ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

WORKING PAPER SERIES

Parents, Patience, and Persistence: A Novel Theory of Intergenerational College Attainment

Authors	David L. Fuller, and Guillaume Vandenbroucke
Working Paper Number	2025-003A
Creation Date	February 2025
Citable Link	https://doi.org/10.20955/wp.2025.003
Suggested Citation	Fuller, D.L., Vandenbroucke, G., 2025; Parents, Patience, and Persistence: A Novel Theory of Intergenerational College Attainment, Federal Reserve Bank of St. Louis Working Paper 2025-003. URL https://doi.org/10.20955/wp.2025.003

Federal Reserve Bank of St. Louis, Research Division, P.O. Box 442, St. Louis, MO 63166

The views expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Federal Reserve System, the Board of Governors, or the regional Federal Reserve Banks. Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment.

Parents, Patience, and Persistence: A Novel Theory of Intergenerational College Attainment*

D. Fuller† G. Vandenbroucke‡

February 2025

Abstract

We present a novel theory of intergenerational persistence in college attainment that does not rely on credit constraints, parental transfers and investments, or persistence in innate ability. The gist of our theory is heterogeneity in time preferences, which are endogenous since parents can teach patience to their children before the children make their college decisions. We show, analytically, that the most patient parents are simultaneously more likely to have a college degree and to educate their children to be patient. Persistence follows. We also show that persistence occurs if and only if there is persistence in patience. So, our model rationalizes the view that long-run family factors matter for educational attainment. We embed our theory in a general equilibrium model which we calibrate to U.S. data. The model matches both college persistence and cross sectional inequality moments. We implement a policy experiment in which a college subsidy is financed via a proportional tax on labor income. The policy distorts intertemporal trade-offs and, hence, affects people differently depending on their patience. We find that the policy raises overall college attainment (individuals earn a BA) and its persistence. Specifically, it raises the likelihood of attaining a BA for individuals whose parents have a BA and it lowers that likelihood for individuals whose parents do not have a BA. This result differs from that of the existing literature where college subsidies alleviate borrowing constraints and tend to reduce intergenerational persistence.

JEL codes: E6, I2, J62.
Keywords: Endogenous preferences; Intergenerational persistence; Inequality; College choice.

*The views expressed in this article are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

†University of Wisconsin-Oshkosh; Email: fullerd@uwosh.edu

‡Research Division, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166, USA; Email: guillaumevdb@gmail.com.

1 Introduction

Educational attainment, which is associated with higher income, better health, and longer lives, critically depends on parental educational attainment. In the United States in the 1980s, for example, the probability of attaining a BA is greater than 60% for the people whose parents have a BA; it is less than 30% for the others (Gonzalez and Langot, 2022). In short, standards of living persist across generations through education. This persistence is a source of inequality and, as such, the subject of a literature that attributes it to a combination of persistence in innate ability, transfers, and borrowing constraints (e.g., Becker and Tomes, 1979, 1986, Restuccia and Urrutia, 2004, Lee and Seshadri, 2019). In this paper, we propose a novel theory of persistence in college attainment.

Our theory centers around heterogeneity in time preferences, which are endogenous as in Becker and Mulligan (1997). Specifically, parents can teach patience to their children before they make their college decisions. The most patient parents are the most likely to have a college degree, and they also value their children's welfare the most. Hence, they are more likely to induce them to attain a college degree by teaching them patience. Persistence follows.

Some aspects of this description deserve comment. First, in our theory, the most patient are the most likely to attain a college degree. Empirically, there exists a well-documented positive association between measures of patience and educational attainment (e.g., Golsteyn et al., 2014, De Paola and Gioia, 2017, Hanushek et al., 2023). Thus, this aspect of our theory is in line with existing evidence. Theoretically, patience is irrelevant to investment decisions, such as attending college, whenever credit markets are perfect and the costs and benefits of investing are exclusively pecuniary.[1] Practically, we assume a utility cost of college (e.g., Heckman et al., 1998, 2003) to generate the positive association between patience and college attainment; however, it is important to note that this role of patience does not depend critically on the utility cost. A borrowing constraint also implies a positive association between patience and college attainment. We adopt a utility cost to demonstrate that our mechanism operates even in the absence of a borrowing constraint. In the context of our model, a borrowing constraint would matter for college attainment via two mechanisms: reduced college attainment for low-income individuals, conditional on patience, and reduced college attainment for low-patience individuals, conditional on income. The latter effect would encapsulate the novelty of our theory.

