On the Intergenerational Transmission of Economic Status

Sang Yoon (Tim) Lee
Toulouse School of Economics

Ananth Seshadri
University of Wisconsin-Madison*

December 2016

Abstract

We present a model in which human capital investments occur over the life-cycle and across generations, à la Becker and Tomes (1986), also featuring incomplete markets and government transfer programs. The human capital technology features multiple stages of investment during childhood, a college decision, and on-the-job accumulation. The model can jointly explain a wide range of intergenerational relationships, such as the intergenerational elasticities (IGE) of lifetime earnings, college attainment and wealth, while remaining empirically consistent with cross-sectional inequality. Much of life-cycle inequality is determined early in life, which in turn is explained in large part by parental background. The model implies that this is mainly due to early investments in children made by young parents, so life-cycle constraints these parents face are important for understanding the persistence of economic status across generations. Education subsidies, especially early on, can significantly reduce the intergenerational persistence of economic status.

*Email: sylee.tim@tse-fr.eu and aseshadr@ssc.wisc.edu, respectively. We are very grateful for constructive comments by Gary Becker, Steven Durlauf, Arthur Goldberger, John Knowles, Nicolas Schutz, Chris Taber, seminar participants at the 2012 MIE conference, the 2012 Vienna Macro Workshop, the University of Mannheim, and the University of Chicago. All remaining errors are our own.

1. Introduction

Intergenerational persistence is high in the United States. Estimates for the intergenerational elasticity or correlation (IGE/IGC) of earnings is as high as 0.4, and wealth, consumption, schooling and poverty are also persistent across generations.[1] Understanding this degree of persistence, along with disentangling the respective contributions of the transmission of innate abilities, family background, or economic policy in generating persistence, has been the subject of much discussion and debate. However, most such analyses are statistical models, and economic models as of yet have come short of rationalizing various patterns we see in the data. The goal of this paper is to provide such a framework to better understand these complex empirical relationships, and to analyze the impact of various policies.

Theories of intergenerational persistence gained traction beginning with the seminal work of Becker and Tomes (1979, 1986) (henceforth BT).[2] They laid out a simple two-period equilibrium model to derive implications for the intergenerational transmission of lifetime earnings and wealth.[3] But while the BT framework is used widely in the literature, it has been met with some empirical skepticism, most notably by Goldberger (1989). He argued that an economic approach added little value relative to mechanical approaches, i.e. those that do not rely on household optimizing behavior. Mulligan (1999) showed that the inability of a parent to borrow against the future income of his child (hereafter "intergenerational borrowing constraint"), a key feature of Becker and Tomes (1986), seems not to matter, and even if it does, is empirically irrelevant. Han and Mulligan (2001) further argue that heterogeneous abilities and intergenerational borrowing constraints are indistinguishable.

One reason for this apparent invalidation of the BT model may be because entire life-

[1]There is a large empirical literature that attempts to measure the magnitude of the IGE/IGC of earnings and/or income, and the number we cite is in the range of by now widely accepted value found by Chetty et al. (2014) using social security records, which in turn is similar to earlier estimates Solon (1992) using the PSID. Throughout the text, we will refer to the intergenerational persistence of earnings as simply "the IGE" unless clarification is needed.

[2]Loury (1981) was a similar model in a dynastic setting with borrowing constraints.

[3]There are several other important papers in the theoretical literature that focus on intergenerational persistence. Benabou (1993) and Durlauf (1996) present models of segregation, Galor and Zeira (1993) focus on poverty traps while Banerjee and Newman (1993) present a model featuring mobility traps.

times are condensed into two periods, each which comprises as many as 20-30 years. Parents make a once-and-for-all investment in children who grow up to earn a once-and-for-all income. Human and physical capital investments (intergenerational financial transfers) are decided upon simultaneously. The only moments predicted by the model are total investments in children's education, lifetime earnings and total financial transfers, which are averaged over extensive time periods. All decisions that are made over multiple periods are ignored, which could affect intergenerational transfer decisions that happen much later in life, such as bequests. There is no consideration for less than perfectly substitutable child investments across periods.[4] Furthermore, data spanning the entire lifetimes of multiple generations are scant at best, making it difficult to validate the key mechanisms of the model.

We explicitly incorporate multi-period decisions into the BT model. Adults accumulate human capital until retirement according to a Ben-Porath (1967) technology, as in Heckman et al. (1998); Huggett et al. (2011) and others. When young, they educate their children over multiple periods according to a technology that features dynamic complementarity, and also face life-cycle borrowing constraints (Heckman and Mosso, 2014). Children grow up and go to college, and have their own children as they continue to accumulate their own human capital in adulthood. Financial transfers from one generation to the next occur only *after* children become fully grown adults. These families are cast in an overlapping generations framework in which learning abilities are imperfectly transmitted over generations, with infinitely-lived dynasties who are altruistic.[5]

We require this model to be consistent with intergenerational moments of earnings, education and wealth, and also cross-sectional earnings inequality over the life-cycle. This is possible since all individuals live through multiple periods of life, unlike in the BT model. In particular, we rationalize the joint distribution of earnings and assets, since a framework that can account for earnings persistence but inconsistent with wealth in-

[4] This criticism also forms part of the bases for studies such as Cunha and Heckman (2007); Cunha et al. (2010); Caucutt and Lochner (2012) who argue that there is strong complementarity across periods.

[5] BT as well as many of the ensuing papers used two period models in which parents care about their own consumption and the income of their children. A dynastic formulation in which parents care directly about descendants' utility, as we do, allows a parsimonious representation of preferences, although is more costly numerically.

equality and its transmission would be unconvincing. While we are certainly not the first to present a quantitative model, many prior formulations were too simplistic to directly compare with available data. Some either assume away important model elements emphasized by BT (e.g. Restuccia and Urrutia (2004) abstract away from asset accumulation) while others ignore the empirical skepticism raised against it.

We find that parental background is a strong predictor of children's lifetime outcomes, and more so for wealth than earnings. The inclusion of life-cycle borrowing constraints which prevent young parents from investing in their children early on, before they transfer any of their wealth, is crucial for this result. Due to complementarity, suboptimal investments early in life cannot be corrected later in life (Heckman and Mosso, 2014). Hence the intergenerational borrowing constraint, faced by parents much later in life, does not matter much for young parents' investments in children's human capital. In fact in our model, parents who are unable to invest enough in their children instead invest in their own human capital, making it more possible for parents who *did not* achieve optimal investment in their children's human capital to make larger financial transfers.

Conversely, the intergenerational component of the model also helps explain life-cycle inequality. Castañeda et al. (2003) demonstrate that accounting for intergenerational relationships is crucial to account for cross-sectional inequality, but takes both the life-cycle earnings process and intergenerational relationship as exogenous. Huggett et al. (2011) estimates the joint distribution of human capital and assets at age 20 to match life-cycle earnings dynamics and distributions later in life. They find that small differences in initial conditions can lead to large differences in earnings and wealth over the life-cycle. In contrast, we take neither the earnings process nor initial conditions as given, and require the intergenerational mechanism of our model to endogenously generate empirically valid distributions of earnings and wealth within and across generations. That is, the life-cycle component of our model helps explain intergenerational data, while intergenerational investments help explain cross-sectional data.

Consistently with previous studies, our model predicts that most of life-cycle inequality is predetermined upon labor market entry (73-74%). What is new in our paper is that we can also quantify how much of this can be explained by parental background.

Indeed, we find that parents' states when children are young can explain about a quarter of children's life-cycle earnings variance later in life, and half of their lifetime wealth variance. This is because parents' states have a large explanatory power over children's initial conditions at age 24. Going further, we quantify that grandparents' states also have a significant impact on the life-cycle earnings and lifetime wealth inequality in the grandchildren's generation: about a fifth and a third of their variances, respectively.

Consequently, our model admits an IGC of lifetime earnings of 0.34, as estimated by Chetty et al. (2014), from an IGC of learning abilities which we calibrate to be 0.23. The borrowing constraints faced by young parents who need to invest in their children's human capital amplify ability persistence, accounting for about a third of earnings persistence. What distinguishes our model from previous BT-type models is that our mechanism is inherently non-linear, and we explicitly differentiate ability persistence from childhood investments that happen in their first stage of life.[6]

To the best of our knowledge, ours is the first attempt at incorporating an endogenous human capital accumulation process into an intergenerational setting in which pre-labor market initial conditions are determined by multiple stages of childhood investments. Combining childhood, life-cycle, and intergenerational elements is not trivial, especially because the child's state variables affect the parent's decision problem. Adding to the complexity is the asset and labor market equilibrium for high school and college and non-college workers. Nonetheless, it is precisely these two features (multiple time periods and equilibrium) that make comparison with empirical counterparts feasible. In the data, typically, we observe parent and child variables at different ages, which we can compare to the end and beginning of life-cycles in our model. The stationary equilibrium setup allows us to discipline parent's behavior early in life, or children's behavior later in life, which is also not simultaneously observed in the data.

These features are well understood in isolation, and parameters we recover are in the range of previous estimates. Following Del Boca et al. (2014), who show that parental time is an important input in the production of a child's human capital, our childhood

[6] "Genes and environment cannot be meaningfully parsed by traditional linear models that assign unique variances to each component." (Heckman, 2008)

human capital production function also features both time and good investments. We estimate this production function using time-use, education expenses, and test score data from the Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID).[7] While our formulation is simpler than theirs, we directly estimate the degree of complementarity over the multiple stages of investment rather than assuming it, following Cunha et al. (2010) and subsequent papers who have shown that the degree of complementarity between early and later child investments is important. Nonetheless, we find that the elasticity of substitution between investments made in different stages of childhood are close to unity, as assumed in Del Boca et al. (2014).

After childhood but before entering the labor market, we assume households decide whether or not children enroll in college, as much of inequality can be attributed to differences in educational attainment. The college choice also allows us to separate cross-sectional inequality into differential skill prices and different levels of skill, a focus of many recent studies. Consistent with Carneiro and Heckman (2002); Carneiro et al. (2011), college is mostly selection and explains little of life-cycle inequality. The Ben-Porath function for post-schooling life-cycle wage growth is calibrated to life-cycle earnings moments from the Panel Study of Income Dynamics (PSID), and its estimated parameter is in the range of estimates in Browning et al. (1999).

In addition, we model borrowing constraints and most of the conventional forms of government intervention. The data we observe comes from an economy in which the government taxes income progressively to subsidize education, fund social security and assist low income households through welfare payments. Thus, while our main objective is to rationalize intergenerational persistence and life-cycle inequality simultaneously, it is also suitable for counterfactual policy analyses. Since childhood investments and borrowing constraints matter, alternative capital markets and government policies may also affect intergenerational persistence and inequality.

As expected, we find that relaxing the intergenerational borrowing constraint has only small effects. Relaxing the life-cycle constraint has a large impact on intergenerational

[7] As a caveat, we should emphasize that we do not use test scores to measure learning ability, which determines the slope of human capital growth, but rather the level of human capital among children.

persistence, reducing the IGC from 0.34 to 0.24, while also reducing inequality. When both constraints are relaxed simultaneously, the drop in the IGC is smaller but there is a further reduction in inequality and also a rise in average earnings.

Among the policies we consider, we find that shifting all education subsidies to the earliest period, when children are ages 0-5, has the largest effect on long-run intergenerational persistence, reducing the IGC to as low as 0.1. This is because under dynamic complementarity, small differences early on generate large differences later in life, and young parents are the most likely to be borrowing constrained. Furthermore, such a policy also raises average earnings by increasing the average human capital level of the entire economy. While a serious policy recommendation will no doubt require richer micro-data, our results give weight to recent discussions that call for more interventions early in life as a remedy to intergenerational persistence, while also raising the overall productivity of the economy (Heckman and Mosso, 2014).

The rest of the paper is organized as follows. Section 2 lays out our model, and Section 3 describes the data we use to discipline our model parameters, in which we also estimate the childhood human capital production function. Section 4 explains the calibration strategy. Sections 5-6 show our main results on sources of inequality and how our model differs from BT. This includes the relevance of our model for understanding parents' influence on children's initial conditions in adulthood, and conversely, how investment in children affects the cross-sectional distribution of earnings and wealth in the long-run. In Section 7 we conduct counterfactuals in which we vary the tightness of borrowing constraints, tax progressivity, and education subsidies. Section 8 concludes.

2. Model

A period is 6 years, and time is discrete and runs forever. At any given point in time, there are 13 overlapping generations, each with a unit mass of individuals (so the demographic structure is uniform). Each individual goes through 13 stages in life, and we let period j represent the stage of the life-cycle between ages $[6j, 6j+5]$:

1. Childhood, $j = 0, 1, 2$: Children make no decisions and are raised/educated by the

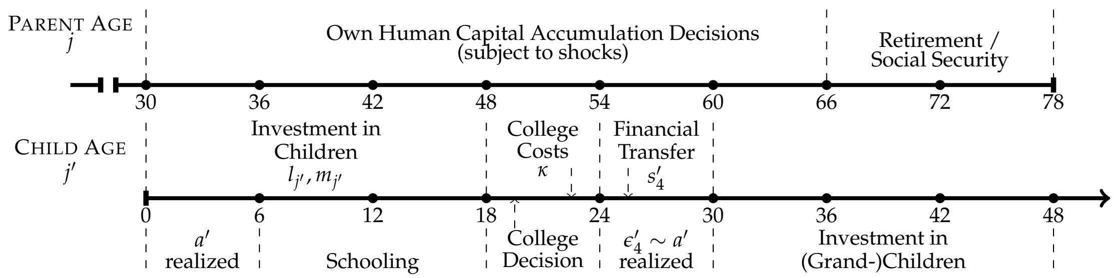


Figure 1: Sequence of Events

parent they are attached to.

2. College, $j = 3$: The child remains attached to the parent, and the parent-child pair makes a joint decision on whether the child should go to college, and how much the child should work. If the child goes to college, the amount of time he can work is reduced by 4 years.

3. Independence, $j = 4$: A child becomes an independent adult, and accumulates human capital on-the-job.

4. Young parent, $j = 5, 6, 7, 8$: Individuals bear children at $j = 5$ while continuing to accumulate human capital. At $j = 5, 6, 7$, they make time and goods investments in their children's human capital. Joint college decision for child at $j = 8$.

5. Old parent, $j = 9, 10$: Continues to accumulate own human capital, makes a financial transfer to child at $j = 9$ and saves for retirement which occurs after $j = 10$.

6. Retirement, $j = 11, 12$: Retired individuals consume their savings and social security benefits, and die after $j = 12$.

The sequence of events is depicted in Figure 1. For the remainder of the text, we will denote all child variables with primes. Since a generation is 30 years, the child of a parent who is in period j is always in period $j' = j - 5$. In periods $j = 5, 6, 7, 8$, we assume that the parent-child pair solves a Pareto problem to maximize period utility:

$$U(C_j) = \max \left\{ u(c_j) + \theta u(c'_{j-5}) \right\}$$

where C_j is the total consumption of the parent-child pair and θ the (dynastic) altruism factor. Assuming a CRRA coefficient χ, since consumption is shared according to a Pareto rule, we can write

$$U(C_j) = qu(C_j), \quad q = \left(1 + \theta^{\frac{1}{\chi}}\right)^{\chi}$$

where q is then interpreted as an adult consumption-equivalent scale.

The essence of our paper is to postulate different human capital production technologies for childhood and adulthood, in a model in which *all* individuals live through both phases. We first describe these production technologies, followed by a detailed description of decisions made at each stage of the life-cycle.

2.1 Adulthood human capital accumulation

From college ($j = 3$) onward, we assume that human capital h evolves as Ben-Porath:

$$h_{j+1} = \epsilon_{j+1} \left[a(n_j h_j)^b + h_j \right], \tag{1}$$

where a is an individual's learning ability determined at birth, and ϵ_j is a market luck shock drawn in period j.[8] It assumes the same exponent, b, for both the time spent accumulating human capital $n_j \in [0, 1]$ and human capital h_j.[9] Shocks to the growth rate of human capital are manifested as permanent earnings shocks. We assume these market luck shocks are i.i.d. starting from a young adult's first period of independence:[10]

$$\log \epsilon_j \underset{i.i.d.}{\sim} \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2) \equiv F(\epsilon), \quad j \in \{4, \ldots, 10\} \tag{2}$$

where μ_a and σ_a are the population mean and standard deviation of abilities.

[8] This specification was used in Huggett et al. (2011), making our results easily comparable to theirs.

[9] Heckman (1976); Heckman et al. (1998); Browning et al. (1999) all find evidence in favor of assuming a single exponent for n and h. These papers also argue that estimates of intertemporal substitution in labor supply over annual or longer horizons are small, so abstracting from a labor-leisure choice will have little impact in our model in which a period is 6 years. In a calibration in which we let b differ between high school and college individuals, they were calibrated to be virtually equal.

[10] We have tried a version in which the initial shock is correlated with own ability or parental states, with zero to little impact on our quantitative results (the correlation was calibrated to be close to zero).

2.2 Childhood skill formation

At $j = 5$, individuals bear children whose abilities a' are drawn according to $a' \sim G(a'|a)$, which we model as an AR(1) process

$$\log a' = (1 - \rho_a)(\mu_a - \sigma_a^2/2) + \rho_a \log a + \eta', \quad \eta \sim \mathcal{N}(0, (1 - \rho_a^2)\sigma_a^2), \tag{3}$$

where η' is an intergenerational shock. An individual's ability remains constant throughout his lifetime, and since it exogenously depends on his parent's ability, will capture the part of intergenerational persistence not explained by economic behavior.

In addition to ability transmission, we include three important aspects on children's skill formation that are well appreciated in the literature: i) parental time and good investments, ii) complementarity between inputs, and iii) multiple stages of investments. Specifically, we assume that the amount of human capital a child attains at the beginning of $j = 3$ (college) is determined by

$$h_3' = \zeta \left\{ \omega_2 X_2^{\phi_2} + (1 - \omega_2) \left[\omega_1 X_1^{\phi_1} + (1 - \omega_1) \left(\omega_0 X_0^{\phi_0} \right)^{\phi_1} \right]^{\frac{\phi_2}{\phi_1}} \right\}^{\frac{1}{\phi_2}} \tag{4}$$

where ϕ_0 captures the returns to initial investments, and $(\omega_{j'}, \phi_{j'})$, $j' \in \{1, 2\}$, capture the relative weights and complementarity between investments in periods $j' - 1$ and j'. The input $X_{j'}$ is a composite of parental time and good investments which we define below.

