/\phi_j} e^\nu, \quad \nu \sim N(0, \sigma_{j,\nu})
$$

(8)

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 $\theta_{PE}$ entered as a perfect substitute for $\theta$) we would expect general welfare benefits to be even larger. Therefore, our results here may be thought as a lower bound in that respect.\textsuperscript{22}

Estimating the cost of and returns to parenting education (in terms of $\theta_{PE}$) is not easy, so we take the following approach in order to estimate a lower bound on the benefits of such policy. We would like to estimate these from parenting education programs in the US but, to the best of our knowledge, this data is not available. In general, parenting education programs have been more popular in research studies from developing countries so we use that evidence instead. Moreover, even though we were not able to find evidence of costs and long-term impacts from the same study, we used evidence of two programs with similar curricula. We estimate the cost of running such policy in the US based on the upper-bound available for Colombia (Attanasio et al., 2016). In particular, we assume that the cost

\textsuperscript{22}An alternative compelling interpretation of the effect of parenting education is to increase $\bar{A}$. Effectively, both alternatives increase the derivative of $\theta'_k$ to $\tau$ or $m$, i.e., the productivity of investments. The key element is how to benchmark the increase of productivity for either alternative. We present here the first approach since our benchmarking is made on the increased income of children from low-income parents. If we were to focus on increases in $\bar{A}$ we would get higher returns for high-income individuals than if we follow our selected approach. This is in line with our objective of estimating a lower bound on the potential impact of parenting education.

53
of running such a program is $11,400 per family in the first period with children.\textsuperscript{23} We also need to estimate the effectiveness of parenting education, i.e., $\theta_{PE}$. 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. (2013). Parents of growth-stunted children were randomly selected to participate in the program when their children were between 0 and 2. Once children were approximately 22 years old, Gertler et al. (2013) estimate that parenting education program led to 25% increase in the children’s income. As shown in Table D4, we choose $\theta_{PE}$ such that if a small share of poor families in our estimated economy were introduced to the parenting education program their children’s income would increase by 25% as well. This is obtained by $\theta_{PE}$ that is 0.2 standard deviations above the average $\theta$. We will refer to this value as the benchmark return to parenting education.

<table>
<thead>
<tr>
<th>$\theta_{PE}$ relative to Avg. $\theta$</th>
<th>Policy benchmark</th>
<th>Change from Baseline (%)</th>
<th>Income Bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.0 SD</td>
<td>-1.2 SD</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>-0.8 SD</td>
<td>-1.0 SD</td>
<td>3.46</td>
<td></td>
</tr>
<tr>
<td>-0.6 SD</td>
<td>-0.8 SD</td>
<td>6.78</td>
<td></td>
</tr>
<tr>
<td>-0.4 SD</td>
<td>-0.6 SD</td>
<td>10.14</td>
<td></td>
</tr>
<tr>
<td>-0.2 SD</td>
<td>-0.4 SD</td>
<td>13.65</td>
<td></td>
</tr>
<tr>
<td>+0.0 SD</td>
<td>-0.2 SD</td>
<td>18.86</td>
<td></td>
</tr>
<tr>
<td>+0.1 SD</td>
<td>-0.1 SD</td>
<td>21.42</td>
<td></td>
</tr>
<tr>
<td>+0.2 SD</td>
<td>0.0 SD</td>
<td>23.69</td>
<td></td>
</tr>
<tr>
<td>+0.3 SD</td>
<td>+0.1 SD</td>
<td>26.06</td>
<td></td>
</tr>
</tbody>
</table>

We now evaluate parenting education in three steps. The first two steps estimate what the return would be a government run program that introduced parenting education. Here the government enrolls (and pays for) every agent to obtain parenting education, independently of whether it is ineffective for them or not—i.e., the government cannot observe or use the skills of the agents to determine their enrollment. Our third step looks at whether such a program would need to be government enforced. We

\textsuperscript{23}Running a similar policy in Colombia has been estimated to cost between $450 and $750 per child (Attanasio et al., 2016). This program actively used a group of women ("Madres Lideres") with average education equivalent to a high-school degree. If we assume running the program in the us would use similar inputs we can try to estimate the costs, we can compare the salaries of similar individuals in the US and Colombia to estimate the cost in the US. In order to estimate an upper bound to such cost, we assume here that in the US they would employ college-educated women instead. In Colombia, $450 represented approximately the average monthly salary of a high-school educated person in Colombia. Since the average salary of a college graduate in the US in the 2000s was approximately $42,000, this would imply that in the US the cost of running a similar program would be between $3,400 and $5,700 per child. Disregarding potential returns to scale of running the program for two children per family, this would imply a cost of up to $11,400 per family in our model.
introduce the program as something that agents can purchase by themselves once children are born and study its effects. For most of the analysis we focus on measures of income inequality, intergenerational mobility, college-graduation rate, and average income. Finding policies that manage to improve these aggregate outcomes is of general interest. As a measure of general welfare, we also report on consumption equivalence.