Second, in our theory, parents can teach patience to their children. Even though such teaching could be taking place within schools, we interpret it as taking place within families. Interestingly, Cameron and Heckman (1998) report that the effect of mothers' educational attainment on their children's educational attainment is stronger than that of fathers. We interpret this finding as evidence that parenting, mostly done by mothers, plays a critical role in the children's development

[1] In such cases, only the net present value criterion matters.

of their time preference. Similarly, a literature on adoptees (see Sacerdote, 2011) documents intergenerational persistence of outcomes, including educational attainment, finding an important role for parents outside of genetic endowments. It is conceivable that in addition to teaching patience, parents teach other non-cognitive skills such as a taste for effort. However, this alone cannot constitute a theory of persistence, since the question remains: Why would college-educated parents teach the importance of effort more than others? Patience is a "natural" candidate to think about persistence because it has two distinct effects: it affects college attainment and it affects parents' valuation of their children. The fact that, in our theory, patient parents value their children the most arises solely from assuming that they value their children's future welfare. Specifically, we do not assume that patient parents are "more altruistic" than others.

Finally, our theory implies intergenerational persistence in patience, which we show is equivalent to intergenerational persistence in college attainment. In sum, patience, which is taught within families, is persistent and critical to college decisions. We therefore interpret it as a long-run family factor in the spirit of Cameron and Heckman (1998) who wrote

> *"It is (...) long-run family factors and not short-run credit constraints that play a decisive role in explaining schooling attainment."*

We implement our theory in a general equilibrium model with perfect credit markets and where overlapping generations of individuals are endowed at birth with an ability to be productive and a discount factor. These traits are drawn from two distributions that are independent of one another and independent of the parents' traits. Parents can exert a costly effort to change the discount factor of their children before the children become adults, at which time the children's birth-ability and their (possibly changed) discount factor become their adult traits. Adults face a college decision before they, themselves, are endowed with one child. College increases lifetime wages but requires time and entails a good and a utility cost. We make some simplifying assumptions: (i) all children complete high-school; (ii) there are no college dropouts, so that attending college and graduating are the same things; and (iii) there are only two possible discount factors.

At the individual level, our model delivers two sets of threshold rules. First, there are college-decision rules: Adults with ability above a threshold that depends on their discount factor choose to attend college. We show that these thresholds decrease with the discount factor, implying that patient adults are more likely to choose college. Second, there are education-decision rules, namely rules for whether parents educate their children to the higher discount factor (if their birth-discount factor was low) or not: Children with ability above a threshold that depends on their parents' discount factor get educated in this way. We show that these thresholds, again, decrease with the discount factor of the parents, implying that patient parents are more likely to educate their children to be patient.

At the aggregate level, our model's equilibrium features a joint distribution of adults by college

attainment and discount factors, as well as a transition function mapping the distribution for one generation into that of the next generation. We compute the stationary distribution associated with this process and analyze it. We find that persistence is not guaranteed because there are "regions of inaction:" For instance, when children have such low ability that college is suboptiomal, even with patience, or when children have such high ability that college is optimal, even with impatience. In both cases, parents cannot affect their children college decision and, hence, persistence cannot hold. For persistence to exist, (i) parents must instill patience in their children, **and** (ii) such choices by the parent must change the children's college decisions, **and** (iii) such effects must differ across parents with different college attainment. Our main theoretical result, therefore, describes the conditions under which persistence is either positive or zero.

We calibrate our model to U.S. data. We target, in particular, moments pertaining to the level of college attainment and its persistence across generations (e.g., the likelihood of attaining a BA for people whose parents have a BA). We also target moments related to the dispersion of earnings between and across graduation groups (e.g., the college premium and the coefficient of variation of earnings conditional on whether a worker holds a BA or not). We find that our model matches the persistence in college attainment across generations while being consistent with cross-sectional inequality. We note that the model implies significantly less persistence in earnings than is typically measured in the U.S. data (4% v. 40%). We do not see this as a failure since we have, on purpose, abstracted from modeling any persistence in innate ability.[2]

We use the calibrated model to simulate a policy experiment in which a college subsidy is financed via a proportional labor tax. We find that the policy raises both college attainment and intergenerational persistence in college attainment. The latter effect obtains because the policy distorts intertemporal tradeoffs via a general equilibrium effect: The subsidy in the present changes prices that apply in the present and the future in the labor market. So, the policy affects people differently depending on their patience. Specifically we explain that the policy discourages impatient individuals to seek a BA. Since impatient individuals are likely to have impatient parents that do not have a BA, this mechanism is conducive to persistence. Furthermore, the policy encourages more patient individuals to seek a BA. Since these individuals are likely to have patient parents that do have a BA, this second mechanism is also conducive to persistence. Interestingly, our results differ from those in the literature since it is often found that education subsidies reduce intergenerational persistence by alleviating borrowing constraints.

Our results are important because they point out a mechanism not often mentioned in the literature: The relevance of patience heterogeneity to the persistence in college attainment **and** to the effects of college-related policies. Models of persistence have often abstracted from patience heterogeneity and emphasized credit market imperfections instead. But credit market imperfection imply that patience matters in individual decisions—it does not under perfect credit markets. Hence, if models with borrowing constraints allowed patience heterogeneity, the mechanism we discuss here would

[2] In Restuccia and Urrutia (2004) the persistence in innate ability accounts for 50% of the persistence in earnings.

apply there as well.