The constant ζ is an anchor that transforms children's human capital, which we will later proxy by test scores in the data, into adult outcomes, which we will measure using earnings (Cunha et al., 2010; Del Boca et al., 2014).[11] Specifically, define $\tilde{h}_{j'} \equiv h_{j'}/\zeta$ as pre-labor market skills in periods $j' \in \{0, 1, 2\}$. Since the production function is HD1, we can write the childhood skill formation process recursively as

$$\tilde{h}_1 \equiv \omega_0 X_0^{\phi_0}, \quad \tilde{h}_{j'+1} = \left[\omega_{j'} X_{j'}^{\phi_{j'}} + (1 - \omega_{j'}) \tilde{h}_{j'}^{\phi_{j'}} \right]^{\frac{1}{\phi_{j'}}}, \ j' = 1, 2, \tag{5}$$

[11] In practice, the exact interpretation is slightly different, mainly because it is not separately identified from ω_0. In our context, in addition to transforming test scores into meaningful units, ζ will also capture the productivity of initial investments.

which makes it explicit that one's future earnings is a function of skills formed in childhood. This technology displays dynamic complementarity and self-productivity across multiple stages (Heckman and Mosso, 2014) and has been investigated in several recent papers (Cunha et al., 2010; Caucutt and Lochner, 2012).

However, most specifications focus only on goods investments, while we also allow for time; conversely, Del Boca et al. (2014) focus only on time, and assume unit elasticities across periods.[12] Specifically, we model the investment $X_{j'}$ as

$$X_{j'} \equiv \left(l_{j'} h_j + \frac{\gamma_{j'} d_{j'}}{w_S} \right)^{\gamma_{j'}} \left(x_{j'} + (1 - \gamma_{j'}) d_{j'} \right)^{1-\gamma_{j'}} . \tag{6}$$

The inputs $(l_{j'}, m_{j'})$ are time and good investments in period j' made by a parent with human capital h_j, and $d_{j'}$ are government expenditures spent in education. This specification implies that higher human capital parents spend more time with their children, a salient feature of the data (Behrman et al., 1999).

Government expenditures are equally distributed across children and taken as given by all parents. Such expenditures can be used as both time and good inputs (e.g., teachers and school libraries, respectively). For lack of a better alternative, the above specification assumes that public investments are split between time and good investments in the same ratio as private parental inputs; the time expenditure component of $d_{j'}$ is divided by the parent's wage, w_S, to be transformed into parental time units. Specifically, let $I_{j'}$ total investment in a child of age j' in dollars:

$$I_{j'} = w_S h_j l_{j'} + x_{j'} + d_{j'},$$

then simple cost minimization implies that $w_S h_j l_{j'} / x_{j'} = \gamma_{j'} / (1 - \gamma_{j'})$ and

$$X_{j'} = \lambda_{j'} I_{j'}, \quad \text{where } \lambda_{j'} = (\gamma_{j'} / w_S)^{\gamma_{j'}} (1 - \gamma_{j'})^{1-\gamma_{j'}} \tag{7}$$

is the implied productivity of one additional dollar of investment (or the inverse of the

[12] However, it turns out that the estimated elasticities across periods are indeed close to one.

implied price of producing one additional unit of $X_{j'}$).[13]

This childhood technology formulation is important as it will capture the significance of parental inputs, as opposed to ρ_a, the persistence of learning abilities. The parametric restrictions chosen are selected parsimoniously so that they can be recovered from the available data, which we discuss later.[14]

2.3 Recursive formulation of decisions

Each period, choices are subject to the human capital production technologies in (1) and (5), which we suppress in the following formulations. Individuals in the working phase of the life-cycle are subject to the market luck shock (2), which are suppressed in the integral over tomorrow's continuation value.

In all periods j, we assume that the individual faces the life-cycle borrowing constraint

$$s_{j+1} \geq -g/(1+r), \tag{8}$$

where r is the period interest rate and g a lump-sum subsidy. That is, individuals can only borrow up to the amount that will be 100% repayable using the government transfer in the next period.

Period $j = 4$: Independence The newly independent adult solves a standard life-cycle savings problem:

$$\begin{aligned} V_4(S, a, h_4, s_4) &= \max_{c_4, s_5, n_4} \left\{ u(c_4) + \beta \int V_5(a'; S, a, h_5, s_5) dF(\epsilon_5) dG(a'|a) \right\} \\ \text{s.t.} \quad & c_4 + s_5 = f_4(e_4, 0) + s_4, \quad e_4 = w_S h_4 (1 - n_4), \quad n_4 \in [0, 1] \end{aligned} \tag{9}$$

[13] Our specification implies that private and public education expenditures are perfect substitutes. Although we could explicitly make private and public goods expenses less substitutable (Heckman and Mosso, 2014), this would be unidentified in the context of our model in which the government subsidies apply equally to all individuals with no variation.

[14] We could let children's skill accumulation explicitly depend on the child's ability, a', in which case parents with high a' children will anticipate this and invest more early on. But in Lee et al. (2015), we find that the estimated direct effect of one's a' on childhood human capital is quantitatively negligible.

along with the borrowing constraint (8). The first state S denotes whether the individual is high-school or college educated ($S = 0$ or 1, respectively), a the ability of the individual, and h_4 his level of human capital determined in the previous period. The last state, s_4, is a financial transfer from the parent that the young adult takes as given.

The function $f_j(e, s)$ denotes income net of a government tax-transfer program, that takes earnings and savings as inputs and is specified below. Note that since s_4 represents financial transfers that are made *within* period, it is not subject to the tax-transfer program. The variable e_4 captures the earnings of the adult, which are subject to taxes. In addition to time spent accumulating his own human capital, n_4, the individual makes consumption-savings decisions (c_4, s_5).

In addition to the i.i.d. luck shock to his own human capital, expectations are taken over the ability of the child the individual knows will be born tomorrow.

Periods $j = 5, 6, 7$: Investment in children During this stage, the parent (or family) faces the budget constraint

$$C_j + s_{j+1} = f_j(e_j, s_j) + s_j, \quad e_j = w_S h_j(1 - n_j - l_{j'}) - x_{j'}, \quad n_j \in [0, 1], \quad l_{j'} \in [0, n_j]$$

and the borrowing constraint (8). We assume that investment in children are deducted from parents' income subject to the government tax-transfer program.[15] In period 5, the parent takes care of his new-born child :

$$V_5(a'; S, a, h_5, s_5) = \max_{C_5, s_6, n_5, l_0, m_0} \left\{ q u(C_5) + \beta \int V_6(a', \tilde{h}'_1; S, a, h_6, s_6) dF(\epsilon_6) \right\} \tag{10}$$

where the child's ability is now included in the parent's state. The parent's own human capital and savings (h_5, s_5) are determined from yesterday's optimal choices, while the latter is also affected by the market luck shock.

When the child goes to primary school in period 6, his human capital accumulated in

[15] In reality, childcare in the U.S. is deductible, so this assumption is not necessarily stringent. Moreover, although later education expenses are not deductible in the U.S., the amount deducted in the model is quantitatively negligible. This is because the model also includes public expenditures in education, captured by $d_{j'}$. Public expenditures are small for childcare, but much larger for primary/secondary school, so private investments are much smaller for children at later ages.

period 5 appears as an additional state in the parent's period 6 value function:

$$V_6(a', \tilde{h}_1'; S, a, h_6, s_6) = \max_{C_6, s_7, n_6, l_1, m_1} \left\{ qu(C_6) + \beta \int V_7(a', \tilde{h}_2'; S, a, h_7, s_7) dF(\epsilon_7) \right\}.$$

In period 7 the child goes to high school, after which he first becomes eligible for work. Between periods 7 and 8, childhood skills are transformed into adulthood human capital, so next period's continuation value now includes h_3' rather than $\tilde{h}_3'$:

$$V_7(a', \tilde{h}_2'; S, a, h_7, s_7) = \max_{C_7, s_8, n_7, l_2, m_2} \left\{ qu(C_7) + \beta \int V_8(a', h_3'; S, a, h_8, s_8) dF(\epsilon_8) \right\}.$$

Period $j = 8$: Child in college The parent-child pair make a college decision:

$$V_8(a', h_3'; S, a, h_8, s_8) = \max_{S'} \left\{ W_8(S', a', h_3'; S, a, h_8, s_8) + \psi_S \cdot S' \right\}$$

where $S' = 1$ if the child goes to college and 0 otherwise. Hence ψ_S is a preference for college that depends on whether or not the parent went to college herself, $S \in \{0, 1\}$.[16] Once the decision is made, the child's education status becomes a new state:

$$\begin{aligned}
&W_8(S', a', h_3'; S, a, h_8, s_8) \\
&= \max_{C_8, s_9, n_8, n_3'} \left\{ qu(C_8) + \beta \int V_9(S', a', h_4'; S, a, h_9, s_9) dF(\epsilon_9) d\tilde{F}(\epsilon_4' | a') \right\} \\
&\text{s.t.} \quad C_8 + s_9 + \kappa S' = f(e_8, s_8) + f(e_3', 0) + s_8, \\
&\qquad e_8 = w_S h_8 (1 - n_8), \quad e_3' = w_{S'} h_3' (1 - n_3'), \quad n_8 \in [0, 1], \quad n_3' \in [\tilde{\kappa} S', 1],
\end{aligned}$$

and the borrowing constraint (8). The constant κ is the pecuniary cost of college, and children who go to college must spend at least 4 years accumulating human capital (rather than working), represented by $\tilde{\kappa} = 2/3$. Earnings of college-aged children are taxed independently, but children are assumed to have zero savings.

[16] We have also let ψ depend on a_k, but once S was included the impact of a_k was calibrated to be zero.

Period $j = 9$: Financial Transfers (Inter-vivos) Now the child is an independent adult, and the parent makes a financial transfer s'_4:

$$\begin{aligned} & V_9(S', a', h'_4; S, a, h_9, s_9) \\ & = \max_{c_9, s_{10}, n_9, s'_4} \left\{ u(c_9) + \theta V_4(S', a', h'_4, s'_4) + \beta \int V_{10}(S, a, h_{10}) dF(\epsilon_{10}) \right\} \\ & \text{s.t.} \qquad c_9 + s_{10} + s'_4 = f(e_9, s_9) + s_9, \quad e_9 = w_S h_9 (1 - n_9), \quad s'_4 \geq 0. \end{aligned} \tag{11}$$

The intergenerational transfer s'_4 is subject to a non-negativity constraint, meaning that parents cannot borrow against their children's future incomes, while s_{10} is still subject to the life-cycle borrowing constraint (8). We assume that the parent makes the transfer *after* observing ϵ'_4, but before the child makes any decisions. So the child takes the transfers as given when making his own decisions for the first time.

Period $j = 10, 11, 12$: Old Age and Retirement The parent expects to be retired in periods $j = 11, 12$ in which he lives off social security benefits, so no longer faces any uncertainty in period 10. Furthermore, the choice for $n_{10} = 1$, since human capital becomes useless after retirement, hence:

$$\begin{aligned} & V_{10}(S, h_{10}, s_{10}) = \max_{c_{10}, s_{11}} \left\{ u(c_{10}) + \beta u(c_{11}) + \beta^2 u(c_{12}) \right\} \\ & \text{s.t.} \quad \sum_{j=10}^{12} \frac{c_j}{(1+\tilde{r})^{j-10}} = f(e_{10}, s_{10}) + s_{10} + \frac{2+\tilde{r}}{(1+\tilde{r})^2} \cdot (p_0 + p_1 e_{10} + g), \quad e_{10} = w_S h_{10}, \end{aligned}$$

where we assume that after retirement, financial income is taxed at a flat rate, resulting in an after-tax effective interest rate of $\tilde{r}$. The parameters (p_0, p_1) capture the social security scheme, which is affine in an individual's last period earnings, and g is a lump-sum government subsidy. This implies that savings in periods 10 and 11 are

$$s_{11} = f(e_{10}, s_{10}) + s_{10} - c_{10}, \quad s_{12} = (1+\tilde{r}) s_{11} + p_0 + p_1 e_{10} + g - c_{11}.$$

2.4 Government, production and equilibrium

All income is taxed progressively, and earnings are subject to a social security tax τ_s. Given earnings and assets (e_j, s_j) in period j, after-tax income is

$$f_j(e_j, s_j) = [1 - \tau_y(y_j)]y_j + (1 - \tau_s)e_j + q_j \cdot g, \quad y_j \equiv e_j + rs_j \tag{12}$$

where $\tau_y(\cdot)$ is a progressive tax schedule, y_j is period income, and the period interest rate $r = (1 + \bar{r})^6 - 1$, where $\bar{r}$ is the annual interest rate. The lump-sum subsidy g is meant to capture welfare transfers, and q_j is an adult equivalent scale equal to 1 in periods $4, 8, \ldots, 12$ (adults with no children in the household); q_A in periods $5, 6, 7$ (households with children); and 2 in period 8 (child is college-age).[17] The constant τ_s is a flat-rate social security tax. The revenue is used to finance a PAYGO social security scheme received by retirees (parametrized by (p_0, p_1) above). As stated in the previous subsection, retirees do not face the progressive income tax, but instead pay a flat rate $\tilde{\tau}_k$ on their interest income from savings (which is their only source of income).

A representative firm uses physical capital, and high school and college human capital to produce the single consumption good. It solves

$$\max_{K, H_0, H_1} \{F(K, H_0, H_1) - RK - w_0 H_0 - w_1 H_1\}$$

where K is aggregate capital and (H_0, H_1) are aggregate quantities of utilized human capital in efficiency units for high school and college labor, which are imperfect substitutes:

$$F(K, H_0, H_1) = K^\alpha (AH)^{1-\alpha}, \quad H \equiv [vH_0^\sigma + (1 - v)H_1^\sigma]^{\frac{1}{\sigma}}. \tag{13}$$

The price of capital $R = (1 + \bar{r} + \delta)^6 - 1$ where δ the annual depreciation rate of capital.[18]

---

[17] Although college-age children are still assumed to make joint decisions with their parents, they are treated as adults in terms of taxation and transfers.

[18] The TFP parameter A is not qualitatively important in our stationary setting, but later calibrated so that average earnings equals 1.

The firm's profit maximization leads to the optimality conditions

$$RK = \alpha Y, \quad WH = (1-\alpha)Y, \quad \frac{w_1}{w_0} = \frac{1-v}{v}\left(\frac{H_1}{H_0}\right)^{\sigma-1} \tag{14}$$

where W is an aggregate wage index for H:

$$W = \left[v^{\frac{1}{1-\sigma}} w_0^{\frac{\sigma}{\sigma-1}} + (1-v)^{\frac{1}{1-\sigma}} w_1^{\frac{\sigma}{\sigma-1}}\right]^{\frac{\sigma-1}{\sigma}}. \tag{15}$$

Now let z_j denote the state space of an adult in period j, $z \equiv [z_4, \ldots, z_{12}]$ the aggregate state space spanning all generations, and $\Phi(z)$ its stationary distribution. Let $\bar{e}$ denote the average earnings in the economy:

$$\bar{e} = \int_S \left[e_3(z_8) + \sum_{j=4}^{10} e_j(z_j)\right] d\Phi(z) \Big/ 8,$$

since at any point in time, there are 8 generations that are working and we assume a uniform demographic structure. To define a stationary equilibrium, let $\Gamma(\cdot)$ denote the aggregate law of motion of z, which is derived from the agents' policy functions.

Definition 1 *In a stationary equilibrium, prices (r, w_0, w_1) and decision rules are such that*

1. *Given prices, agents of all ages make optimal choices;*

2. *The representative firm maximizes profit;*

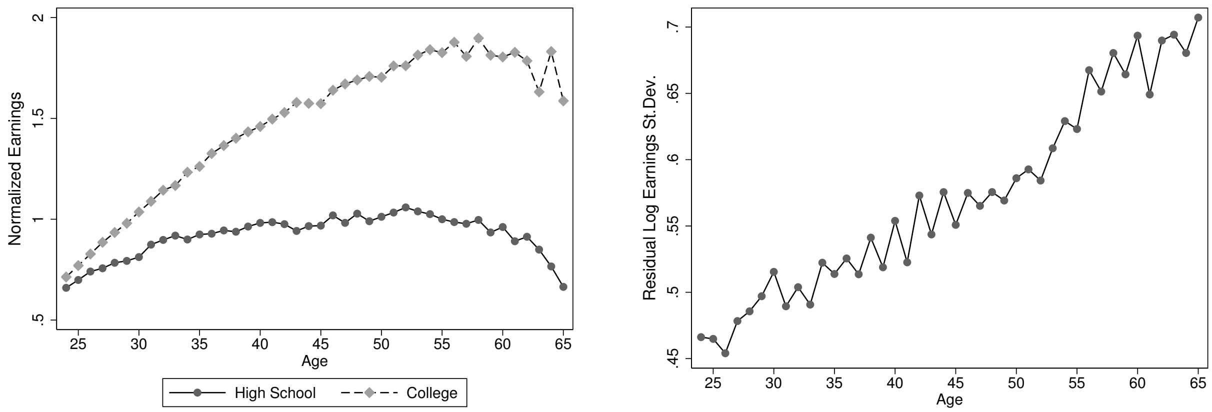
3. *Capital and labor markets clear:*

$$K = \int \left[\sum_{j=5}^{12} s_{j+1}(z_j)\right] d\Phi(z), \quad w_S H_S = \int_S \left[e_3(z_8) + \sum_{j=4}^{10} e_j(z_j)\right] d\Phi(z),$$

   which implies that the goods market clears;

4. *The social security budget balances:*

$$2\left(p_0 + p_1 \int e_{10}(z_{10}) d\Phi(z)\right) = 8\tau_s \bar{e}$$



(a) Average earnings profile

(b) Variance of residual log earnings

Figure 2: Life-cycle earnings

Source: PSID, 1969-2007. In the left panel, high school earnings at age 55 is normalized to one. Both panels are computed from a fixed effects regression including education, age and time (16), and the right panel plots the standard deviation of the residuals. The sample consists of 31,486 earnings observations from 1,981 heads of households.

5. *The distribution of z is stationary:* $\Phi(z) = \int \Gamma(z) d\Phi(z)$.