D.1.1 Parenting Education as a Government Program

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

The top panel of Table D5 shows the results in the short-run partial-equilibrium case. Effects are reported as percent changes from the baseline economy. Each row shows the outcome changes for different levels of $\theta_{PE}$. For example, -0.1SD means that the effectiveness of the program is 0.1 standard deviations (of $\theta$) less 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., at or close to 1 standard deviation below), parenting education seems to provide large welfare increase. 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 15% increase in tax revenue is approximately equal to a $750 increase in the tax revenue per household each year. Parenting education, as long as the policy is at least 0.1 SD less effective than the benchmark, is able to increase the resources available for the government in the short run.
We now evaluate the parenting education program in a long-run general-equilibrium environment: it is introduced permanently and we consider the economy in the new steady state. We adjust government (labor) taxes so that its budget does not change. We evaluate what the long-run effects would be—taking into account the interactions between taxation, education, and parental investments towards children. The model now provides evidence that is harder to obtain empirically. The bottom panel of Table D5 shows the results.

Parenting education remains highly beneficial. Once again, even for the cases in which the policy is 1 standard deviation less effective than the estimated benchmark, we find the consumption equivalence measure for welfare to be larger than zero. The effect on intergenerational mobility is almost equivalent to the partial equilibrium case. If parenting skills can be improved as much as the literature suggests, it would lead to an improvement in the intergenerational mobility log-income persistence coefficient of 0.14.
points. Similarly, the effect on average income remains almost as strong, with the reduction being driven by the wages adjustment as well as the increased taxes. The effect on college graduation is smaller than in partial equilibrium. Parenting education also proves to be a policy that would increase tax revenue in the long run. By increasing the share of high-skilled, the growth in income, consumption, and savings even allows the government to reduce labor tax rates.

### D.1.2 Parenting Education Market

Next we look at whether such a program would need to be government enforced. We introduce the program as something that agents can purchase by themselves once children are born and study its effects. In other words, when their children is born they have one more choice to make: acquire parenting education at the price of $13,500 (same as the government estimate) or now. Table D6 shows the results for such exercise, for different levels of productivities $\theta_{PE}$ relative to the estimated benchmark. At the benchmark the consumption equivalence is 9.75%, which is associated with a 100% take-up among the low-skilled parents. The largest effect is on intergenerational mobility which would increase by almost 40%. Similarly, average income would increase by 7% while inequality would be reduced by almost 7%. Relative to the benchmark, the productivity of the program would need to be below 1 SD for the take-up to be very small and aggregate effects to be minimal. For lower values, no agent acquires parenting education and hence the economy does not change relative to the initial steady state.

<table>
<thead>
<tr>
<th>$\theta_{PE}$</th>
<th>Cons. Equiv.</th>
<th>Bottom Income</th>
<th>Avg. Income</th>
<th>Inequality</th>
<th>Mobility</th>
<th>College</th>
<th>Tax Rate</th>
<th>Take-Up Low</th>
<th>Take-Up Medium</th>
<th>Take-Up High</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.2 SD</td>
<td>-0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>-1.0 SD</td>
<td>0.09</td>
<td>0.09</td>
<td>0.26</td>
<td>-0.05</td>
<td>0.65</td>
<td>0.49</td>
<td>-0.01</td>
<td>2.56</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>-0.8 SD</td>
<td>1.68</td>
<td>5.92</td>
<td>1.95</td>
<td>-1.81</td>
<td>12.46</td>
<td>2.35</td>
<td>-0.24</td>
<td>29.99</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>-0.6 SD</td>
<td>5.47</td>
<td>13.08</td>
<td>3.86</td>
<td>-5.00</td>
<td>24.87</td>
<td>7.48</td>
<td>-1.01</td>
<td>77.40</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>-0.4 SD</td>
<td>6.14</td>
<td>15.50</td>
<td>4.12</td>
<td>-5.71</td>
<td>29.92</td>
<td>8.50</td>
<td>-1.11</td>
<td>77.67</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>-0.2 SD</td>
<td>6.90</td>
<td>18.25</td>
<td>4.64</td>
<td>-6.22</td>
<td>35.02</td>
<td>9.09</td>
<td>-1.22</td>
<td>81.16</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>-0.1 SD</td>
<td>7.48</td>
<td>19.22</td>
<td>4.88</td>
<td>-6.51</td>
<td>36.61</td>
<td>9.78</td>
<td>-1.35</td>
<td>88.94</td>
<td>0.63</td>
<td>0.00</td>
</tr>
<tr>
<td>Benchmark</td>
<td>9.75</td>
<td>21.07</td>
<td>7.43</td>
<td>-8.12</td>
<td>39.97</td>
<td>12.33</td>
<td>-1.78</td>
<td>100.00</td>
<td>36.51</td>
<td>0.00</td>
</tr>
<tr>
<td>0.1 SD</td>
<td>11.78</td>
<td>22.82</td>
<td>8.86</td>
<td>-9.39</td>
<td>43.69</td>
<td>15.05</td>
<td>-2.15</td>
<td>100.00</td>
<td>45.33</td>
<td>0.00</td>
</tr>
</tbody>
</table>