1.1 Literature

Our paper is at the intersection of two strands of the literature: One that studies the endogenous determination and transmission of preferences (e.g., time and risk preferences, altruism) and cultural traits (e.g., attitudes towards work, family, religion, marriage, and sexuality); and one that studies the intergenerational persistence of economic outcomes.

One focus of the theoretical literature on preferences and cultural transmission is that of the persistence of (possibly minority) traits in a population or, conversely, their assimilation (e.g., Bisin and Verdier, 2001, 2011, Saez-Marti and Zenou, 2012). In our theory the endogenous traits are time preferences (i.e., patience) in the spirit of Becker and Mulligan (1997). We show that patience is persistent across generations and exhibit a non-degenerate stationary distribution, but we do not argue this to be an innovation. Instead, we emphasize an implication of our model that persistence in patience occurs if and only if there is persistence in college attainment. Hence the link we draw between the aforementioned strands of the literature.

Models with endogenous preferences have been developed to tackle a variety of issues: Bisin and Verdier (2000) build a model of marital segregation in a multi-ethnic society; Fernández et al. (2004) analyze the labor force participation and education of women; Doepke and Zilibotti (2008) study the socioeconomics of the Industrial Revolution; Becker et al. (2016) present a model where parents can manipulate children's preferences toward supporting social security; Doepke and Zilibotti (2017) connect economic development and parenting styles (i.e., from permissive to authoritarian); in Fernández-Villaverde et al. (2014), parents' can shame children into avoiding pre-marital sex. In the spirit of this literature, we develop a model where parents influence their children's patience. We argue this yields a novel theory of the intergenerational persistence of economic outcomes.

Thus, our main contribution is to the literature on the intergenerational persistence of economic outcomes, specifically college attainment, and to a lesser extent earnings.[3] Building on the seminal work of Becker and Tomes (1979, 1986), authors such as Keane and Wolpin (2001), Restuccia and Urrutia (2004), and Lee and Seshadri (2019) ascribe persistence in earnings to a combination of persistence in innate ability, parental transfers and investments, and borrowing constraints. Their focus is largely quantitative: They show that, with these elements, their models generate observed intergenerational persistence while being consistent with observed cross-sectional inequality. We complement their work since our novel mechanism does not require the elements to which they ascribe persistence. Instead, it relies on endogenous heterogeneity and persistence in patience.

The literature on the role of patience for a range of lifetime outcomes, particularly, formal edu-

[3]There exist a literature, to which we do not contribute, on intergenerational transfers and their role for macroeconomic questions such as capital accumulation (Kotlikoff and Summers, 1981) and growth (Owen and Weil, 1998).

cation and income, is vast. Sutter et al. (2013)'s experimental result is that high ability students (as measured by math grades) are more patient. They further find that patience measured in experiments are strong predictors of actual behavior in school. Golsteyn et al. (2014), using Swedish data, find a positive relationship between patience and academic performance, health, labor supply, and lifetime income. De Paola and Gioia (2017) also report a positive association of patience with academic performance for Italian undergraduate students; and document higher drop-out rates for the impatient students. Sunde et al. (2022) and Hanushek et al. (2022) use data from the Global Preference Survey where patience is measured (for a large number of countries) through answers to hypothetical choices between immediate and later rewards. They find patience to be strongly associated with student achievement and educational attainment. Hanushek et al. (2023) use social-media data to measure patience, and find similar results across regions of Italy and the United States. Angerer et al. (2023) find patience to be positively associated with students choosing academic high-school track (instead of vocational school track). Finally, and closely related to our topic, Albanese et al. (2016) report that values and traits, such as patience, are closely associated with those received from parents and that they are often transmitted to children.[4]

Intergenerational preferences can affect persistence. We adopt a "standard" approach in the literature. That of altruistic parents who value their children's welfare (e.g., Becker and Tomes, 1986). We do not assume "warm glow" preferences, where parents value transfers to their children (e.g., Vandenbroucke, 2021); nor do we assume "paternalistic" preferences, where parents value a trait of their children such as human capital of college attainment (e.g., Glomm and Ravikumar, 1992). Although standard, our approach together with heterogeneity in patience has an important implication since it yields heterogeneity in the value of children to their parents even though we do not assume heterogeneity in altruism.