3. Data

Some of the model parameters are estimated from the PSID and the CDS. The data is also used to generate target moments for other parameters that are separately calibrated. Specifically, the adulthood part of the model is disciplined using the earnings of the heads of households in the PSID 1969-2007 family files, and the childhood part using the CDS and its Time Diary data files, which have three waves: 1997, 2002 and 2007.

3.1 Life-cycle Earnings

The PSID collects data on a representative sample of more than 5,000 American families, oversampling low-income families. Importantly for our purposes, the data includes earnings (labor income) and annual hours information for each family member.

Our empirical analysis is similar to Huggett et al. (2011), but we also differentiate between high school and college, which is defined as whether an individual's final education outcome is at least one year beyond high school graduation (or GED).[19] For individuals over 30, we keep only those years in which they work 520 hours or more and earn 1500 dollars or more, and for those 30 and below, only those those who work 260 hours or more and earn 1000 dollars or more (in 1968 prices). We also drop all observations in which an individual works more than 5820 hours per year. Top-coded earnings are multiplied by 1.5, which is a common ad-hoc correction procedure (Autor et al., 2006).

All earnings are then inflated to 2000 dollars using the GDP PCE deflator, after which we smooth individual earnings profiles using a 5-year moving average. Lastly, we only keep all heads of households aged 20 and above, and 65 and below.[20] This leaves us with 31,486 earnings observations from 1,981 heads of households.

Education Specific Age-Earnings Profiles First, we construct target moments for our quantitative model. Let E_{iat} denote the observed earnings of an individual i of age a at time t. We run the regression:[21]

$$\log E_{iat} = S_s + A_a + T_t + S_s A_a + S_s T_t + \varepsilon_{iat} \tag{16}$$

where (S_s, A_a, T_t) are education, age and time fixed effects, respectively. We also include interaction terms to completely separate education-specific earnings profiles by S_s (high school or college). We then compute the education-specific earnings profiles using the estimated coefficients as the average marginal treatment effect of each education-age category, assuming a balanced distribution for each education-age-time category.

Figures 2(a)-(b) depict the estimated profiles and residual log earnings variance (the variance of ε_{iat} by a). The earnings profiles are normalized so that high school earnings at age 55 equals one. As is well known, earnings follows a hump-shape with a much steeper profile for college workers, and residual log earnings variance rises with age. In our quantitative analysis, we target all three profiles, averaged by 6-year bins. Summary

[19] This is the same criteria as in Heckman et al. (1998). Increasing the education categories would be interesting, but in the context of our model becomes numerically infeasible.

[20] We do not differentiate between gender.

[21] As well known, time and cohort effects are not separately identified; we use the time effects approach as the benchmark following Huggett et al. (2011).

moments are tabulated in Table 1: the college earnings premium (an average marginal college treatment effect) of 49%, the education-specific earnings profiles, and the age-specific log earnings variance. We also target a college enrollment rate of 43%: the fraction of college workers in the sample.

Earnings volatility We use old age individuals in the PSID to compute the mean and variance of the market luck shocks, $(\mu_\epsilon, \sigma_\epsilon^2)$. In the Ben-Porath model, workers stop accumulating human capital toward the end of their working life, which in our model corresponds to the last two period $j = 9, 10$. If agents spend zero time investing in their own human capital, earnings become functions of only the market wage rate and earnings shocks, that is:

$$\log e_{10} - \log e_9 = \log w_S h_{10} - \log w_S h_9 = \log \epsilon_{10}.$$

Since ϵ_j is assumed to be i.i.d. across individuals and age, we can estimate the the mean and variance of ϵ_j simply by looking at the sample mean and variance of the growth rate of old workers' earnings:

$$\hat{\mu}_\epsilon = \hat{\mathbb{E}}[\log e_{10}] - \hat{\mathbb{E}}[\log e_9], \quad \hat{\sigma}_\epsilon^2 = \hat{\mathbb{V}}[\log e_{10}] - \hat{\mathbb{V}}[\log e_9], \tag{17}$$

where $\hat{\mathbb{E}}$ and $\hat{\mathbb{V}}$ denote, respectively, the sample mean and variance operators. We compute these statistics for high school and college separately.

In practice, for μ_ϵ, we simply compute the mean slope of the earnings profile in Figure 2(a) at the end of the life-cycle, separately for high school and college. For σ_ϵ^2 we first compute hourly wage rates in the data.[22] Then we smooth annual hours worked using 5-year moving averages, and compute the wage rate for each observation, W_{iat}, as smoothed earnings divided by smoothed hours. Since a model period is 6 years, we keep all individuals aged 60 to 65. For these individuals, we compute their log-wage difference

[22]This is a common approach in models using Ben-Porath technologies.

Parameter	μ_ϵ	σ_ϵ	# Obs.
High School	-0.13	0.13	754
College	-0.10	0.21	712

Table 1: Earnings Volatility Parameters estimated from PSID

μ_ϵ is recovered from the mean slope of old age individuals in Figure 2(a), and σ_ϵ from the variance of the residuals from regressing equation (18), assuming human capital is no longer accumulated from age 54 onward. Refer to text for details.

between ages 54 and 60, 55 and 61, and so forth. Then we run the fixed effects regression

$$d \log W_{iat} = u_i + S_s + A_a + T_t + S_s[A_a + T_t] + \varepsilon_{iat}, \tag{18}$$

where $d \log W_{iat}$ is the 6-year log-wage difference, and cluster standard errors at the individual level. This filters out individual and time fixed effects, and different trends in each 6-year distanced pair. We then take the variance of the residual as an estimate for σ_ϵ^2, for high-school and college separately. The results are tabulated in Table 1, which is based on 507 old individuals with 1,466×2 wage observations. As shown in the table, the shocks are declining on average but less for college, and the volatility is higher for college.

3.2 Investment in children

In 1997, 2002 and 2007, the PSID collected detailed data on investment in children and children's outcomes for families with children aged 12 and below. The initial wave consisted of approximately 3,500 children in 2,400 households. The range of information collected was large, of which we use three: the time diaries, private and public money expenditures for children, and children's Letter-Word test scores. These are used to estimate the childhood skill formation technology (5).

For the purposes of our study, we take a unit of observation as a child.[23] The CDS contains information on primary and secondary caregivers, who may or may not be a parent, and also may or may not be in the child's household. We merge information on adults in the CDS into the PSID using household and individual identifiers, and only keep

[23] That is, we do not take into consideration that time and money expenses reported by a parent may be spent on more than one child.

those children who live with at least one biological parent, and for whom both caregivers in the CDS correspond to the head or wife in a PSID family unit.[24] Then we use the same criteria as in Section 3.1 and drop observations if parents' earnings are too low, or hours too low or high. We also drop families in which a parent is less than 18 or more than 42 years older than the child. This leaves us with 4,402 observations over the 3 waves of the CDS.

Time spent with children Each child in the CDS submitted a detailed 24-hour time diary for one weekday and one weekend-day.[25] For each activity listed, children were also asked to list whether or not a parent (or another adult) was present, and if so, whether the parent was just around or participating in the given activity. Del Boca et al. (2014) refer to this as "active" and "passive" time, respectively.

We follow their strategy and first aggregate the time each parent spent with a child, resulting in 8 categories (2 parents×2 days×(active,passive)). When doing so, we adjust weekday hours so that average hours spent in each category is equal across children of the same age. Weekend hours are similarly adjusted. Specifically, for each category, adjustments are made to raw data following

$$l_i(\text{Adj. Day}) = l_i(\text{Raw Day}) \times \frac{\bar{l}(\text{Mon-Fri})}{\bar{l}(\text{Day})}$$
$$l_i(\text{Adj. Sat}) = l_i(\text{Raw Sat}) \times \frac{\bar{l}(\text{Sun})}{\bar{l}(\text{Sat})} \qquad l_i(\text{Adj. Sun}) = l_i(\text{Raw Sun}) \times \frac{\bar{l}(\text{Sat})}{\bar{l}(\text{Sun})}$$

where l_i is an individual observation, "Days" run from Monday through Friday, and $\bar{l}$(X) denotes the average hours spent per day during X.[26]

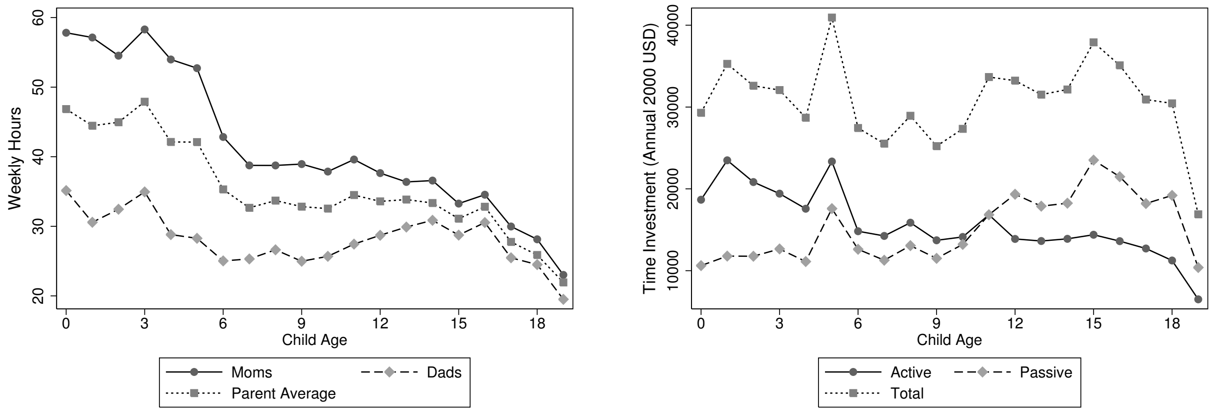
We then compute weekly hours spent with children for moms and dads, and their average, by multiplying weekday hours by 5 and weekend hours by 2 and then adding them. The mean values by children's age are plotted in Figure 17 in Appendix C.[27] Figure

[24]More than 90% of primary caregivers are biological mothers. We keep single parents as long as they are the primary caregiver.

[25]For younger children, the diary was filled by a caregiver.

[26]For this normalization, we use the raw CDS data before merging with the PSID. There are a total of 6,915 non-missing observations over the 3 waves. We use CDS-provided sampling weights to compute means, also adjusting the weights so that the sum of weights within a wave is equal.

[27]As expected, moms spend more time with children than dads. It is also not surprising that active time



(a) Total time (b) Time cost

Figure 3: Opportunity cost of time investment in children
Source: PSID, Child Development Supplement Time Diaries. "Active" time is defined as when children report a parents participated in their activity, and "passive" time when they report a parent was around but not participating. Time costs are computed by multiplying parents' weekly time spent with children by their annual wage rate (earnings/hours)×52. All values are in 2000 USD.

3(a) shows that the sum of active and passive time still shows a clear downward trend for moms, but the trend for dads appears to be flat.

Next we compute the opportunity cost of parents' time spent with children, using parents' labor market information in the PSID. We obtain the hourly wage rate for each parent by dividing parents' earnings by annual hours. By multiplying these rates by parents' weekly time spent with children×52, we compute the annual cost of time investments, separately for active and passive time. The result is shown in Figure 3(b). Interestingly, while the active time cost is clearly declining with children's age, passive time costs are increasing. This is most likely due to dads spending more passive time with children as they age, who also tend to earn more than moms.

In the model, time spent with children comes at the expense of parents working or accumulating their own human capital. When parents are around but not participating, it could mean that parents are engaged in either activity at the same time rather than ed-

spent with children declines with age. However, passive time spent with children does not seem to follow a particular trend, although it seems moms tend to spend less and dads more time as their children age. Despite differences in sample selection criteria—they focus only on one- or two-child families in which both biological parents are present, and further analyze children and one- and two-child families separately—all these features are similar to results in Del Boca et al. (2014).

ucating their children.[28] Furthermore, passive time is not only noisier but also estimated to have much lower productivity than active time in Del Boca et al. (2014). For all these reasons, we will use only active time costs as the measure of parental time inputs into children's skill formation in our estimation below.

Letter-Word test scores To estimate the child skill formation technology (5), we need a measure of children's outcomes $\tilde{h}$. As is common in the literature, we use children's test scores as our outcome variable. Specifically, we use the Letter-Word test score outcomes administered to all children in the CDS.[29] The standard LW-test comprises 57 questions, and the CDS records whether the child answered each question correctly or not (1-0).

Since we want to capture children's cognitive development as they age in addition to heterogeneity, we adjust the raw test score as follows. Let d_q denote the fraction of children who answer question q correctly, regardless of age.[30] We then assume that question q is worth d_q points, and sum up across questions to obtain a child's adjusted test score.

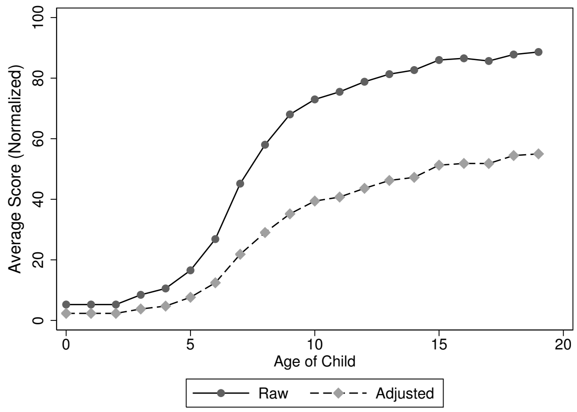
In Figure 4(a), both raw and adjusted test scores are normalized to lie between 0-100.[31] The figure shows that test scores increase with age, flattening out at later ages. Adjusted scores are less steep than raw scores at earlier ages but as steep at later ages, resulting in a smoother average increase in scores over age. More interestingly, the standard deviation of raw test scores (after being normalized to lie between 0-100) displays a spike in early ages. This is likely because among very young children, some children mature earlier and begin to answer many easy questions correctly, while others lag behind. As children age, the variance declines as most of them get all the easy questions.

[28] Passive time could also be parents' leisure time.

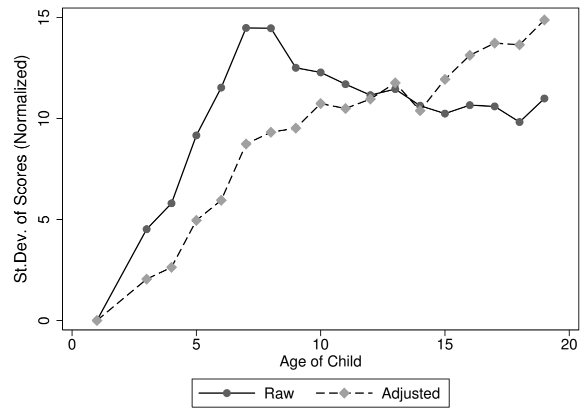
[29] Clearly, this is a measure of children's cognitive skills. While non-cognitive skills are also important inputs into adult outcomes, including multiple skills is beyond the scope of this paper. Furthermore, test scores such as AFQT typically have strong prediction power for earnings (Belley and Lochner, 2007), which maps into adult human capital h in the model. The CDS contains other test scores as well, but we use LW-scores as it is the only test administered to all children of all ages.

[30] We use the same sampling weights as when we weight parental time usage. As shown in Figure 18 in Appendix C, higher number questions are designed to be much more difficult, with almost no children getting the last question correct.

[31] Alternatively, one could run an age-time fixed effects regression as we did for earnings. However, the LW-test is a standardized test so that each question addresses a similar question, with the difficulty progressing with question number but not varying much across tests. Regardless, including time fixed effects barely changes the data.



(a) Letter-Word test scores



(b) Test score variance

Figure 4: Letter-Word Test Scores
Source: PSID, Child Development Supplement. Letter-Word scores are normalized to lie between 0-100. Adjusted test scores are computed by weighting each of the 57 questions by the inverse of the fraction of children who answered correctly.

When weighted by difficulty, questions that any child would answer correctly after a certain age becomes less important for the adjusted test score. Strikingly, the data then shows that children's test score variance increases with age, just as earnings variance increases monotonically with adult's age. This suggests that human capital differences across individuals begin at a much earlier age than college or labor market entry; Indeed, the variance increases monotonically already from age 2 (test scores for even younger children are unavailable; the values in Figure 4 are predicted values). We take these normalized, adjusted test scores as $\log \tilde{h}$ when we estimate technology (5) below.

Childcare and educational money expenditures For money investments in children, we focus only on expenditures related to children's cognitive skills such as costs of childcare, money spent on schooling (this includes private school tuition and school-related supplies) and extracurricular activities (such as private tutoring and lessons). Unlike time and test score data, this data is extremely noisy in the CDS: Only about 10% of the sample has reliable expenditure data on the costs of childcare, and among children above age 5, only about half of the parents report expenses on extracurricular activities.

Moreover, since most childcare is irregular, the survey asks for at least 4 types of childcare arrangements, and for each type, the amount of money spent, how frequently the

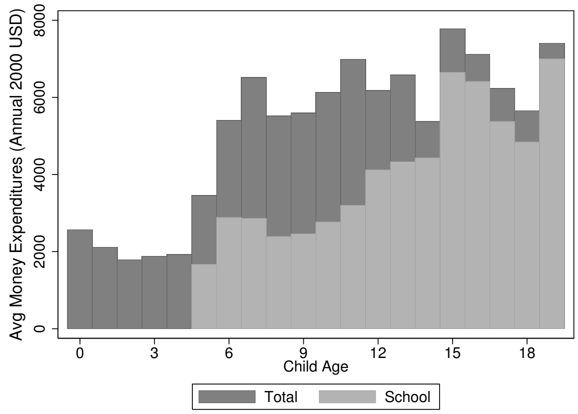


Figure 5: Money investments in children
Source: PSID, Child Development Supplement Time Diaries. All values are in 2000 USD. Money investments include childcare, schooling tuition and supplies, and extracurricular activities.