There are differences between the case in which the program is government run (and every household
is paid to go through it, even if it provides no benefits) and when parenting education is market good that households decide to acquire. On the one hand, the government program is wasteful in the sense that it pays for people who do not benefit from the program. On the other hand, the government gets larger take-up. For low values of $\theta_{PE}$ the larger take-up is important for low-skilled individuals. Once $\theta_{PE}$ is close to its benchmark value it also starts being beneficial for mid-skilled agents. Around the benchmark $\theta_{PE}$ this two forces are close to balanced as seeing by comparing the welfare outcomes between the parenting education program (in general equilibrium) and the parenting education market. This suggests that, if parenting education is as effective as the evidence implies, it does seem that the program should be provided by the government.

### D.2 Endogenous Parental Investments and the Welfare Evaluation of Tax Progressivity

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

The model introduced here is useful to answer this question as it adds endogenous childhood development and parental transfers to a standard life-cycle macroeconomic model. In order to evaluate the importance of endogenous intergenerational links, we compare the effects in our model with endogenous links to the effects obtained by the same model but with exogenously fixed links. The model with exogenous childhood development is equivalent to the original model, but where the intergenerational transition matrix of skills is fixed to be equal to one obtained (endogenously) in the original steady state. This matrix defines that the distribution of children’s skills depends (exclusively) on parents’ skills and education group. Hence, the tax system cannot affect the development of skills directly, though it may affect it through education choices. The model with exogenous childhood development is re-estimated to match the same set of moments (excluding the childhood development related ones) from the full
model. The estimated parameters and moments are shown in Table D7.

Table D7: Estimation: exogenous childhood development model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
<th>Moment</th>
<th>Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preferences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>b</em></td>
<td>0.26</td>
<td>Altruism</td>
<td>Parent-to-Child Transfer (k)</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td><em>μ</em></td>
<td>356</td>
<td>Mean Labor Disutility</td>
<td>Avg. hours worked</td>
<td>65</td>
<td>64</td>
</tr>
<tr>
<td><strong>School Taste:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>α</em></td>
<td>8.6</td>
<td>Avg. taste for College</td>
<td>College share</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td><em>αθ</em></td>
<td>-1.9</td>
<td>College taste-Skills relation</td>
<td>College-Skills slope</td>
<td>0.30</td>
<td>0.28</td>
</tr>
<tr>
<td><em>σφ</em></td>
<td>2.7</td>
<td>SD of College taste shock</td>
<td>College: residual variance</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td><em>δφ</em></td>
<td>-2.9</td>
<td>Draw of school taste: mean by parent’s education</td>
<td>IGE persistence of education</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>Interest rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>ι</em> (×10^2)</td>
<td>2.6</td>
<td>Borrow-save wedge</td>
<td>Share of borrowers</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td><strong>Government</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>ω</em> (×10^2)</td>
<td>2.4</td>
<td>Tax rebate</td>
<td>Income Var Ratio: Disposable to Pre-Gov</td>
<td>0.70</td>
<td>0.68</td>
</tr>
</tbody>
</table>

For both models, we evaluate modifying the labor income tax *τ_y* and adjusting the lump-sum transfer *ω* such that the government’s budget remains constant. For the welfare analysis we focus on consumption equivalence under the veil of ignorance as defined in Section 6. Figure D1 shows the results. The model with exogenous childhood skills predicts small welfare increases from increasing tax progressivity, in line with the literature that uses this kind of models to evaluate tax progressivity (e.g., Floden and Lindé, 2001; Conesa and Krueger, 2006; Heathcote et al., 2014). On the other hand, the full model developed here predicts large welfare gains from substantially reducing tax progressivity.
Figure D1: (Constrained) Optimal Tax Progressivity

- Consumption Equivalents (% - All)
  - Exogenous
  - Endogenous
  - w/o Changing Distribution

- Consumption Equivalents (% - Low SES)

- Consumption Equivalents (% - High SES)

- Parental transfer ($1,000)

- Share Change (% - Low SES)

- Share Change (% - High SES)

- Hours with child

- Labor Tax Rate (%)

- Rebate ($1,000)

- Redistribution
Figure D2: Transition Dynamics: Tax Rate – Sudden Adjustment
Figure D3: Transition Dynamics: Tax Rate – Delayed-Fast Adjustment
Figure D4: Transition Dynamics: Tax Rate – Delayed-Slow Adjustment