We make two further observations. First, our abstracting from borrowing constraints is not a statement on their irrelevance. Recent work has shown the opposite (Lochner and Monge-Naranjo, 2012, Caucutt and Lochner, 2020). We note, however, that borrowing constraints would not be an orthogonal addition to our model, as impatient individuals are more likely to be constrained than patient ones. Because of such interaction between borrowing constraints and the mechanism in our model, the abstraction serves to illustrate the workings of our theory. Second, we abstract from persistence of ability for the same reason: simplicity. We do not argue that persistence in innate ability is irrelevant. We note, however, that an empirical literature finds significant persistence between the college attainment of adopted children and that of their adoptive parents (e.g. Sacerdote, 2000, 2007, 2011). Furthermore, recall that Cameron and Heckman (1998) find a positive effect mothers' education on their children, after controlling for ability. Thus, we argue that a model that can deliver persistence of outcomes without persistence in ability is of value.

[4] A broader empirical literature attempting to measure cultural traits (not just patience) and investigating whether they matter for economic outcomes is quite related (e.g., Guiso et al., 2006, Fernández, 2011).

2 Model

2.1 Environment

Time is discrete. The economy is populated by overlapping generations of individuals and a representative firm. The firm hires capital and labor from high school and college graduates in competitive factor markets, and produces output via a constant-returns-to-scale technology. The individuals live for L periods. Childhood lasts from age 1 to $A-1$ and adulthood from A to L. Only adults are economically active. It is convenient to define adult life to last from (adult) age $j=1$ to $j=J$ where $J=L-A+1$. One child is born to each adult whose preferences are defined over consumption and the child's future welfare. Each newborn is endowed with a type (a,b). The term $a\in\mathbb{R}_+$ is labor market ability and its distribution is described by the probability measure F. The term b is a pair (β,δ) comprising two discount factors: β, which discounts the future utility from consumption; and δ, which discounts the future welfare of children. There are two pairs of discount factors: $b_L\equiv(\beta_L,\delta_L)$ and $b_H\equiv(\beta_H,\delta_H)$. Newborn are endowed with b_L with probability μ and with b_H with probability $1-\mu$. We assume $\beta_H\geq\beta_L$ and $\delta_H\geq\delta_L$.

Age-1 adults make two decisions besides consumption and saving. First, before their children are born, they decide whether to attend college or not. If they do, they pay a goods cost $e>0$, a utility cost $\psi\geq 0$, and they do not work for S periods. When they work, from age $S+1$ to J, they supply their market ability a at the rate w^{co} per period. If they do not attend college, they work from age 1 to J and supply ability at the rate w^{hs} per period. There is a perfect credit market where the interest rate is r.

Second, after their children are born and their type (a',b') revealed, age-1 adults can choose to exert an effort, with utility cost $\chi\geq 0$, to change their children's discount factors. For example, some age-1 adults with b_L-born children may choose to change their children's discount factors to b_H so that, at the start of their adults lives, these children are b_H-adults. If they decide to not change their children's discount factors age-1 adults do not need to exert effort and their children's adult discount factor are that with which they were born.

We assume that college attendance implies graduation. Thus, we refer to individuals who attend college as college graduates and to the others as high school graduates. Furthermore, we refer to the decision to attend college as the "college decision," and to the decision to change a child's birth-b as the "education decision."

2.2 Production

The representative firm seeks to maximize profit:

$$\max_{K,L^{\text{co}},L^{\text{hs}}} \; F(K,L^{\text{co}},L^{\text{hs}}) - (r+\delta)K - w^{\text{co}}L^{\text{co}} - w^{\text{hs}}L^{\text{hs}}, \tag{1}$$

where L^{co} and L^{hs} are the demands for efficiency units of labor from college and high school graduates, respectively. The demand for capital is K and δ is the rate of depreciation of capital. The first-order optimality conditions are

$$\begin{aligned} K &: \; 0 = r + \delta - F_K(K,L^{\text{co}},L^{\text{hs}}), \\ L^{\text{co}} &: \; 0 = w^{\text{co}} - F_{L^{\text{co}}}(K,L^{\text{co}},L^{\text{hs}}), \\ L^{\text{hs}} &: \; 0 = w^{\text{hs}} - F_{L^{\text{hs}}}(K,L^{\text{co}},L^{\text{hs}}). \end{aligned}$$

We adopt the following functional form for the production function:

$$F(K,L^{\text{co}},L^{\text{hs}}) = K^{\alpha}[z^{\text{co}}(L^{\text{co}})^{\rho} + z^{\text{hs}}(L^{\text{hs}})^{\rho}]^{\frac{1-\alpha}{\rho}}$$

where $\alpha \in (0,1)$ is the capital share of income and $\rho \leq 1$. The elasticity of substitution between the two types of labor is $1/(1-\rho)$.