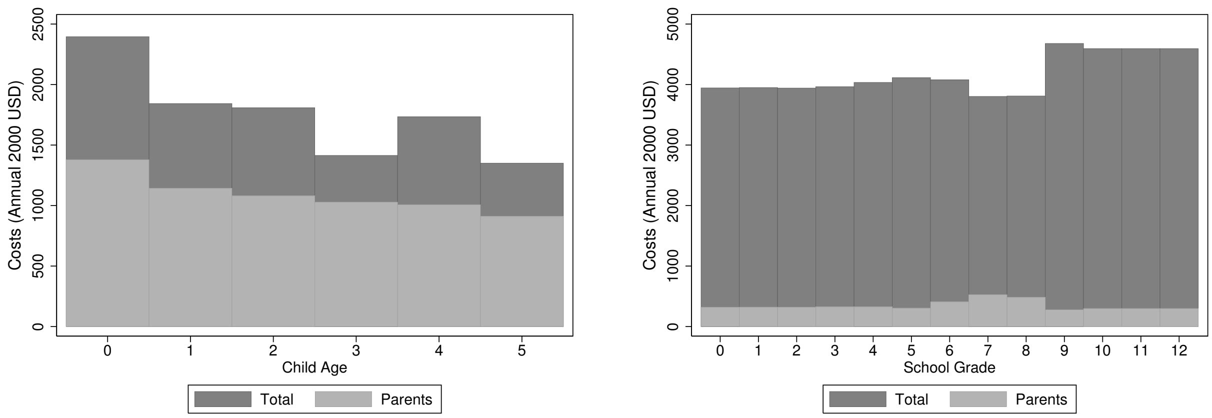
cost is incurred and how long the arrangement was used. There is also data on special arrangements, such as during weekends or summers. We drop all observations if no costs were incurred or we are unable to transform the costs into annual dollars. For extracurricular activities, we only include those that are at least indirectly related to education, such as tutoring or participation in community programs.[32]

Average private expenditures by age are shown in Figure 5. Although the data is noisy, its shows an increasing pattern for money expenditures as children grow older. At the same time, schooling costs, mostly in the form of private school tuition, crowds out other expenditures as children grow older.

For most children above schooling age, however, public schooling is the dominant source of money expenditures. For public schooling expenses, we refer to the 1997 CDS school administrator files. The CDS asked administrators of the daycare center or school the child was attending questions such as how many students the institution had by age (for daycare centers) or by grade (for schools), the average amount of dollars spent per child/student, and average fees charged to parents.

Using this information, we construct average dollars spent per child for each age, by averaging over the mean dollars spent and fees charged by all institutions weighted

[32]While the final, reliable number of observations is small, the amount of available data is sparse but large, requiring an immense amount of cleaning. More details on how we construct the money expenditure data are available upon request.



(a) Daycare costs

(b) Schooling costs

Figure 6: Institutional cost of children's education
Source: PSID, Child Development Supplement. All values are in 2000 USD.

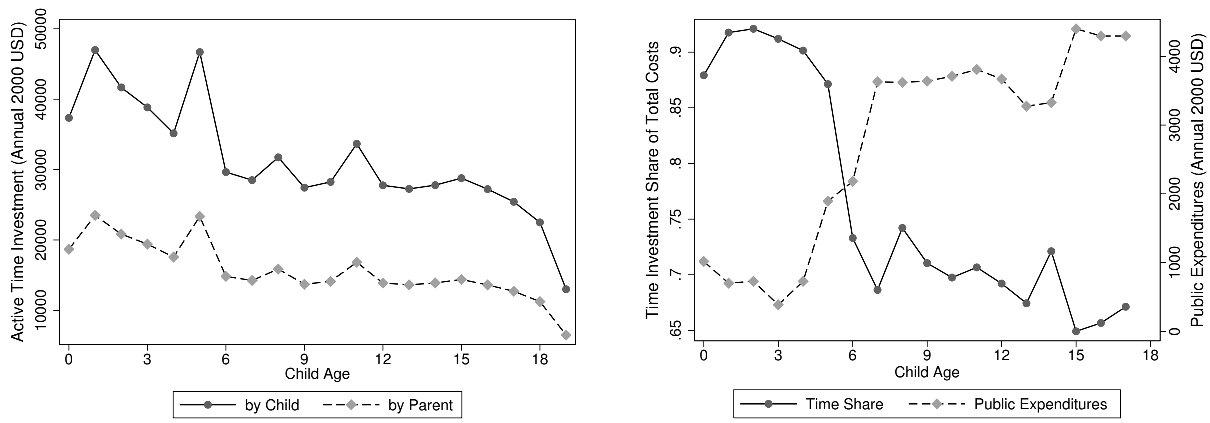
by the number of children they report having in each age/grade and the CDS sampling weights.[33] As shown in Figure 6, both daycare and schooling costs do not vary much over age, with discrete jumps when children begin kindergarten (grade 0 in Figure 6(b)) and high school (grade 9). But parents bear most of the cost burden for daycare, while they pay less than a few hundred dollars on average for schooling.

3.3 Estimating children's skill formation technology

Armed with information on parents' time spent with children, children's test scores, and private and public money expenditures, we recover the three important parameters of the childhood skill formation technology in (5), namely (ϕ_0, ϕ_1, ϕ_2), the complementarity between different stages of child investments. We will exploit the relationship between parents' active time investments and children's adjusted test scores between different waves of the CDS, while using private expenditure data to discipline aggregates and public expenditure data to back out government subsidies.

The model assumes single parent, single child households. Since our unit of observation in the data was per child, we first normalize the active time investments in children by the number of parents. As shown in Figure 7(a), the average cost per parent is slightly

[33]This assumes that the distribution of schooling costs is similar to the distribution of children.



(a) Average cost of time investments

(b) Recovering $(\tilde{\gamma}_a, \tilde{d}_a)$

Figure 7: Moments used for Childhood Skill Production Estimation
Source: PSID, Child Development Supplement. In the left panel, average cost of time investments takes only into consideration the active time costs in Figure 3(b). In the right panel, $\tilde{\gamma}_a$ denotes the time share of total private investments in children, and $\tilde{d}_a$ the dollar amount of government expenditures, at age a.

above half of the average cost per child. About one-fifth of the sample are single parent households, of whom 95% are single moms.

Now skill-formation technology (5) and the technology for producing the effective input $X_{j'}$, (6), imply that the share of time and good investments are determined by

$$w_S h_j l_{j'} = \gamma_{j'} \left(X_{j'}/\lambda_{j'} - d_{j'} \right), \qquad x_{j'} = \left(1 - \gamma_{j'}\right) \left(X_{j'}/\lambda_{j'} - d_{j'} \right).$$

In the data, $w_S h_j l_{j'}$ corresponds to the opportunity cost of time spent with children, and $x_{j'}$ to money costs. For each age a, we compute the ratio of mean time and money investments in children shown in Figures 5 and 7(a), and denote this value $\tilde{\gamma}_a$. The values of $\tilde{\gamma}_a$ are shown in the left axis of Figure 7(b), and as expected, the weight on time investments declines as children age, most of which happens when the child begins school.

Next, we fix the values of $d_{j'}$ using information in Figure 6. First, we assume that kindergarten is age 6 and grade 12 is age 18, and merge the institutional cost data for daycare centers and schools to construct a unified series for ages 0-17.[34] For each age, we then subtract the average fees paid by parents from the average dollars spent per student,

[34] While not shown in the figures, a small number of daycare centers provide pre-K and kindergarten education, and some schools provide pre-K education as well.

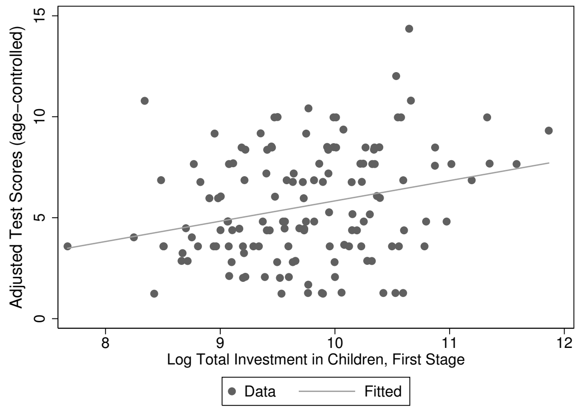


Figure 8: Childhood Skill Production Estimates: ϕ_0
Source: PSID, Child Development Supplement. Refer to text for the dependent and independent variables used for the regression in the right panel.

which we take as a dollar measure for public subsidies, denoted $\tilde{d}_a$.[35] The dollar amounts of $\tilde{d}_a$ are shown in the right axis of Figure 7(b) and is increasing with age, mostly when the child begins school.

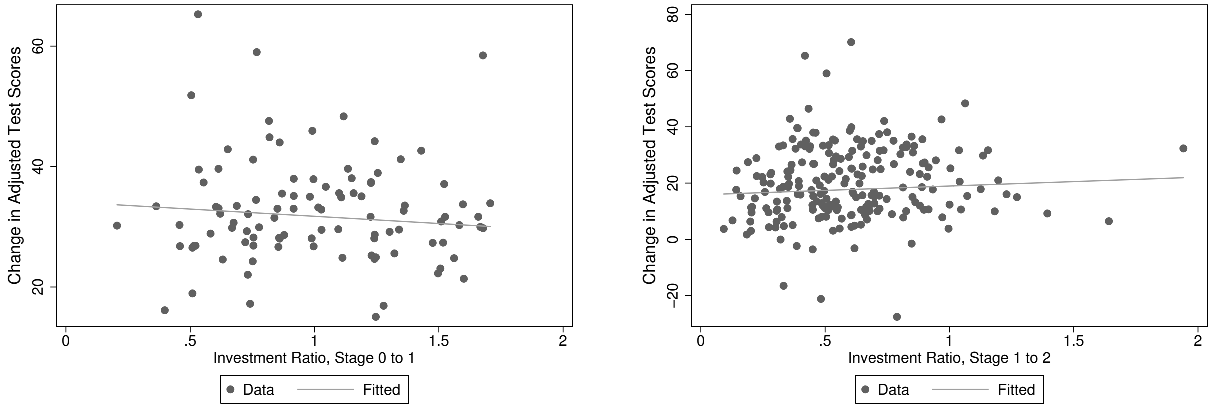
We now construct an auxiliary measure $\tilde{I}_{i,a}$, the total investment in children, for each child i of age a. Let $T_{i,a}$ denote the implied time cost investment (the average opportunity cost of parents' active time investment in children) observed in the data. Then given the values of $(\tilde{\gamma}_a, \tilde{d}_a)$, implied total investment in children is

$$\tilde{I}_{i,a} = T_{i,a} / \tilde{\gamma}_a + \tilde{d}_a,$$

and since $\tilde{X}_{i,a} = \lambda_a \tilde{I}_{i,a}$ from (7), the skill formation technology (5) between periods 0 and 1 (child ages 0-5 to 6-11) implies

$$\left(\frac{\tilde{h}_{i,a}}{\tilde{h}_{i,a-6}} \right)^{\phi_1} = \omega_1 \left(\frac{\tilde{X}_{i,a}}{\tilde{h}_{i,a-6}} \right)^{\phi_1} + 1 - \omega_1$$

[35]As noted earlier, private money expenses are largely comprised of private school tuition for older children, while the implied school fees paid by parents are small in the administrator files. Nonetheless, we continue to include private schooling costs as shown in Figure 5, because we believe that most parents indeed to spend more money on their children at later ages, which is not reflected in other costs simply due to the lack of quality data. Implied money expenses are also much larger than in the data in Del Boca et al. (2014), who do not make any use of this data for similar reasons. Moreover, our estimation strategy only exploits variation in time investments and is insensitive to the value of the $\gamma_{j'}$'s, although our later quantitative results potentially are.



(a) Recovering ϕ_1

(b) Recovering ϕ_2

Figure 9: Childhood Skill Production Estimates: ϕ_j
Source: PSID, Child Development Supplement.

while cost minimization across periods implies

$$\frac{(1+r)\tilde{I}_{i,a}}{\tilde{I}_{i,a-6}} = \frac{\omega_1}{1-\omega_1} \cdot \left(\frac{\tilde{X}_{i,a}}{\tilde{h}_{a-6}}\right)^{\phi_1}$$

from which we obtain

$$\underbrace{\log\left[\frac{(1+r)\tilde{I}_{i,a}}{\tilde{I}_{i,a-6}} + 1\right]}_{\Delta Y_{i,a}} = -\log(1-\omega_1) + \phi_1\left(\log\tilde{h}_{i,a} - \log\tilde{h}_{i,a-6}\right) \tag{19}$$

and similarly between ages 6-11 and 12-17 we obtain

$$\underbrace{\log\left[\frac{(1+r)\tilde{I}_{i,a}}{\tilde{I}_{i,a-6} + \frac{\tilde{I}_{i,a-12}}{1+r}} + 1\right]}_{\Delta Y_{i,a}} = -\log(1-\omega_2) + \phi_2\left(\log\tilde{h}_{i,a} - \log\tilde{h}_{i,a-6}\right), \tag{20}$$

where the denominator on the left-hand side is now the cost of human capital production spanning both periods $a-6$ and $a-12$. Assuming that $r = 1.04^6 - 1$, as we do in the calibration, the $\Delta Y_{i,a}$'s on the left-hand sides are observed in the data.

Parameter	ϕ_0	ϕ_1	ϕ_2
Estimate	1.01^{***}	0.00	0.00
	(0.32)	(0.00)	(0.00)
# Obs.	140	101	215

Table 2: Parameters estimated from the CDS (3 waves)

Estimates are recovered from an OLS regression of equations (19)-(20). Refer to text for details on how these equations are constructed.

For the first period, children aged 0-5, we have from (5) and (7) that

$$\log \tilde{h}_{i,a} = \log \omega_0 + \phi_0 \log \tilde{X}_{i,a} = \log \omega_0 + \phi_0 \log \lambda_a + \phi_0 \underbrace{\log \tilde{I}_{i,a}}_{Y_{i_a}}$$

Assuming childhood human capital $\tilde{h}_{i,a} = b \exp(LW_{i,a})$, where $LW_{i,a}$ is the adjusted LW-score of child i of age a, we can recover the three parameters (ϕ_0, ϕ_1, ϕ_2) from

$$Y_{i,a} = A_a + \phi_0 \tilde{I}_{i,a} + \epsilon_{i,a} \qquad \text{for ages 0-5,} \tag{21}$$

$$\Delta Y_{i,a} = B_j + \phi_j \Delta LW_{i,a} + \epsilon_{i,a} \quad \text{for period } j = 1, 2 \text{ and ages } 6j \text{ to } 6j + 5, \tag{22}$$

where A_a is a complete set of age dummies for the first period, $\Delta LW_{i,a} = LW_{i,a} - LW_{i,a-6}$, and B_j are regression constants.

Since test scores are only available for children age 2 and above, the sample size for estimating ϕ_0 in (21) is rather small at 140. Similarly, to conduct regressions (22), we need children for whom we observed time investments and test scores for consecutive waves in the CDS.[36] Consequently, the estimation of ϕ_1 is also small at 101, since we lose additional observations due to data attrition. The sample size for estimating ϕ_2 is 215, for which we need 3 consecutive observations of time investments, but only 2 for test scores. All estimates are tabulated in Table 2, and the slopes of $\phi_{j'}$ are visualized in Figure 9.

While investments across periods are far from perfectly substitutable, as emphasized by Heckman and Mosso (2014), the estimates of (ϕ_1, ϕ_2) are close to 0 with small standard errors, implying that the technology is close to Cobb-Douglas. Moreover, the estimate for

[36] Unfortunately, the CDS was conducted every 5 years while our model period is 6 years, but for lack of a better alternative we ignore this 1 year difference.

ϕ_0 is equal to unity, but more importantly, is close to linear.[37] Despite the restriction in sample size, since all estimates are tightly estimated, in our subsequent calibration we set the values of the ϕ_j's to $(1, 0, 0)$: that is, initial human capital is linear in inputs and the dynamic skill formation technology is Cobb-Douglas.

4. Calibration

Given the estimates and targets from the PSID and CDS in Tables 1-2, we now turn to our calibration and quantitative analysis. We first describe how we discipline the remaining parameters of the model, followed by an analysis of its benchmark properties.

4.1 Parameters Set Exogenously

Several parameters are fixed at standard values in the literature, while policy variables are set to their empirical counterparts. The rest are computed from a method of moments by numerically simulating the model. These parameters are summarized in Tables 3-7.

Preferences and technology The CRRA coefficient χ is fixed at 2, which is approximately the midpoint of values used in human capital models with Ben-Porath technologies. Capital income share α and depreciation rate of capital δ are set to $(0.32, 0.07)$, which are the values used in Huggett et al. (2011) and consistent with the U.S. data on long-run capital income shares and return to capital. The elasticity parameter σ is set to (1-1/1.44), which are estimated in Katz and Murphy (1992); Heckman et al. (1998) to aggregate time trends in labor supply.

The parameters that govern the time share of investments in children, $(\gamma_0, \gamma_1, \gamma_2)$, are set to the mean values of their empirical counterparts $\tilde{\gamma}_a$ depicted in Figure 7(b), averaged over the corresponding age intervals of 0-5, 6-11, and 12-17.

[37] If we had set $\tilde{h} = b \exp(\tilde{h}^a)$, the parameter a would not be separately identified from the ϕ_j's, which we estimate, and the ω_j's, which we calibrate. Hence, we have normalized it to 1. What is really striking about the estimate of ϕ_0 is the strong evidence of linearity, not so much the slope itself.

Parameter	*Value*	*Description*
χ, α, δ	1.5, 0.32, 0.07	previous literature (Browning et al., 1999; Huggett et al., 2011)
σ	0.31	skill elasticity 1.44 (Katz and Murphy, 1992; Heckman et al., 1998)
$\gamma_0, \gamma_1, \gamma_2$	0.90, 0.71, 0.68	Time investment in children's skill formation [Figure 7(b)]
Set in equilibrium		
$\bar{\beta} = \beta^{1/6}$	0.98	interest rate 4%
v	0.70	college enrollment rate 48%
A	1.674	average earnings controller

Table 3: Fixed Parameters

Parameters set in equilibrium The discount factor β is calibrated to an implied annual interest rate of $\bar{r} = 4\%$ in the benchmark equilibrium, consistent with historical data on long-run asset returns. Skill prices (w_0, w_1) are not directly observed in the data, so we instead calibrate v, the weight on high school human capital, as follows. Let $L_0 \equiv 1 - L_1$ denote the employment share of high school workers. Using (14), we can write the college earnings premium as

$$EP = \frac{w_1 H_1 / L_1}{w_0 H_0 / L_0} = \frac{1 - v}{v} \left(\frac{H_1}{H_0} \right)^{\sigma} \cdot \frac{L_0}{L_1} \quad \Rightarrow \quad \frac{H_1}{H_0} = \left(\frac{v}{1 - v} \cdot \frac{L_1}{L_0} \cdot EP \right)^{\frac{1}{\sigma}}$$

and plugging this back into (14) we obtain

$$\frac{w_1}{w_0} = \left(\frac{1 - v}{v} \right)^{\frac{1}{\sigma}} \left(\frac{L_1}{L_0} \cdot EP \right)^{\frac{\sigma - 1}{\sigma}}. \tag{23}$$

Next, in a stationary equilibrium, the aggregate wage index in (14) must satisfy

$$W \equiv (1 - \alpha) \left[\frac{\alpha}{(1 + \bar{r} + \delta)^6 - 1} \right]^{\frac{\alpha}{1 - \alpha}}$$

and since $(\alpha, \delta, \bar{r})$ are parameters, W also becomes a parameter. But then from (15),

$$w_0 = W \cdot \left[v^{\frac{1}{1 - \sigma}} + (1 - v)^{\frac{1}{1 - \sigma}} (w_1 / w_0)^{\frac{\sigma}{\sigma - 1}} \right]^{\frac{1 - \sigma}{\sigma}}. \tag{24}$$

Parameter	*Value*	*Description*
τ_0, τ_1	0.10, 0.04	progressive income tax Guner et al. (2014)
g	0.03	welfare transfers 2% of GDP
q_A	1.70	OECD adult equivalence scale
p_1	0.33	median social security replacement bracket
p_0, τ_s	0.08, 0.11	social security budget balance and 40% replacement rate (Diamond and Gruber, 1999)

Table 4: Government tax-transfer system parameters

The empirical values of EP and (L_0, L_1) in (23) are observed in the data (Table 1), so if we know v, (24) determines (w_0, w_1) separately. Hence, we calibrate v so that the model implied high school employment share is equal to its empirical value in equilibrium. Later, other parameters are calibrated so that the model implied college premium is equal to its empirical counterpart, which implies labor market clearing.

We also ensure that mean earnings, $\bar{e}$, is equal to 1 in our benchmark equilibrium. We do so by varying the level of TFP, A, which simply shifts wages uniformly for all individuals. This normalization is useful when setting the model's policy parameters below, most of which are normalized by mean earnings.[38]

Tax-transfer system We parametrize the progressive income tax function in (12) as

$$\tau(y) = \tau_0 + \tau_1 \log(y/\bar{y}), \tag{25}$$

where $\bar{y}$ is mean income, following Guner et al. (2014), and use their estimates for (τ_0, τ_1). Since mean savings $\bar{s}$ equals capital per worker in equilibrium, using (14) we can write mean income as

$$\bar{y} = \bar{e} + r\bar{s} = \bar{e}\left(1 + \frac{r}{R} \cdot \frac{\alpha}{1-\alpha}\right),$$

which is a function of the fixed parameters (r, R, α).

We view the lump-sum transfer g primarily as welfare payments for the poor, and set it to 3% of mean earnings. The size of welfare transfers in the U.S. was approximately

[38] Mean earnings in our PSID sample is 38,338 in 2000 U.S. dollars.

1-2% of total GDP throughout the late 1980s to mid-1990s, of which we take the upper-bound 2% and divide by $1-\alpha$ to make it a fraction of mean earnings rather than GDP per worker.[39] The adult equivalent scale, q_A, which increases transfers for families with children, is set to 1.7. This is the correction used by the OECD to compare the consumption of two-adult households with two-children against those without children.

Social security payments in the U.S. is based on the average of a retired individual's highest 36 years of earnings. The replacement scheme is affine with kinks, with most individuals receiving 32 cents more per a 1$ increase in their average earnings. Since the model system is based on last period earnings only, we adjust this factor and set $p_1 = 0.33$.[40] Given this, we choose p_0 and the payroll tax τ_S to balance the social security budget and match a replacement rate of 40%, as reported in Diamond and Gruber (1999). Since all working adults pay social security taxes, this implies

$$p_0 + p_1 \bar{e}_{10} = 4\tau_s \bar{e}, \quad p_0 + p_1 \tilde{e}_{10} = 0.4 \tilde{e}_{10},$$

where $\bar{e}_{10}$ is mean earnings from ages 60-65 and can be computed from the PSID. This results in $\tau_S = 0.11$ and $p_0 = 0.08\bar{e}$. All policy parameters are summarized in Table 4.

Education Similarly as the time investment shares, public subsidies in children's education (d_0, d_1, d_2) are obtained from the mean dollar values of their empirical counterparts $\tilde{d}_a$ in Figure 7(b), averaged over the age intervals of 0-5, 6-11, and 12-17, then divided by meaning earnings in the PSID. The exact values are summarized in Table 5.

We refer to *Trends in College Pricing*, published annually by the College Board, to construct the college cost parameter κ. We exclude room and board and only include tuition and fees, since all individuals incur living costs (through consumption) regardless of college attendance. Table 5 shows the average costs of attending a 2-year public, 4-year public, and 4-year private college in the two years 1980 and 2005. 2-year colleges are

[39] We take the upper-bound since the model transfers are not means-tested while real-world welfare transfers, such as AFDC, are. Hence the transfers need to be larger to match the amount received by the poorest households.

[40] The adjustment is made according to mean age-earnings profiles in the PSID. As is common in the literature, we make social security payments depend only on last period earnings due to numerical complexities.

Parameter	*Value*	*Description*			*Mean*
d_0	0.02	Public expenditures in children's			908
d_1	0.09	skill formation [Figure 7(b)]			3,494
d_2	0.10				3,846
		Cost of College ($)	1980	2005	
		Two-year public	789	1,965	
κ	0.19	Four-year public	1,641	4,925	7,059
		Four-year private	7,170	19,046	

Table 5: Education subsidies and cost of college
Public expenditures and costs of college are in 2000 USD.

much cheaper than 4-year colleges, which are in turn substantially cheaper than private colleges. The table also shows that the cost of college has been rising over time: from 1980 to 2005, average costs more than doubled for all types of colleges. Since we do not differentiate between these types of colleges, we simply assume that the first two-years of college costs the average of all three types, and the latter two-years the average of 4-year public and private colleges. Then we take the mean of these two values, and divide it by mean earnings to obtain the model cost of college κ.

4.2 Method of Moments

The 10 remaining parameters are chosen to minimize the distance between 27 equilibrium moments simulated by the model and empirical moments from the PSID, CDS, and three intergenerational persistence moments we take from previous literature, summarized in Table 6. Specifically, the parameter vector is

$$\Theta = [\theta \; \rho_a \; \mu_a \; \sigma_a \; b \; \zeta \; \omega_1 \; \omega_2 \; \psi_1 \; \psi_2]'$$

whose calibrated values are summarized in Table 7.

Altruism and the exogenous persistence of abilities, (θ, ρ_a), govern intergenerational persistence of wealth and earnings. Empirical estimates of the intergenerational persistence of earnings and income have been studied by Solon (1992); Chetty et al. (2014),

Target Moments	*Source*	*Data*	*Model*
IG persistence of earnings	Chetty et al. (2014)	0.34	0.35
IG persistence of college attendance	NLSY97	0.73	0.75
IG transfer share of net worth	Gale and Scholz (1994)	0.30	0.31
Mean earnings* (normalized)	PSID, Figure 2 Table 8	1.00	1.00
College premium*		1.57	1.58
Residual log earnings st. dev.*		0.57	0.59
Time investment in children*	CDS, Figure 17(a) Table 8	0.11	0.11

*Averages across all ages. See Figures 2 and 17 for age profiles.

Table 6: Target Moments

among others. We follow the latter and target a rank-rank slope of 0.34 when regressing children's lifetime average earnings on parents'. Consistently with that paper, we also find that there is no significant difference between the rank correlation and the intergenerational elasticity of earnings.

Intergenerational transfers as a share of an economy's total net worth are usually estimated from the SCF, by transforming transfer flows into a stock of "transfer wealth" under a steady state assumption (Kotlikoff and Summers, 1981). Although the exact estimate varies depending on the assumed demographic structure, Gale and Scholz (1994); Brown and Weisbenner (2004) obtain estimates in the range of 30% in the 1986 and 1998 SCF, respectively. While their transfers combine inter-vivos transfers and bequests, our model only includes a once-and-for-all transfer, s_4. So we choose the ratio of s_4 over total savings as the simulated moment.[41]

The mean and variance of abilities, (μ_a, σ_a^2), along with the Ben-Porath parameter b, govern the distribution of age-earnings profiles. These are calibrated to 21 moments from the PSID: the mean age-earnings profiles, the life-cycle profile of the college premium, and the residual earnings variance from age 24 to 65, which we compute from Figure 2 by averaging over 6 year brackets.

In Section 3.3, we estimated that the skill formation technology (5) is linear in the child's first period of life and Cobb-Douglas across periods. The remaining parameters,

[41]Specifically, the transfer wealth-net worth ratio $\int s_4 d\Phi / \int \sum_{j=5}^{12} s_j d\Phi = 0.3$

Parameter	*Value*	*Description*
θ	0.32	Parental altruism
ρ_a	0.23	Persistence of learning abilities
μ_a	0.83	Mean of learning abilities
σ_a	0.30	Variance of learning abilities
b	0.81	Ben-Porath HC accumulation
ζ	3.11	Child to adult human capital anchor
ω_1	0.56	Productivity of investment in children, primary
ω_2	0.30	Productivity of investment in children, secondary
ψ_1	0.23	Preference for children going to college, high school parents
ψ_2	0.24	Preference for children going to college, college parents

Table 7: Calibrated Parameters

$(\zeta, \omega_1 \omega_2)$, are calibrated to the means of parents' fraction of (active) time spent with children in Figure 17(a), averaged over 6 year intervals.[42] A parent with children aged 0-5 spends between 20-30 weekly hours actively, corresponding to about 16% of their time a week as shown in Table 2. This declines to about 8% when the child becomes 12-17.

The taste for college parameters, (ψ_1, ψ_2), govern how many children go to college in aggregate and how much it differs by parents' education status. Hence we target $L_1 = 0.45$, the fraction of individuals who attain at least 1 year of college education in the PSID, and its persistence. For the latter, we refer to the NLSY97, according to which about 73% of high school graduates whose father had some college education or more was enrolled for at least one year in post-secondary education.[43] Note that $\psi_1 \approx \psi_2$, meaning that tastes for college need only differ slightly to generate the degree of college persistence observed in the data. So college in our model will be mainly due to selection, and we will see later that college plays little role in explaining life-cycle inequality.

Let M_s denote the vector of the 27 empirical moments. We find the point estimate $\hat{\Theta}$

[42]Note that ω_0 is not separately identified from these three parameters. While the ω_{j}''s capture the productivity of each stage's investment, ζ boosts the productivity in all periods $j' = 0, 1, 2$.

[43]This number is the same whether we condition on biological or residential fathers, but slightly lower at 69% for both biological and residential mothers. Since our earnings data is from heads of households in the PSID, of which more than 90% are the husband, we chose the number conditional on fathers. The average enrollment rate in NLSY97 is approximately 49%, slightly higher than the 45% in the PSID.

Period	Mean Earnings		Col. Premia		SD($\log e_j$)*		Time wt Child**	
j	Data	Model	Data	Model	Data	Model	Data	Model
4	0.59	0.23	1.15	1.30	0.47	0.49		
5	0.75	0.46	1.30	1.56	0.51	0.44	0.16	0.15
6	0.86	0.59	1.48	1.66	0.53	0.44	0.10	0.10
7	0.93	0.69	1.63	1.69	0.56	0.54	0.08	0.09
8	0.99	0.85	1.69	1.74	0.59	0.72		
9	1.00	1.00	1.87	1.95	0.65	0.78		
10	0.89	0.99	2.09	1.86	0.69	0.76		

* Residual standard deviation after controlling for college; in the data, also for time effects
** Average of active time spent with children by both parents

Table 8: Data and Calibrated Moments by Age
Mean earnings in period 9 (ages 54-60) are normalized to 1, both in the data and model. In the model, children of adults in periods 5-7 are in periods 0-2, or age 0 to 17.

by numerically solving

$$\hat{\Theta} = \arg\min_{\Theta} \left[M(\Theta) - M_s\right]' \left[M(\Theta) - M_s\right], \tag{26}$$

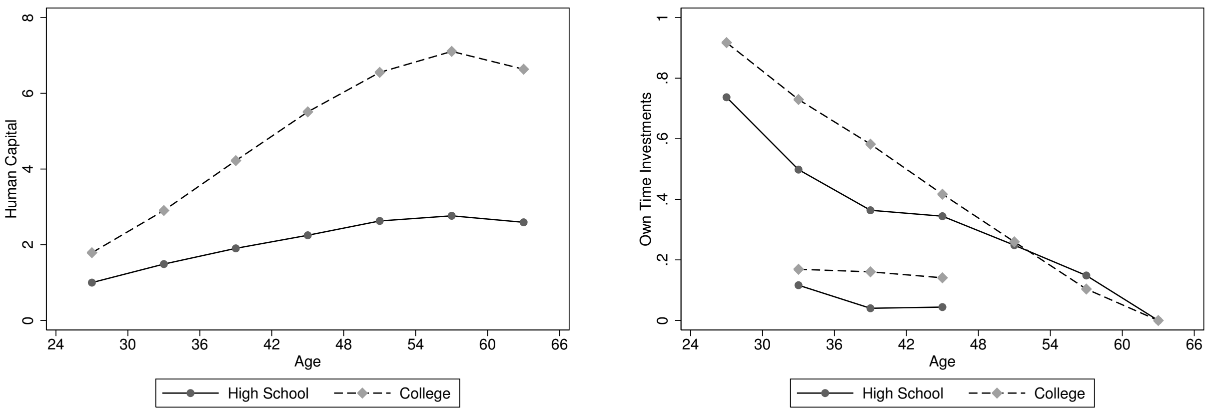
where $M(\Theta)$ are the simulated model moments.[44] Note that this is a large nested fixed point problem, since for every Θ we must also satisfy the two equilibrium and one budget balance conditions in Definition 1. Furthermore, taxation, subsidies and college costs are parameterized as fractions of mean earnings, so there is an additional fixed point in which we must ensure that $\bar{e}$ equals 1. Numerical details are given in Appendix A.

4.3 Model fit

Performance of the model can be seen in Tables 6 and 8. We attain a near exact fit for all the average moments, in particular the three intergenerational moments, and also the time profile of parents' time spent with children. The age profiles of the college earnings premia and log-earnings variances are close to the data, which one might expect given the similarity of the adulthood part of the model to Heckman et al. (1998); Huggett et al. (2011). However, the mean earnings profile is steeper than in the data.

While the PSID earnings profiles are estimated from adults who work, in reality many

[44] We weight all average moments in Table 6 equally, but give Table 8 smaller weights.



(a) Human Capital

(b) Time Investments

Figure 10: Human Capital and Time Investments
In the second plot, the longer lines lying above are investments in adults' own human capital, and the shorter lines lying below are investments in children.

children are taken care of more intensively by a parent who does not work. But in our model, the single parent needs to both work and spend time with the child. This makes his measured earnings when young lower than in the data, when he spends more time with the child than at work. In related vein, the model implies slightly lower inequality in periods 5-7, and higher inequality later in life. Since all parents spend more time with children and less at work, inequality is suppressed when young; as adults make up for the lost time by working more later in life, inequality becomes larger.

This effect can also be seen in the age profiles of adulthood human capital and own time investments, in Figure 10. The shapes of these profiles are not unusual, but own time investments, especially in periods 5-7, are somewhat lower than in standard models, especially for high school. Adults invest more in themselves even as they invest in their children. Because they know their working time is reduced when young, they invest more in themselves when young to reap higher earnings when they are older and children have left the household. So both because human capital grows more rapidly and also because they work less when young, the age profile is steeper than in the data.

Nonetheless, we match the mean and variance of lifetime earnings and time investments in children almost exactly, giving us confidence in how we formulated the connection between children's human capital $\tilde{h}$ to adulthood human capital h in (5). Morever,

Variance explained by (%):	(1) (S, a, h_4, s_4)	(2) (a, h_4, s_4)	(3) (S, h_4, s_4)	(4) (S, a, s_4)	(5) (S, a, h_4)
LFE	74%	74%	50%	52%	74%
LFW	73%	73%	57%	56%	67%

Table 9: Variance conditional on individual state at age 24-29
LFE and LFW stand for, respectively, lifetime earnings and lifetime wealth, defined in footnote 45.

most of our subsequent analysis will be based on lifetime earnings, so the age-earnings profile will play only a minor role.

5. Sources of Inequality

There are two important departures our paper makes from earlier works: initial conditions themselves are a product of investments from parents at earlier ages, and subsequent life-cycle earnings vary not only because of shocks to one's own human capital but also through future decisions made based on their (unborn) children's learning abilities.

5.1 Adulthood Inequality

Similarly as in Huggett et al. (2011), we first decompose how much of lifetime outcomes can be explained by differences at age 24. The individual states at this age are (S, a, h_4, s_4): whether or not an individual went to college, one's learning ability and human capital accumulated up to this age, and wealth transfers received from one's now elderly parent. The outcome variables we consider are lifetime earnings (LFE), defined as the present-discounted sum of earnings at all ages up to retirement, and lifetime wealth (LFW), which is simply lifetime earnings plus the initial transfer received from the parent.[45]

To compute the contribution of initial conditions, we first divide individuals into three equally sized groups, separately for each state (a, h, s). We then compute the fraction of lifetime earnings and wealth variance that can be attributed to various combinations of these initial conditions by computing conditional variances.

[45]That is, $LFE = \sum_{j=4}^{10} e_j/(1+r)^{j-4}$ and $LFW = LFE + s_4$.

Variable	Mean	Variance	Corr. wt $\log a$	$\log h_4$
$\log a$	-0.23	0.09	1.00	
$\log h_4$	-0.11	0.42	0.36	1.00
$\log s_4^*$	-0.73	1.56	-0.09	0.15
$\log s_4 > 0$	-0.72	1.34	-0.09	0.17

Table 10: Moments of initial distribution at age 24

$\log s_4^*$ is computed by adding machine zero to s_4 to avoid taking logs of zeros. The row $\log s_4 > 0$ shows the statistics when excluding zeros. Less than 0.02% of parents make zero transfers.