2.3 Individuals

Let $V(a,b)$ denote the lifetime utility of age-1 (a,b)-adults, before their children are born:

$$V(a,b) = \max\left\{\mathcal{V}^{\text{hs}}(a,b), \mathcal{V}^{\text{co}}(a,b)\right\} \tag{2}$$

where

$$\begin{aligned} \mathcal{V}^{\text{x}}(a,b) &= \mathcal{U}^{\text{x}}(a,b) - \psi\mathbb{I}_{\text{x=co}} + E\left[\mathcal{W}\left(a',b';b\right)\right], & (3)\\ \mathcal{U}^{\text{x}}(a,b) &= \max\left\{\sum_{j=1}^{J}\beta^{j-1}U\left(c_j\right) : \sum_{j=1}^{J}c_j\left(1+r\right)^{1-j} = I^{\text{x}}(a)\right\}, & (4)\\ \mathcal{W}\left(a',b';b\right) &= \max_{z\in\{b_L,b_H\}}\delta V\left(a',z\right) - \chi\mathbb{I}_{z\neq b'}, & (5)\\ E\left[\mathcal{W}\left(a',b';b\right)\right] &= \int_{a'}\left[\mu\mathcal{W}\left(a',b_L;b\right) + (1-\mu)\,\mathcal{W}\left(a',b_H;b\right)\right]F\left(da'\right). & (6) \end{aligned}$$

Equation (3) describes the value of graduating college (x = co) or not (x = hs). It comprises the utility $\mathcal{U}^{\text{x}}(a,b)$ from lifetime consumption described in Equation (4), the utility cost of college when

x = co, and the expected utility of children described in Equations (5) and (6). The function $I^{\mathrm{x}}(a)$ in Equation (4) denotes the net present value of labor income for high school and college graduates:

$$I^{\mathrm{hs}}(a) = aw^{\mathrm{hs}} \sum_{j=1}^{J} \left(\frac{1}{1+r}\right)^{j-1} \equiv aw^{\mathrm{hs}} D^{\mathrm{hs}}, \tag{7}$$

$$I^{\mathrm{co}}(a) = -e + aw^{\mathrm{co}} \sum_{j=S+1}^{J} \left(\frac{1}{1+r}\right)^{j-1} \equiv -e + aw^{\mathrm{co}} D^{\mathrm{co}}. \tag{8}$$

The expected utility of children is derived from $\mathcal{W}(a', b'; b)$ which is the discounted value, to b-parents, of a newborn with type (a', b'). Note that the cost χ of educating children to discount factor z is borne only if z differs from the children's discount factors at birth. Equation (6) describes the expected value of $\mathcal{W}(a', b'; b)$.

We adopt the following functional form for the period utility index:

$$U(c) = \frac{c^{1-\sigma}}{1-\sigma},$$

where $\sigma \in (0, 1)$ to ensure that utility is positive as in Becker and Mulligan (1997). Such assumption is similar to that of models with endogenous life expectancy (see Rosen, 1988). In these models, the survival probability acts as a discount factor. If utility is negative, an increase in the survival probability or in the discount factor reduces lifetime utility. Thus, positive utility implies that individuals prefer longer lives to shorter ones in models with uncertain life expectancy. In our model, positive utility implies that individuals prefer to be patient rather than impatient.

We differentiate β and δ to permit a distinction between discounting and altruism. To see this, suppose that individuals were endowed with an altruism parameter ω such that adults' preferences were

$$V_{\text{parent}} = \underbrace{U(c_1) + \cdots + \beta^{J-1} U(c_J)}_{\text{Utils from own consumption}} + \underbrace{\overbrace{\beta^{A-1}\omega}^{\delta} V_{\text{child}}}_{\text{Utils from child's welfare}}.$$

Then, it is immediate that our (β, δ) formulation conflates discounting and altruism. Suppose that there exists two possible values for the altruism parameter: $\omega \in \{\omega_L, \omega_H\}$. Then

$$\beta_H > \beta_L \Rightarrow \delta_H > \delta_L \quad \text{if and only if} \quad \frac{\omega_H}{\omega_L} > \underbrace{\left(\frac{\beta_L}{\beta_H}\right)^{A-1}}_{<1}.$$

So, our assumption that the discount factors β and δ are positively assorted in the pairs $b_L \equiv (\beta_L, \delta_L)$ and $b_H \equiv (\beta_H, \delta_H)$ does not imply an assumption about the ordering of altruism since ω_H/ω_L may be above or below 1. In the quantitative application of Section 3, we assume away heterogeneity in altruism and, hence, set $\omega_H = \omega_L = 1$.

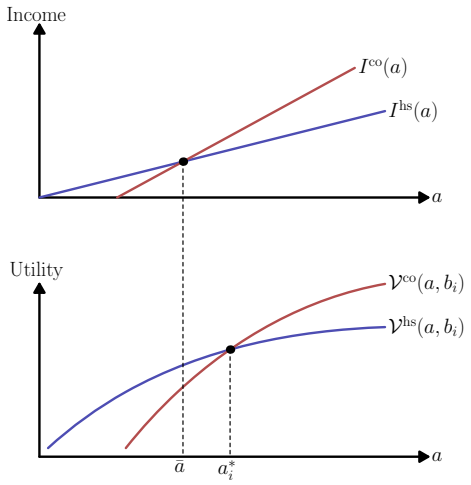


Figure 1: The college decision

2.3.1 *College decision*

We define $\bar{a}$, as the ability equating the lifetime incomes of high school and college graduates:

$$I^{hs}(\bar{a}) = I^{co}(\bar{a}). \tag{9}$$

There exists one $\bar{a} > 0$ if and only if $w^{co}D^{co} > w^{hs}D^{hs}$. This condition is not guaranteed to hold for any set of parameters since w^{co} and w^{hs} are endogenous. We assume it holds, however, since $\bar{a} < 0$ implies that any individual would be financially worse off from graduating college. In such case there would be no college graduates and the model's solution would be trivial.