The variance of lifetime earnings and wealth explained by initial conditions are sizable in our model, at 73-74% (column 1). So despite life-cycle uncertainty (the human capital shocks and future investment in children), a large portion of individuals' lifetime outcomes can be explained by initial conditions when they become independent.[46] But conditional variances barely change when leaving out college (column 2), indicating the college choice margin can be explained almost entirely by the other variables. Indeed, the rank-rank correlation between h_3 and h_4, the level of human capital before individuals make the college choice, is 0.9, implying college more or less reflects selection. It is the learning abilities a and human capital at age 24, h_4, that play a large role: Without them, the contribution of initial conditions falls by 22-24 percentage points (columns 3-4).

What is somewhat surprising from Table 9 is that initial wealth, which is the once-and-for-all transfer the individual receives from his parent, does little to explain lifetime earnings inequality but does affect lifetime wealth inequality (column 5). Since $LFW = LFE + s_4$, this implies a disconnect between s_4, the financial transfer received from the parent, and (a, h_4), which is the main determinant of lifetime earnings inequality. To see this, it is useful to consider the distribution of (a, h_4, s_4). A major departure of our framework from the previous literature is that the distribution of this vector is fully endogenous and determined by intergenerational linkages, of which the mean, variance and correlation structure we present in Table 10.

[46]This number is close to Huggett et al. (2011), who compute this number as ranging from 61-67%. In Keane and Wolpin (1997), as much as 91% of lifetime utility differences are explained by initial conditions. As they note, this does not mean that inequality is exogenously predetermined, since we assume forward looking, individually rational individuals.

Variable	Compared to group:	Change in LFE percentile
College S	-	-6.38
Learning ability a	Low	-18.62
	High	24.98
Human capital h_4	Low	-20.35
	High	23.41
Transfers s_4	Low	-1.35
	High	-2.15

Table 11: Average Lifetime Earnings Differences Across Groups
The first row shows the lifetime earnings rank difference between college- and high school-educated individuals, holding all else equal. The following rows show the average differences between the high and low groups compared to the medium group, holding all else equal.

We note that the variance of learning abilities and human capital at age 24 is much larger, and their correlation smaller, than in Huggett et al. (2011), in which the initial distribution is exogenously calibrated. This is due to our inclusion of childhood human capital formation. A large variation in h_4 is required to explain lifetime inequality, but in order to arrive at this level in the first place the model requires enough variation in learning abilities, which is the only exogenous source of heterogeneity before labor market entry. This also results in a higher correlation between ability and h_4.

Interestingly, s_4 is negatively correlated with learning abilities, and only weakly positively correlated with h_4. This suggests a compensatory mechanism that we investigate further in the next section: parents with high learning ability children invest more in their human capital and transfer less financial assets, especially among poorer parents.

Table 9, while useful, does not tell us the exact contribution of each state variable at age 24 since they are intercorrelated. In particular, learning abilities seem to play a large role, but some of it is because the other variables are outcomes of investments made by parents who take into consideration their children's learning ability. To analyze the importance of each separately, we first compute the lifetime earnings of high school individuals ($S = 0$) whose (a, h_4, s_4) are in the median group. We then compute the lifetime earnings percentile difference between this group and other groups which only differ in terms of one state variable. The results are shown in Table 11.

The college effect is small and negative: Holding all else equal, going to college moves individuals along the lifetime earnings percentile down by 6.38 percentage points. In our model, the college premium is generated by positive selection on tastes for college, so the controlled effect is negative due to the opportunity cost of attendance. Both learning abilities and human capital play a large role. Having higher ability or human capital moves individuals up the lifetime earnings rankings by more than 20 percentage points. The effect of having lower ability or human capital are similar, moving individuals down the percentile rank by 18-23 percentage points. Financial transfers have a small but non-monotonic effect: both smaller and larger transfers reduce lifetime earnings. This is not due to parental effects, but because children themselves later invest in their own children (the grandchildren). Individuals with large assets invest more in their children than themselves, increasing their children's human capital at the expense of their own. This is further analyzed in the next section.

5.2 Children's outcomes

Given the important of the distribution at age 24, we now decompose how much is due to an individual's own learning ability or due to one's *parental background*. We follow a similar exercise as above. We take a parent's state at age 30 (when the child is born) as the initial condition, and decompose how much of the child's ability can explain *the child's* outcomes: his lifetime earnings and wealth, and initial human capital and assets (h'_4, s'_4). The results are summarized in Table 12.

By construction, the child's (S', h'_4, s'_4) is a function of the parents (S, a, h_5, s_5) and the child's ability a'. So the contribution of the parent's state at age 30 (when the child is born) relative to the child's initial conditions is only imperfect due to the shocks the parent receives while raising the child. Just like an individual's human capital shocks later in life has little explanatory power for lifetime inequality, it also turns out that the parents' shocks have little explanatory power for children's outcomes as well. As seen in column (1) of Table 12, parents' states at age 30 explain virtually as much as the the children's own states when they become independent.

Variance explained by %	(1) $(S, a, h_5, s_5; a')$	(2) (S, a, h_4, s_4)	(3) $(S, a, h_4, s_4)_g$
children LFE'	71%	22%	19%
children LFW'	72%	49%	34%
children h'_4	63%	55%	48%
children s'_4	65%	58%	20%

Table 12: Variance conditional on intergenerational states
LFE' and LFW' stand for, respectively, the children's future lifetime earnings and lifetime wealth, defined in footnote 45. (h'_4, s'_4) are the children's initial endogenous states when they become age 24-29 (become independent). Column 1 conditions on the parents' states at age 30-35; Children are aged 0-5 at this stage. Columns 2-3 conditions on the parents' and grandparents' states, respectively, when they are aged 24-29.

So parental states at age 30 explain most of their children's outcomes, almost as much as the children's initial conditions at age 24. Our model allows us to extend this further to even before the child is born: we can analyze how much of children's outcomes are affected by parental states at age 24, and even by the *grandparents'* states at age 24. These results are summarized in columns 2-3 of Table 12.

Parents' states have stronger explanatory power for children's lifetime wealth than earnings differences. Even before a child is born (when parents are age 24), parents' states can explain 49% of children's lifetime wealth, but only 22% of their earnings (column 2). Young parents with high ability children are unable to invest enough in their human capital due to life-cycle borrowing constraints, but these same children can quickly accumulate human capital as an adult. This makes earnings depend less on parents. In contrast, children's lifetime wealth are more affected by parents because s'_4 is decided later in life when life-cycle constraints matter less.

The reason parents' states have such large explanatory power can be see in the third and fourth rows. Parents' states before the child is born explains a large amount of children's eventual human capital and asset levels at age 24, meaning that parents with higher states both invest more in their children and also transfer more assets. Of course, this is because the model postulates that the investment parents make in children, both in terms of human capital and financial assets, are not only functions of the children's states but the parents' states.

The grandparents' contribution to explaining grand-children's lifetime earnings and

		Parent's a		
		Low	Med.	High
Child's a':	Low	0.05	0.11	0.17
	Med.	0.03	0.07	0.15
	High	0.03	0.05	0.11

Table 13: Fraction of parent's lifetime wealth transferred to child
For each group, we compute $s_4'/(1+r)^5 LFW$, where s_4' is discounted to parent's age 24 and LFW refers to the parent's lifetime wealth, also measured at age 24.

age 24 human capital inequality is still high compared to the parents'. This is due to the assumed mean reversion in learning abilities. Although uncertainty of future offspring's learning abilities matter from one generation to the next, this effect is washed out within two generations. On the other hand, because financial transfers are used to compensate for ability differences, without knowledge of the learning abilities of the next two generations, the grandparent generation's conditions when young lose explanatory power for wealth differences (through s_4) for the grandchild generation.

To see this more clearly, in Table 13 we tabulate how much of parents' lifetime wealth are transferred to the next generation by nine groups of families, divided by whether the parent and child's learning ability is low, medium or high. Clearly, high ability parents transfer more of their wealth to their children. However, note that they also pass down less to high ability children, in anticipation of their higher earnings later in life. That is, there is a compensation mechanism in place, because childhood investment and transfer decisions are made *after* observing their children's abilities. When a parent realizes that a child's ability is higher (lower) than expected, she can compensate for this by leaving less (more) assets and investing more (less) in his education.

6. Comparison to Becker-Tomes

To illustrate the importance of our multi-period model, we compare it against the following simple 2-generation model. A parent has at his disposable $(1-\tau)h + s$, where h corresponds to his lifetime earnings and s a transfer he received from a grandparent. The parameter τ is a flat tax rate that we will calibrate to equal the average tax rate in the benchmark model. He must allocate his net resources to his own consumption c, investment in his child's human capital x, and financial transfers s':

$$\max_{c,c',x,s'} \{u(c) + \bar{\theta}u'(c')\} \quad \text{subject to} \tag{27a}$$

$$c + \frac{s'}{(1+r)^5} = (1-\tau)(h-x) + s, \quad c' = (1-\tau)h' + s', \tag{27b}$$

$$h_k = \bar{\zeta}a'(x+\bar{d})^{\bar{\gamma}}, \quad s' \geq 0 \tag{27c}$$

where $u(\cdot)$ is CRRA, c' is the consumption of the child, and $\bar{\theta}$ the degree of altruism. Children's human capital, h', is produced through (27c), with productivity $\bar{\zeta}$ and decreasing returns, $\bar{\gamma} < 1$. The learning ability of the child, a', is heterogeneous across the population and correlated with a.[47] Investment in children is subsidized by a lump-sum government transfer $\bar{d}$. We will assume that the transfer is taken as given by the parents, but that the government ensures that it equals a fraction π_d of their average earnings. When the child grows up, he can consume $(1-\tau)h'$, which captures his lifetime earnings, plus any financial transfers received from his parents that accrue interest, $r_{BT} = (1+r)^5 - 1$. Intergenerational transfers are subject to a non-negativity constraint.

This is a standard version of the BT model ubiquitously employed in the literature. Some features of this model are:[48]

BT1: When the constraint $s \geq 0$ is not binding, the optimal choice for x is to equalize the returns to investment to the gross interest rate $1 + r_{BT}$.

BT2: If no households are subject to the borrowing constraint, the stationary IGE of earnings is equal to the persistence of abilities, ρ_a.[49]

BT3: If all households are subject to the borrowing constraint, the stationary IGE of earnings is equal to $\dfrac{\rho_a + \bar{\gamma}}{1 + \rho_a\bar{\gamma}}$ if $u(c) = \log c$.

[47] In this simple setup, the wage per unit of human capital is normalized to 1, or can be assumed to be subsumed in $\bar{\zeta}$. To make the BT model as similar to our benchmark model as possible, we deduct education expenses x from taxable earnings.

[48] See Appendix B for derivations.

[49] Under stationarity, the IGE=IGC.

	Benchmark	BT
γ	0.81	0.80
ζ	3.11	4.41
θ	0.32	0.39

Table 14: Calibrated Parameters for BT model

BT stands for the version of Becker and Tomes (1986) employed in the text. In the text, we differentiated the parameters by putting bars over them in the BT model.

BT2-BT3 imply that if the economic behavior of households is to add anything to a mechanical approach in terms of explaining intergenerational persistence, the intergenerational borrowing constraint must play a large role. However, Mulligan (1999) finds that the IGE's of constrained and unconstrained households barely differ, casting doubt on the relevance of intergenerational human capital investments.

Our main innovations upon BT is to feature life-cycle human capital accumulation in adulthood, and multi-period investments in children using a dynamically complementary technology. In particular, the latter creates a role for borrowing constraints faced by young parents to alter child outcomes. Thus, we are able to downplay the significance of the intergenerational borrowing constraint faced by old parents, which was the only market incompleteness in BT. Consequently, our model is able to remain empirically consistent while still attributing a significant role to economic mechanisms in terms of explaining intergenerational persistence.

6.1 Implied IGE's

To operationalize the comparison, we first fix the distribution of parents' (h, s) to equal the distribution of parents earnings and savings at age 24 in the benchmark model. When doing so, we assume that h is the present discounted sum of all future earnings, but before investment in children. Then we compute the one-generation ahead decisions of the BT model. We will call this the "short-run" model. From the short-run model, we calibrate the returns and productivity of investments $(\bar{\gamma}, \bar{\zeta})$, and altruism $(\bar{\theta})$, so that the average level of investments, children's lifetime earnings and intergenerational transfers received in the BT model is equal to their levels in our benchmark model. The calibrated

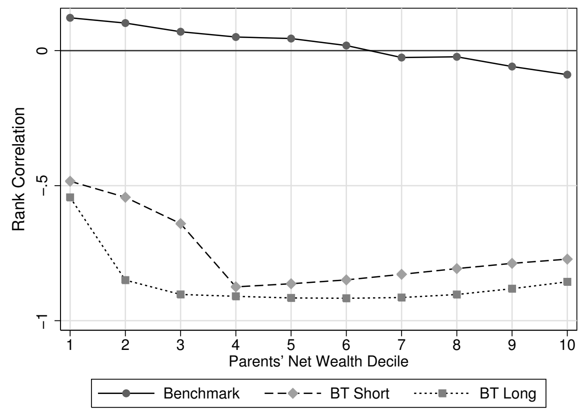


Figure 11: Rank Correlation between Children's Earnings and Transfers Received
For the benchmark model, the figure plots the rank correlation between average earnings and s_4. Parents' net wealth in the benchmark model are defined as the present value of discounted income, net of tax and transfers, plus their wealth in period 4 (age 24). For the BT model, the figure plots the rank correlation between (h', s'), and parents' net wealth is $(1-\tau)h+s$. "BT Short" and "BT Long" are the one-generation ahead and steady state outcomes. See text for description of the BT model.

parameters are tabulated in Table 14. All other parameters are held fixed, in particular the intergenerational process for learning abilities, (3).

We then compute the steady state of the BT model, assuming that each successive generation faces the same decision problem but with an updated distribution of (h, s). We will call this the "long-run" model. The distributions of (h', s'), and the resulting IGE, in both the short- and long-run BT models are compared against our model.[50]

According to BT1, in the BT model, unconstrained parents invest in their children up to point where the returns equal the interest rate. So among parents with a similar amount of resources, there will be a compensation mechanism: those with high a' children will invest more in children's human capital (h') and transfer less (s'), and vice versa for those with low a' children. Hence, if more parents are unconstrained, the more negative will be the conditional correlation between (h', s') in the children's generation.

As we see in Figure 11, when families are split into 10 groups according to the level

[50]Since we are comparing a single human capital outcome of the child (h') in the BT model to the lifetime earnings of children, it may be of concern that we are comparing different objects. However, we have already seen that lifetime inequality is determined early on, so that earnings ranks do not change much over the life-cycle. Appendix Figure 19 shows that the rank correlation between children's lifetime earnings and their human capital, $h'_{j'}$, is high and stable for all j', at around 0.85. Consequently, the rank correlation of h'_j with parents' lifetime earnings is also quite stable throughout their lifetimes, at around 0.5.

	Benchmark	BT Short	BT Long
Corr(LFE, s_4) or Corr(h, s)	0.05	- 0.37	-0.67
IGE of earnings	0.34	0.37	0.23
IGE of wealth	0.41	0.34	0.62
Pr(IG transfer$\leq \bar{e}$)	0.69	0.44	0.34
IG transfer$\leq \bar{e}$	0.52	0.67	0.37
IG transfer$> \bar{e}$	0.60	0.20	0.18

Table 15: Comparison across Benchmark and BT models
"BT Short" and "BT Long" are the one-generation ahead and steady state outcomes. See text for description of the BT model. In the top panel, the first row shows the rank correlation between children's earnings and transfers received. The IGE of wealth is conditional on both the parent and child having positive wealth. The bottom panel shows the IGE of lifetime earnings for households in which parents make transfers below the level of average (annualized) earnings, $\bar{e}$, in each model. Wealth in the benchmark model is measured as $\sum_{j=4}^{12} s_j$. The last row shows the model-implied fraction of such households.

of parents' net wealth, both the short- and long-run BT models result in a negative rank correlation of around -0.7 across net wealth deciles. But in the benchmark model, the rank correlation between children's average earnings and transfers, s_4, is close to zero. As shown in the first row of Table 15, these patterns also bear out in aggregate. The correlation between earnings and transfers is -0.40 in the BT model, but 0.05 in our model. While data on intergenerational transfers are limited to inheritance data, which is very noisy, the PSID and HRS suggest the negative correlation by the BT model is unlikely.[51] Note that it is always the children of poor parents who have the largest correlation, meaning they are the least likely to achieve efficient investment in children.

In the short-run BT model, this still results in a realistic level of an IGE, as shown in the second row of Table 15. But in the long-run model, as many of the poor households escape the constraint (lower deciles in Figure 11), the IGE falls to an unrealistic level of 0.23. According to BT2 and BT3, this means that the borrowing constraint barely matters in aggregate, and the IGE more or less follows ability persistence ($\rho_a = 0.23$). As shown in the fourth row of Table 15, may households make larger transfers in the long-run.[52]

[51] Available evidence shows a highly skewed inheritance distribution with a large fraction of families leaving very little bequests Hurd and Smith (2003); De Nardi and Yang (2014).

[52] Mulligan (1999) reports only about 12% of households expecting bequests above $25,000 in 1982 dollars, which is around the mean earnings in his PSID sample. In the HRS, 43% of respondents respond affirmatively to the question on whether they "expect to leave a sizable inheritance." In comparison, McGarry (1999) documents detailed information on intergenerational transfers using data from the HRS and

So in our model, the IGE of earnings does not simply reflect the correlation of abilities (Goldberger, 1989), nor is much explained by the intergenerational borrowing constraint (Becker and Tomes, 1986). Rather, it is the life-cycle borrowing constraints faced by parents, coupled with dynamic complementarities across multiple periods of invsetment in chidlren, that generates a sizeable part of the IGE of earnings (Heckman and Mosso, 2014). In similar vein, the IGE of wealth is at an unrealistically high level of 0.62 in the BT model, since more parents transfer wealth to their children rather than educating them. In our model it is 0.41, close to the the empirical estimate of 0.37 in Charles and Hurst (2003).[53]

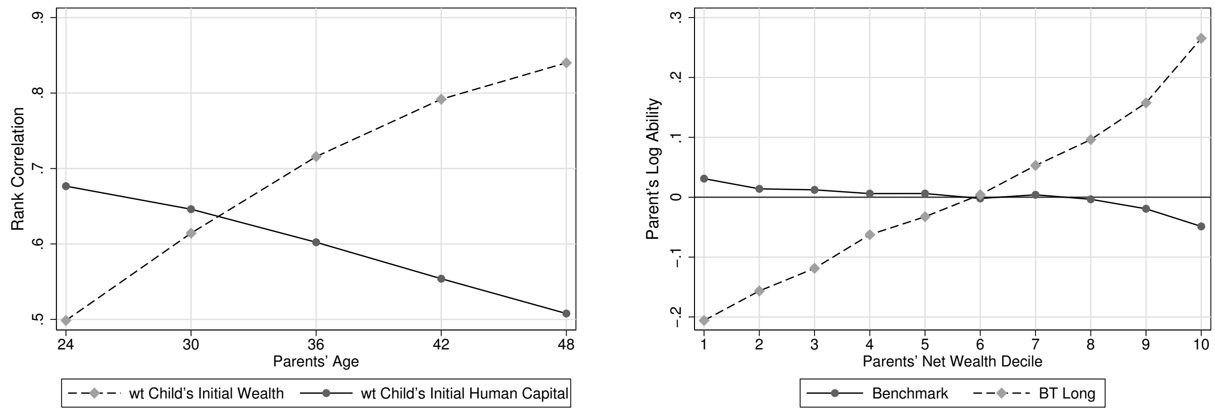
6.2 Constrained vs. Unconstrained Households

In the last two rows of Table 15, we split households according to whether parents leave transfers above or below the mean wage, similarly as in Mulligan (1999). As implied by BT2-BT3, the IGE is higher among constrained households for both BT models, although in the long-run, the difference becomes less prominent as some poor households escape the constraint. However, Mulligan (1999) finds no evidence for such a difference in the PSID sample, concluding the BT model cannot matter much for the IGE. In fact, he finds suggestive evidence that the IGE is higher among unconstrained households.

In our model, the IGE is in fact similar between the two groups, and slightly larger for unconstrained households. With dynamic complementarity in childhood investments and parents facing life-cycle borrowing constraints, whether or not a parent is bound by the intergenerational borrowing constraint says little about the IGE. First, when young parents realize they cannot invest in their children, they instead increase their own human capital to leave more transfers later. Conversely, this means that young parents who invest more in their children invest less in themselves. This is shown in Figure 12(a): the rank correlation between parents' h_j and the transfer they leave, s'_4, increases with their age j, while the correlation with children's initial human capital, h_4, declines.

AHEAD. The AHEAD survey elicits the subjective probability of leaving *any* bequests, of which the sample average is 0.55.

[53]While their estimate is based on children's wealth before receiving bequests and hence not directly comparable, it is close to our benchmark model where we measure wealth as the average level of savings (s_j) across all periods $j = 5, \dots, 10$. Hurd and Smith (2003) also present evidence that much of wealth is not bequeathed, so a large level of intergenerational persistence is unlikely.



(a) Parents' Trade-off

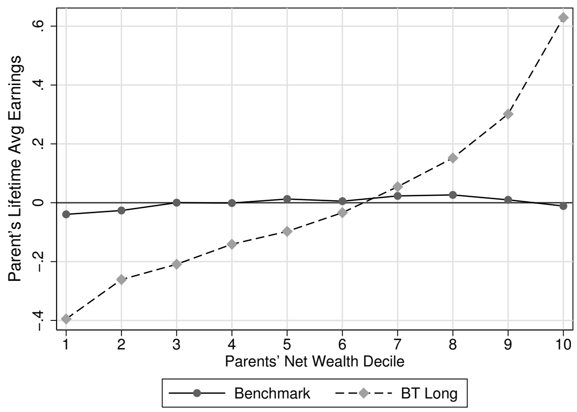
(b) Average (Log) Abilities

Figure 12: Parents' Intergenerational Transfer Decisions

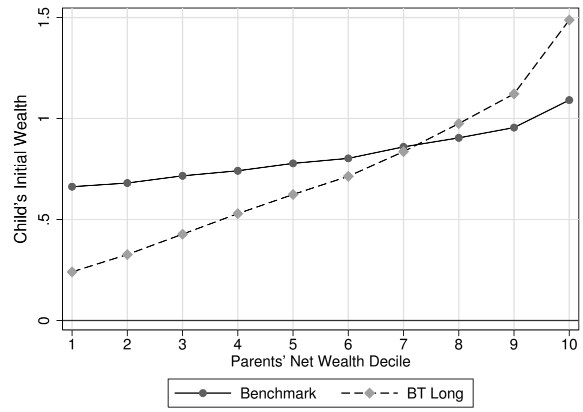
For each period j, the Figure (a) plots the rank correlation between parents' h_j and children's initial human capital and assets, (h_4, s_4). Figure (b) plots average of intergenerational transfers by parents' net wealth decile. Parents' net wealth in the benchmark model are defined as the present value of discounted income, net of tax and transfers, plus their wealth in period 4 (age 24). For the BT model, parents' net wealth is $(1-\tau)h + s$. "BT Long" is steady state outcomes. See text for description of the BT model.

Second, this dissociates children's earnings and wealth, which was also seen in Table 10 in the previous section. That is, there are children with high earnings but low wealth, and vice versa, generating much more heterogeneity in the cross-sectional distribution of earnings and wealth. In our steady state, there are many parents with high net wealth but low ability, and low net wealth but high ability, as shown in Figure 12(b). Consequently, many high net wealth parents are wealthy not because of high earnings but high wealth passed down from previous generations, and vice versa, as shown in Figure 13(a). Although it is still the case that high net wealth parents transfer more wealth to their children, the association is much looser than in the BT model, as shown in Figure 13(b). In contrast, the BT model results in strong sorting in the steady state, so all wealth-rich parents are also earnings-rich parents.

In our model, wealth-rich parents are much more similar to each other than in the standard BT model. Consequently, the IGE's also look more similar across different groups. So rather than concluding that borrowing constraints don't matter and that abilities and/or preferences may be more important for explaining intergenerational persistence, our results point toward the importance of borrowing constraints that matter *earlier in life* cou-



(a) Average (Log) Lifetime Earnings



(b) Average Transfers

Figure 13: Parents' Lifetime Earnings and Transfers

In Figure (a), the average of log lifetime earnings are normalized to mean 0. In Figure (b), the y-axis is in multiples of the average annual earnings. Parents' net wealth in the benchmark model are defined as the present value of discounted income, net of tax and transfers, plus their wealth in period 4 (age 24). For the BT model, parents' net wealth is $(1-\tau)h+s$. "BT Long" is steady state outcomes. See text for description of the BT model.

pled with dynamic complementarity in childhood investments.

7. Counterfactuals

Given the importance of borrowing constraints for childhood investments and intergenerational persistence, we focus on 7 counterfactuals. The first three involve relaxing the borrowing constraints faced by parents, and the rest are policy experiments with respect to taxation and education subsidies.

7.1 Market Incompleteness

Parents in our model face two types of borrowing constraints: the life-cycle constraint (8), which applies to all periods in life, and the intergenerational constraint (11), which prevents parents from borrowing against their children's future income. In particular, our previous results suggest that life-cycle constraints do not explain much of life-cycle inequality, but is important because it affects how parents invest in children. And since investment in children are completed by the time they make their intergenerational trans-

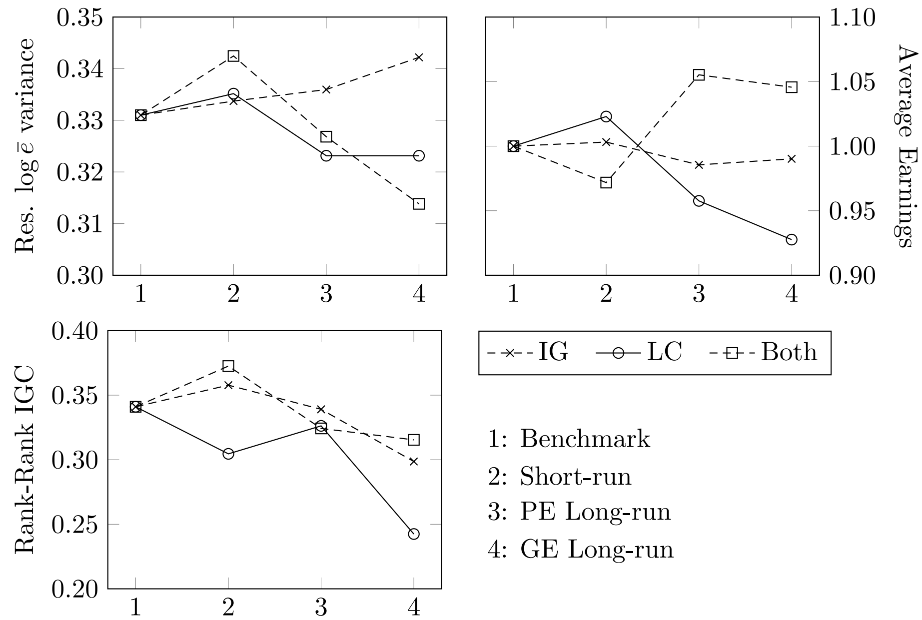


Figure 14: Relaxing Borrowing Constraints

Three policy experiments: relaxing the intergenerational borrowing constraint, the life-cycle borrowing constraint, and both. The short-run result is the IGC and average earnings one generation after implementing the policy change. "PE Long-run" is when the economy reaches a new steady state, but prices (interest rate and wages) are still fixed at their initial steady state levels. "GE Long-run" is the new steady state equilibrium.

fer decisions, the intergenerational constraint had less bite.

To investigate these channels, we first compute a counterfactual in which we relax the intergenerational borrowing constraint to $-g/(1+r)$, so that it is equal to the life-cycle borrowing constraint. Then, we instead relax the life-cycle borrowing constraint to $-2g/(1+r)$.[54] Lastly, we relax both constraints.

For each of these counterfactual scenarios, we focus on three measures of children's outcomes: the variance of (residual) log earnings, average level of earnings, and the resulting IGC. In Figure 14, we show the results under different timing assumptions:[55]

1. a short-run PE: a one-generation ahead transition, with no change in prices. Parents start period 4 at the same states as in the benchmark stationary equilibrium, but now face the new borrowing constraints.

2. a long-run PE: the stationary distribution, but still with no change in prices

[54]When doing so, we must also double the lump-sum subsidy to guarantee that agents can repay their debt in all realized states.

[55]The results are also tabulated in Appendix Table 16.

3. a long-run GE: the new stationary equilibrium with new prices.

As expected, relaxing the intergenerational constraint alone has only small effects. Inequality rises as measured by the variance of (residual) log earnings, while intergenerational persistence falls in the long-run. There is barely any change to average earnings. Relaxing the life-cycle constraint has qualitatively different effects from relaxing the intergenerational one. Inequality rises in the short-run, while mobility increases. But in the long-run, inequality also drops while persistence drops even further.

In both cases, inequality rises because constrained parents tend to be those with higher ability. So investment in children becomes more efficient, and persistence drops. But when life-cycle constraints are relaxed, even poorer parents are able to take advantage of the dynamically complementary technology in the long-run, so earnings become more equal while persistence continues to drop. This is consistent with the previous subsection in which the borrowing constraints faced by young parents when they also need to invest in their children could explain a large fraction of intergenerational persistence in the long-run. But as parents allocate more resources toward their children and less to market labor, average earnings drop.

Qualitatively, it is unclear what would happen when both constraints are relaxed. The outcome is a purely quantitative one that depends on the calibrated parameters. It turns out that persistence is lower in the long-run when both constraints are relaxed simultaneously, but not as much when they are relaxed separately. But keep in mind that the earnings difference between high and low individuals are much smaller, since earnings become much more equal. Moreover, average earnings also rise, since instead of sacrificing market labor, parents are able to pull some of their children's future resources toward the children's education.

7.2 Government policies

In this subsection we implement 4 counterfactual policies: eliminating the progressivity of income taxation in (25), and focusing all education subsidies (d_0, d_1, d_2) into only one of the three periods of childhood. We label these policies "FT," "P0," "P1" and "P2," respec-

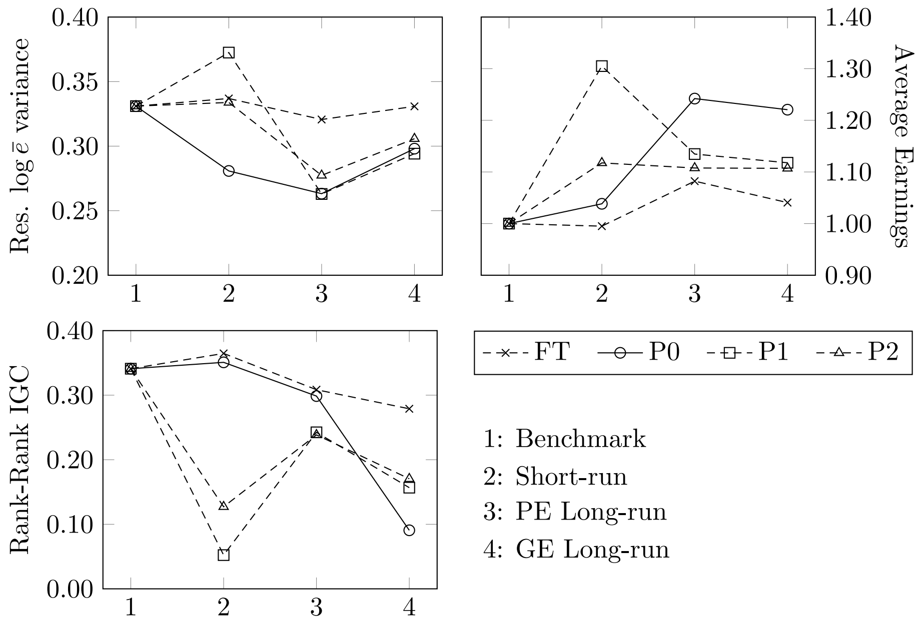


Figure 15: Policy Experiments

Four policy experiments: eliminating tax progressivity ("Flat Tax"), and giving all education subsidies only for children in their first, second or third periods of education (ages 0-5, 6-11, and 12-17, respectively, labeled "P0," "P1" and "P2"). The short-run result is the IGC and average earnings one generation after implementing the policy change. "PE Long-run" is when the economy reaches a new steady state, but prices (interest rate and wages) are still fixed at their initial steady state levels. "GE Long-run" is the new steady state equilibrium. Time with children is shown for each stage in childhood in the long-run steady state equilibrium.

tively. As in the previous section, for all 4 policies, we consider a short- and long-run PE, in addition to the long-run GE. The resulting variance of (residual) log earnings, average level of earnings, and IGC of the children's generation are depicted in Figure 15.[56]

First consider the flattening of the income tax schedule. We assume that the tax rate is equal the average tax rate in the benchmark calibration (about 24%). Of course, we are not carefully modeling top incomes, for whom tax progressivity matters the most. Nonetheless, our results show that while there is little change in the variance of earnings, there is a significant improvement in intergenerational mobility (the IGC drops from 0.37 to 0.29), associated with a small rise in average earnings (about 3 percentage points). The rise in earnings is expected, since individuals accumulate more human capital when less of their earnings are taxed away. Parents invest more in their children for similar reasons, which affects intergenerational persistence.

[56]The results are also tabulated in Appendix Table 16.

But flattening the tax schedule has a negligible effect compared to the education subsidies. While there is no change in the absolute amount of education subsidies across the benchmark calibration and the P0, P1, P2 counterfactuals, the effects are large and differ significantly in the short-run. In the long-run, all policies generate large drops in the IGC, while lowering inequality and increasing average earnings.

In the long-run, P0 lowers the IGC the most, to about 0.1, along with a 21 percentage point increase in average earnings. P1 lowers the IGC to about 0.15, while raising average earnings by about 11 percentage points. Since the long-run (residual) earnings variances are similar, there is a sense in which earlier subsidies improve both mobility and efficiency.

Interestingly, P0 is the only policy which has little short-run impact. Broadly speaking, focusing education subsidies shift private investment into the other periods, as seen in Figure 16.[57] But how much this matters for intergenerational persistence also depends on the distribution of parental states. In the short-run, high human capital parents benefit more from the earliest subsidy: Even though it is better to subsidize education early on under dynamic complementarity, low human capital parents cannot invest the additionally required amount in their children in later periods. In contrast, later subsidies serve as an "equalizer": The optimal levels of investment are lowered for all parents due to the absence of the early subsidy, and the later subsidies boost the human capital levels of children of low human capital parents.

But as the change in policy becomes anticipated (for the first generation of parents, it is unanticipated), the human capital of young parents become higher (that is, they enter adulthood with higher levels of human capital). This makes early investments even more important and later investment less important over time. In particular, as prices adjust, the larger supply of human capital reduces the stock of physical capital, driving up the equilibrium interest rate.[58] This makes it costlier to borrow against future income to invest in children, further raising the importance of early subsidies.

Education subsidies are attractive in the sense of lowering inequality and raising aver-

[57] As seen there, flat taxes barely shift parents' investment compared to the benchmark.

[58] Last column in Appendix Table 16.

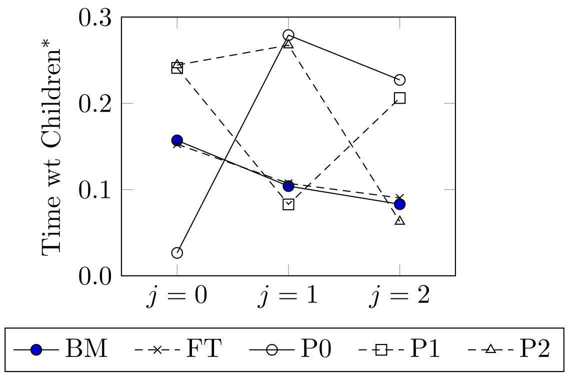


Figure 16: Policy Experiments, Time with Children
Four policy experiments: eliminating tax progressivity ("Flat Tax"), and giving all education subsidies only for children in their first, second or third periods of education (ages 0-5, 6-11, and 12-17, respectively, labeled "P0," "P1" and "P2"). The figure plots the mean time spent with children for each stage in childhood in the new steady state equilibrium.

age earnings. Early subsidies may not have a visible impact on mobility in the short-run, while later subsidies can. But in the long-run, early subsidies can have the largest effect on mobility, especially in general equilibrium: i) the earliest stage of investment is the most important when the childhood human capital accumulation displays dynamic complementarity, and ii) it is also when parents are the most financially constrained. Moreover, this is associated with a long-run rise in average earnings, as more human capital is accumulated in the economy.