We define a_i^*, for each $i \in \{L, H\}$, as the ability threshold such that (a_i^*, b_i)-individuals are indifferent between being high school or college graduates:

$$\mathcal{V}^{hs}(a_i^*, b_i) = \mathcal{V}^{co}(a_i^*, b_i). \tag{10}$$

First, in Lemma 4 (Appendix A), we establish that a_i^* exists, that it is unique, and that

$$\begin{array}{l}\text{College decision of} \\ (a, b_i)\text{-adult}\end{array} : \begin{cases} \text{college} & \text{if } a \geq a_i^*, \\ \text{high school} & \text{if } a < a_i^*. \end{cases} \tag{11}$$

Second, we establish that $a_i^* \geq \bar{a}$ with equality whenever $\psi = 0$ (Lemma 5). Thus, when the utility cost of college is zero, the college decision is income maximizing because $I^{co}(a) > I^{hs}(a)$ whenever $a > \bar{a}$. Figure 1 represents the college decision when the utility cost of college is positive. Age-1 adults with discount factors b_i and ability $a \in [\bar{a}, a_i^*]$ choose to be high school graduates even though their income would be higher if they became college graduates. That is because their additional

income from college would not compensate the utility cost.

Third, we note that a_i^* depends on the discount factor β_i but it is independent of the discount factor δ_i. That is because the expected utility of children is independent from a parent's college decision, as transpires from Equations (3) and (10). Furthermore, Equation (10) implies $a_H^* = a_L^*$ whenever $\beta_L = \beta_H$. Fourth, we establish that $a_H^* < a_L^*$ whenever $\psi > 0$ and $\beta_L < \beta_H$ (Lemma 6).

In sum, we have established the following properties of the college-decision thresholds:

$$\begin{cases} a_H^* < a_L^* & \text{whenever } \psi > 0 \text{ and } \beta_L < \beta_H, \\ a_H^* = a_L^* & \text{otherwise.} \end{cases} \tag{12}$$

It is worth taking stock of these results. The first row of Equation (12) indicates that b_H-adults, i.e., patient adults, are more likely to graduate college than b_L-adults only if college entails a utility cost ($\psi > 0$). Absent the utility cost, b_H- and b_L-adults are equally likely to graduate college (second row of 12). These results stem from the net present value (NPV) criterion, which applies whenever credit markets are perfect and the cost and benefits of investments are exclusively pecuniary. So, when $\psi = 0$ the NPV criterion applies: Individual compare $I^{\text{co}}(a)$ and $I^{\text{hs}}(a)$ and their discount factors are irrelevant to college decisions. When $\psi > 0$, however, the NPV criterion does not apply. There are no markets from where to "borrow utils." Individuals must bear the utility cost at age 1 and assess whether they value the future utils afforded by college enough. The higher their β, the more they do. Unlike income, utility cannot be traded intertemporally leading to the role of β.

2.3.2 *Education decision*

After children are born and their type (a', b') revealed, parents may decide to change their children's discount factors. Recall that doing so imposes a utility cost χ on the parents. We establish that parents would not change their children's discount factors from b'_H to b'_L because it would lower the children's welfare and, hence that of the parents (Lemma 7). Consider children born with type (a', b'_L). Their value depend on their future college decision. Three cases must be considered:

1. $a' < a_H^*$ – They will not graduate college. So,

$$\mathcal{W}\left(a', b'_L; b\right) = \max\left\{\delta \mathcal{V}^{\text{hs}}\left(a', b_H\right) - \chi, \delta \mathcal{V}^{\text{hs}}\left(a', b'_L\right)\right\}.$$

2. $a_H^* \leq a' \leq a_L^*$ – They will graduate college only if educated. So,

$$\mathcal{W}\left(a', b'_L; b\right) = \max\left\{\delta \mathcal{V}^{\text{co}}\left(a', b_H\right) - \chi, \delta \mathcal{V}^{\text{hs}}\left(a', b'_L\right)\right\}.$$

3. $a' > a_L^*$ – They will graduate college. So,

$$\mathcal{W}\left(a', b_L'; b\right) = \max\ \left\{\delta \mathcal{V}^{\text{co}}\left(a', b_H\right) - \chi, \delta \mathcal{V}^{\text{co}}\left(a', b_L'\right)\right\}.$$