8. Conclusion

In this paper, we present a rich, quantitative model that combines a life-cycle and intergenerational model of human capital. The model features multiple generations with altruistic dynasties. Parents make investments in their children's human capital and also decide on financial transfers. Consistent with prior work, we assume complementarity between early and later investments in children, and consider both time and goods investments. We also model the college enrollment decision, and model life-cycle wage growth via investment in one's own human capital. We cast this environment in an equilibrium setting with various government policies. Most importantly, individuals face life-cycle borrowing constraints, preventing them from achieving optimal investment in young children which are difficult to correct later in life.

We find that parental states can explain as much as half of children's lifetime wealth inequality and a quarter of lifetime earnings inequality. Dynamic complementarity in investments in children coupled with life-cycle borrowing constraints may account for as much as a third of intergenerational persistence, while intergenerational borrowing constraints matter less. Consequently, whether or not a parent transferred financial assets to their children contains little information on whether he achieved efficient investment in children. Lastly, we find evidence suggesting that early education subsidies are the most effective tool with which to reduce intergenerational persistence, while taxation plays only a small role. While more work needs to be done, our model has the potential to account for US-Europe differences in inequality and mobility by assigning a first-order role to policy differences.

Appendices

A. Numerical Solution Method

Numerical grids For each stage of the life-cycle, we set a square-grid on the continuous state variables $(h_j, \tilde{h}'_j, s_j)$. When solving for optimal policies, we linearly interpolate over next period's value functions. The AR(1) process for abilities are approximated using the Rouwenhorst method in Kopecky and Suen (2010), and i.i.d. grids for the luck shocks ϵ_j using the equal-mass approach in Kennan (2006).

Value functions Except for the first period of an individual's working life, in which there are 2 choice variables, all periods involve solving for 3 choice variables for each grid point on the state space. Furthermore, all of these choices involve bound constraints that are potentially binding. We optimize over the objective function in each period using a simplified version of Kim et al. (2010), which is a projected quasi-Newton method with subsequent BFGS updates modified to check for boundary constraints. Given a guess V_4^n, we can solve all for all value functions $V_9, V_8, \ldots$ by backward induction, obtaining a new guess V_4^{n+1} (V_{10} is deterministic and not subject to the dynastic continuation values). We iterate until $|V_4^{n+1} - V_4^n|$ falls below a specified tolerance criterion.

Equilibrium and SMM For a given guess of the parameter vector $\hat{\Theta}$, we obtain individual decision rules, the stationary distribution, and find (β, v, A) that matches $\bar{r} = 4\%$, the share of high school workers in (23), and a mean earnings of $\bar{e} = 1$. We obtain the stationary distribution via Monte-Carlo simulation by simulating $N = 120,000$ households for $T = 200$ generations using the optimal decision rules computed above. Given the simulated moments, we solve (26) using a Nelder-Meade downhill simplex routine. The choice of (N, T) is arbitrary, but increasing to $N = 240,000$, $T = 300$ had negligible effects.

Further details are available upon request.[59]

[59]Many compromises were made to make the numerical problem manageable. For example, we do

B. Proof of Claims BT1-BT3

BT1: The first order conditions for program (27) are:

$$
\begin{aligned}
s' : \quad & u'(c)/\bar{\theta}u'(c') \geq 1 + r_{BT} \qquad \text{with equality if } s > 0 \\
x : \quad & u'(c)/\bar{\theta}u'(c') = \bar{\gamma}\bar{\zeta}a'/(x+\bar{d})^{1-\bar{\gamma}}
\end{aligned}
\tag{28}
$$

hence when $s > 0$, the parent simply equates the marginal investment of x to the gross interest rate $1 + \bar{r}$, as claimed in BT1.

BT2: When the constraint does not bind, clearly

$$
\begin{aligned}
& x^* + \bar{d} = \left(\frac{\bar{\gamma}\bar{\zeta}a'}{1+r_{BT}}\right)^{\frac{1}{1-\bar{\gamma}}} \\
\Rightarrow \quad & \log h' = \frac{1}{1-\bar{\gamma}} \cdot \log\left(\bar{\zeta}a'\right) + \frac{\bar{\gamma}}{1-\bar{\gamma}} \cdot \log\left(\frac{\bar{\gamma}}{1+r_{BT}}\right).
\end{aligned}
$$

Hence if no households are constrained, the IGE of h' is simply equal to the IGE of a'. So under stationarity, the IGE of earnings, as measured by $\bar{w}h$, is ρ_a (IGE=IGC).

BT3: Conversely, if the constraint binds for all households, and utility is log, the optimal choice of x simply solves (28):

$$
\begin{aligned}
& \bar{\zeta}a'(x+\bar{d})^{\bar{\gamma}}/(h-x) = \bar{\gamma}\bar{\theta}\bar{\zeta}a'/(x+\bar{d})^{1-\bar{\gamma}} \\
\Rightarrow \quad & x^* = \frac{\bar{\gamma}\bar{\theta}h - \bar{d}}{1+\bar{\gamma}\bar{\theta}} = \frac{\left[\bar{\gamma}\bar{\theta} - \pi_d\right]h}{1+\bar{\gamma}\bar{\theta}} \\
\Rightarrow \quad & \log h' = \left[\log\bar{\zeta} + \bar{\gamma}\log\left(\frac{\bar{\theta}\bar{\gamma}(1+\pi_d)}{1+\bar{\theta}\bar{\gamma}}\right)\right] + \bar{\gamma}\log h + \log a'.
\end{aligned}
$$

So assuming stationarity and subtracting $\rho_a \log h$ from both sides yields

$$
\begin{aligned}
\log h' - \rho_a \log h =& (1-\rho_a)\left[\log\bar{\zeta} + \bar{\gamma}\log\left(\frac{\bar{\theta}\bar{\gamma}(1+\pi_d)}{1+\bar{\theta}\bar{\gamma}}\right) + \mu_a - \sigma_a^2/2\right] \\
& + \bar{\gamma}\log h - \rho_a\bar{\gamma}\log h_{-1} + \eta
\end{aligned}
$$

not use policy function iteration due to the number of choice variables, nor use directly approximate the distribution due to the size of the state space.

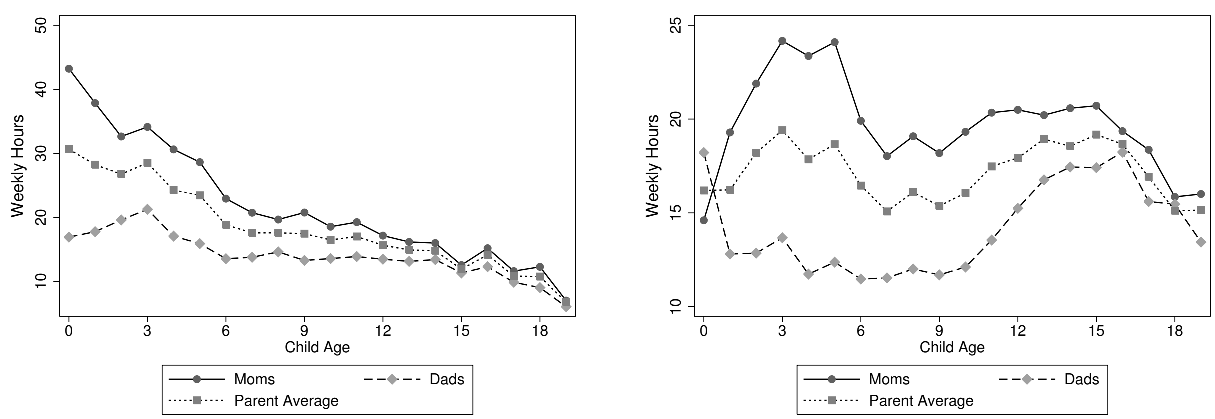
$$\Rightarrow \qquad \log h' = B + (\rho_a + \bar{\gamma}) \log h - \rho_a \bar{\gamma} \log h_{-1} + \eta \tag{29}$$

where h_{-1} is the human capital of the grandparent, and B is a constant. The regression coefficient of $\log h'$ on $\log h$, which is the implied IGE, is easily solved for following Ch. 20 in Greene (2011). Since

$$\text{IGE} \xrightarrow{p} \frac{\text{Cov}(\log h, \log h')}{\text{Var}(\log h)},$$

under stationarity, we can simply take the covariance of both sides of (29) with $\log h$, resulting in BT3.

C. Additional Tables and Figures



(a) Active Time

(b) Passive Time

Figure 17: Time investment in children

Source: PSID, Child Development Supplement Time Diaries. "Active" time is defined as when children report a parents participated in their activity, and "passive" time when they report a parent was around but not participating.

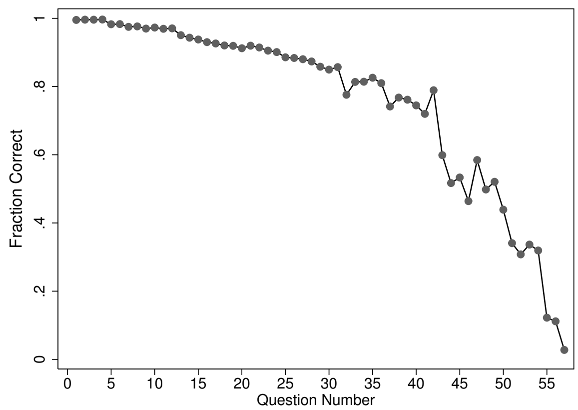


Figure 18: Letter-Word Test Question Difficulties
Source: PSID, Child Development Supplement. For each question, we compute the fraction of children who answer correctly, regardless of age. We use the CDS provided samplying weights, normalized so that the sum of weights in each of the 3 waves are equal.

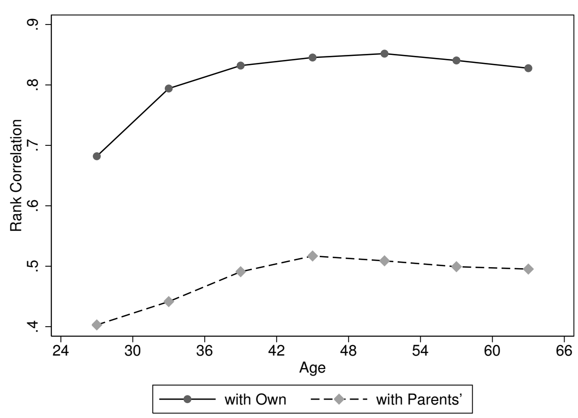


Figure 19: Average Earnings and Human Capital
For each age interval j', the top and bottom lines plot the rank correlation of human capital at age j' with children's own lifetime earnings, and with their parents' lifetime earnings, respectively.

	Short-run			PE Long-run			GE Long-run			
	IGC	Var	$\bar{E}$	IGC	Var	$\bar{E}$	IGC	Var	$\bar{E}$	r
Benchmark	0.34	0.33	1.00	0.34	0.33	1.00	0.34	0.33	1.00	4.00
IG	0.36	0.33	1.00	0.34	0.34	0.99	0.30	0.34	1.00	4.00
LC	0.30	0.34	1.02	0.33	0.32	0.96	0.24	0.32	0.93	4.54
Both	0.37	0.34	0.97	0.32	0.33	1.06	0.32	0.31	1.05	4.40
Flat Tax	0.36	0.34	1.00	0.31	0.32	1.08	0.28	0.33	1.04	4.16
P0	0.35	0.28	1.04	0.30	0.26	1.24	0.09	0.30	1.22	4.29
P1	0.05	0.37	1.30	0.24	0.26	1.13	0.16	0.29	1.12	4.00
P2	0.13	0.33	1.12	0.24	0.28	1.11	0.17	0.31	1.11	4.00

Table 16: Relaxing Borrowing Constraints

Seven experiments. Top panel: relaxing the intergenerational borrowing constraint, the life-cycle borrowing constraint, and both. Bottom panel: flattening the tax function, and focusing all education subsidies in period 0, 1, or 2. The short-run result is the IGC and average earnings one generation after implementing the policy change. “PE Long-run" is when the economy reaches a new steady state, but prices (interest rate and wages) are still fixed at their initial steady state levels. “GE Long-run" is the new steady state equilibrium. The last column is the new equilibrium interest rate in the long-run GE

References

Autor, D. H., L. F. Katz, and M. S. Kearney (2006). The polarization of the u.s. labor market. *American Economic Review 96*(2), 189–194.

Banerjee, A. V. and A. F. Newman (1993, April). Occupational choice and the process of development. *Journal of Political Economy 101*(2), 274–98.

Becker, G. S. and N. Tomes (1979, December). An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of Political Economy 87*(6), 1153–89.

Becker, G. S. and N. Tomes (1986, July). Human capital and the rise and fall of families. *Journal of Labor Economics 4*(3), S1–39.

Behrman, J. R., A. D. Foster, M. R. Rosenzweig, and P. Vashishtha (1999). Women's schooling, home teaching, and economic growth. *Journal of Political Economy 107*(4), pp. 682–714.

Belley, P. and L. Lochner (2007). The Changing Role of Family Income and Ability in Determining Educational Achievement. *Journal of Human Capital 1*(1), 37–89.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*, 352.

Benabou, R. (1993, August). Workings of a city: Location, education, and production. *The Quarterly Journal of Economics 108*(3), 619–52.

Brown, J. and S. Weisbenner (2004, March). Intergenerational Transfers and Savings Behavior. In *Perspectives on the Economics of Aging*, NBER Chapters, pp. 181–204. National Bureau of Economic Research, Inc.

Browning, M., L. P. Hansen, and J. J. Heckman (1999). Micro data and general equilibrium models. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1 of *Handbook of Macroeconomics*, Chapter 8, pp. 543–633. Elsevier.

Carneiro, P. and J. J. Heckman (2002). The evidence on credit constraints in post-secondary schooling. *The Economic Journal 112*(482), 705–734.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011, October). Estimating marginal returns to education. *American Economic Review 101*(6), 2754–81.

Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull (2003). Accounting for the u.s. earnings and wealth inequality. *Journal of Political Economy 111*, 818–857.

Caucutt, E. M. and L. Lochner (2012, October). Early and late human capital investments, borrowing constraints, and the family. Working Paper 18493, National Bureau of Economic Research.

Charles, K. K. and E. Hurst (2003, December). The correlation of wealth across generations. *Journal of Political Economy 111*(6), 1155–1182.

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

Cunha, F. and J. J. Heckman (2007, May). The technology of skill formation. *American Economic Review 97*(2), 31–47.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010, 05). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica 78*(3), 883–931.

De Nardi, M. and F. Yang (2014). Bequests and heterogeneity in retirement wealth. *European Economic Review 72*, 182 – 196.

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

Diamond, P. and J. Gruber (1999). Social security and retirement in the united states. In *Social Security and Retirement around the World*, NBER Chapters, pp. 437–473. National Bureau of Economic Research, Inc.

Durlauf, S. N. (1996, March). A theory of persistent income inequality. *Journal of Economic Growth 1*(1), 75–93.

Gale, W. G. and J. K. Scholz (1994, Fall). Intergenerational transfers and the accumulation of wealth. *Journal of Economic Perspectives 8*(4), 145–60.

Galor, O. and J. Zeira (1993, January). Income distribution and macroeconomics. *Review of Economic Studies 60*(1), 35–52.

Goldberger, A. S. (1989, June). Economic and mechanical models of intergenerational transmission. *American Economic Review 79*(3), 504–13.

Greene, W. H. (2011). *Econometric Analysis* (7 ed.). Pearson.

Guner, N., R. Kaygusuz, and G. Ventura (2014). Income taxation of u.s. households: Facts and parametric estimates. *Review of Economic Dynamics 17*(4), 559 – 581.

Han, S. and C. B. Mulligan (2001, April). Human capital, heterogeneity and estimated degrees of intergenerational mobility. *Economic Journal 111*(470), 207–43.

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

Heckman, J. J. (1976). A life-cycle model of earnings, learning, and consumption. *Journal of Political Economy 84*(4), pp. S11–S44.

Heckman, J. J. (2008, 07). Schools, skills, and synapses. *Economic Inquiry 46*(3), 289–324.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annual Review of Economics 6*(1), 689–733.

Huggett, M., G. Ventura, and A. Yaron (2011, December). Sources of lifetime inequality. *American Economic Review 101*(7), 2923–54.

Hurd, M. and J. P. Smith (2003, April). Expected Bequests and Their Distribution. Working Papers 03-10, RAND Corporation.

Katz, L. F. and K. M. Murphy (1992, February). Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics 107*, 35–78.

Keane, M. P. and K. I. Wolpin (1997, June). The career decisions of young men. *Journal of Political Economy 105*(3), 473–522.

Kennan, J. (2006). A note on discrete approximations of continuous distributions.

Kim, D., S. Sra, and I. S. Dhillon (2010). Tackling box-constrained convex optimization via a new projected quasi-newton approach. *SIAM Journal on Scientific Computing*.

Kopecky, K. and R. Suen (2010, July). Finite State Markov-chain Approximations to Highly Persistent Processes. *Review of Economic Dynamics 13*(3), 701–714.

Kotlikoff, L. J. and L. H. Summers (1981, August). The Role of Intergenerational Transfers in Aggregate Capital Accumulation. *Journal of Political Economy 89*(4), 706–32.

Lee, S. Y. T., A. Seshadri, and N. Roys (2015). The causal effect of parents' education on children's earnings. Technical report.

Loury, G. C. (1981, June). Intergenerational transfers and the distribution of earnings. *Econometrica 49*(4), 843–67.

McGarry, K. (1999, September). Inter vivos transfers and intended bequests. *Journal of Public Economics 73*(3), 321–351.

Mulligan, C. B. (1999, December). Galton versus the human capital approach to inheritance. *Journal of Political Economy 107*(S6), S184–S224.

Restuccia, D. and C. Urrutia (2004). Intergenerational persistence of earnings: The role of early and college education. *American Economic Review 94*, 1354–1378.

Solon, G. (1992, June). Intergenerational income mobility in the united states. *American Economic Review 82*(3), 393–408.