Then, it is useful to define

$$\Delta\left(a\right) = \begin{cases} \mathcal{V}^{\text{hs}}\left(a, b_H\right) - \mathcal{V}^{\text{hs}}\left(a, b_L\right) & \text{when } \ a \leq a_H^*, \\ \mathcal{V}^{\text{co}}\left(a, b_H\right) - \mathcal{V}^{\text{hs}}\left(a, b_L\right) & \text{when } \ a_H^* \leq a \leq a_L^* \\ \mathcal{V}^{\text{co}}\left(a, b_H\right) - \mathcal{V}^{\text{co}}\left(a, b_L\right) & \text{when } \ a \geq a_L^*. \end{cases}$$

If follows from (10) that Δ is continuous. Furthermore, it is monotonically increasing (Lemma 8). We define the thresholds

$$\Delta(\hat{a}_i) = \frac{\chi}{\delta_i}, \tag{13}$$

such that b_i-parents with $(\hat{a}_i, b_L')$-born children are indifferent between educating their children or not. Since Δ is increasing the decision to educate is

$$\begin{array}{l} \text{Education decision of} \\ b_i\text{-parents of } (a', b_L')\text{-born children} \end{array} \ : \ \begin{cases} \text{educate} & \text{if } a' \geq \hat{a}_i, \\ \text{do not educate} & \text{if } a' < \hat{a}_i. \end{cases} \tag{14}$$

and

$$\begin{cases} \hat{a}_H < \hat{a}_L & \text{whenever } \delta_L < \delta_H, \\ \hat{a}_H = \hat{a}_L & \text{otherwise.} \end{cases} \tag{15}$$

Equation (14) and (15) convey two distinct, albeit related, messages. Equation (14) indicates that parents educate their b_L-born children if the children have high enough ability. Equation (15) indicates that b_H-parents are more likely than b_L-parents to educate their children.

The logic of (14) follows from the fact that welfare is increasing in both ability and the discount factor. When parents educate their children they benefit, therefore, from the higher future welfare of the children. The higher the ability of the children, the more likely the benefit to the parents is to offset the cost of education χ. Hence the threshold rule in (14). Naturally, the value of the children's welfare to the parents is increasing with the parents' δ. Hence, the threshold ability at which children are educated is lower when the parents are patient. Hence, b_H-parents are more likely than b_L-parents to educate their b_L-born child as indicated in Equation (15).

There is a total of six combinations of the a_i^*s and $\hat{a}_i$s satisfying $a_H^* < a_L^*$ and $\hat{a}_H < \hat{a}_L$. Figure 2 illustrates the education decision of parents with (a', b_L')-born children in two cases. Consider Panel A. The grey bars indicate the set of children getting educated by their parents: $a' \geq \hat{a}_i$. The pink bar shows the set of b_L-born children such that b_H-parents educate them **and** such that this education induces them to be college graduate. We denote this set $\mathbb{B}_H$. All children with ability above $\hat{a}_H$ are educated by b_H-parents, but only those with ability less than a_L^* are future college

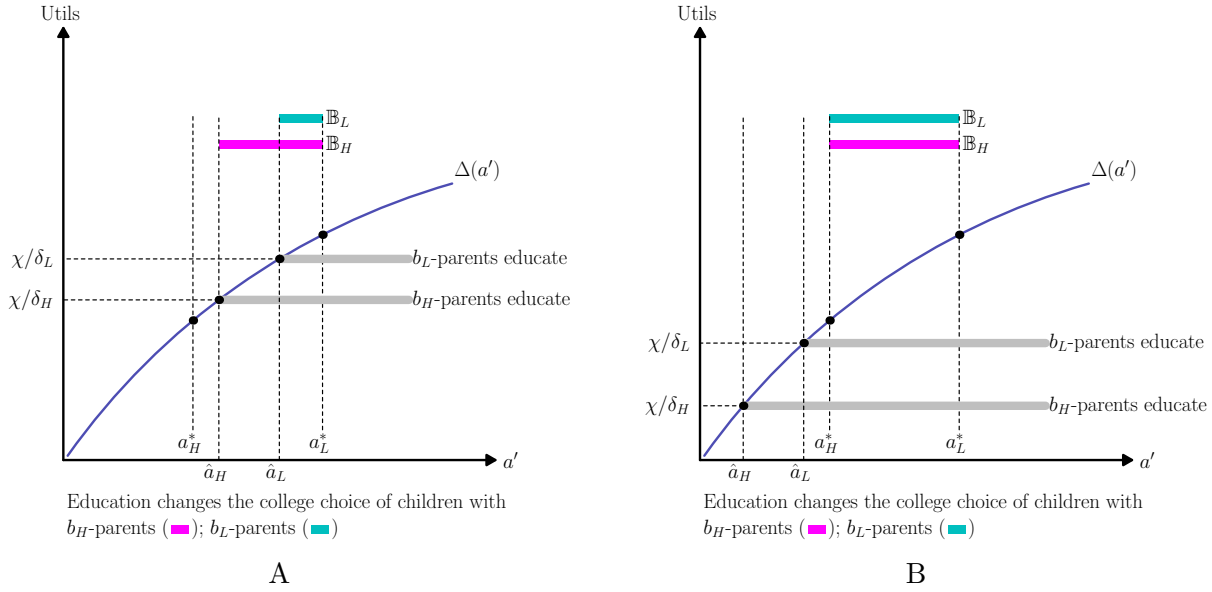


Figure 2: The education of (a', b'_L)-born children

graduates because of parental education. Children with ability above a_L^* become college graduate with or without parental education. Similarly the green bar shows the set $\mathbb{B}_L$ of b_L-born children such that b_L-parents educate them and such that this education induces them to graduate college.

In Panel B, children with ability lower than a_H^* do not graduate college whether or not their parents educate them. Thus, b_L-born children with ability between $\hat{a}_H$ and a_H^*, for example, do get educated by b_H-parents, but do not graduate college as a result of this education. Only the children with ability between a_H^* and a_L^* may be induced by their parents to graduate college.

Formally, the sets $\mathbb{B}_H$ and $\mathbb{B}_L$ are described as

$$\mathbb{B}_i = \underbrace{\{a \geq \hat{a}_i\}}_{\text{educate the child}} \cap \underbrace{\{a_H^* \leq a \leq a_L^*\}}_{\text{education changes college decision}} \tag{16}$$

and their measure is

$$F(\mathbb{B}_i) = \max(0, \min(P_i, C_H) - C_L), \tag{17}$$

where C_i is the probability that b_i-adults decide to graduate college and P_i is the probability that they educate their b_L-born children:

$$\begin{aligned} C_i &= F(\{a > a_i^*\}), \\ P_i &= F(\{a > \hat{a}_i\}). \end{aligned}$$

The sets $\mathbb{B}_H$ and $\mathbb{B}_L$ are of first-order importance in our analysis. On the one hand, Panel A of Figure 2 show that b_H-parents, who are more likely to have a college degree (Lemma 6), are also more likely to induce their children to become college graduate. Thus, the difference between $F(\mathbb{B}_H)$ and $F(\mathbb{B}_L)$ creates intergenerational persistence in educational attainment. On the other hand, in Panel B, the measure of children for whom parental education matters for college attainment is independent of the parents' type. This is conducive to an absence of intergenerational persistence.

2.4 Equilibrium

Let $\Pr(b, \mathrm{x}, da)$ denote the stationary measure of adults with discount factor $b \in \{b_L, b_H\}$, graduation status $\mathrm{x} \in \{\mathrm{co}, \mathrm{hs}\}$, and ability in the interval $da \subseteq \mathbb{R}_+$:

$$\Pr(b_H, \mathrm{co}, da) = \overbrace{(1-\mu)F(da)\mathbb{I}_{da\subseteq[a_H^*,\infty]}}^{b_H\text{-born}} + \overbrace{\mu \sum_i \Pr(b_i)F(da)\mathbb{I}_{da\subseteq[\max\{a_H^*,\hat{a}_i\},\infty]}}^{b_L\text{-born, } b_i\text{-parents, educated}}, \tag{18}$$

The first element is the measure of b_H-born individuals (probability $1-\mu$) with ability in the interval da. They become college graduates if their ability exceeds a_H^* (Equation 11), hence the indicator function. The second element sums the measures of b_L-born individuals (probability μ) with b_i parents (probability $\Pr(b_i)$) and ability in the interval da. They become b_H-college graduates if their parents teach them patience, i.e., their ability exceeds $\hat{a}_i$ (Equation 14), and if college is optimal, i.e., their ability exceeds a_H^*.[5] The expressions below have similar interpretations:

$$\Pr(b_H, \mathrm{hs}, da) = \overbrace{(1-\mu)F(da)\mathbb{I}_{da\subseteq[0,a_H^*]}}^{b_H\text{-born}} + \overbrace{\mu \sum_i \Pr(b_i)F(da)\mathbb{I}_{da\subseteq[\hat{a}_i,a_H^*]}}^{b_L\text{-born, } b_i\text{-parents, educated}}, \tag{19}$$

$$\Pr(b_L, \mathrm{co}, da) = \overbrace{\mu \sum_i \Pr(b_i)F(da)\mathbb{I}_{da\subseteq[a_L^*,\hat{a}_i]}}^{b_L\text{-born, } b_i\text{-parents, not educated}}, \tag{20}$$

and

$$\Pr(b_L, \mathrm{hs}, da) = \overbrace{\mu \sum_i \Pr(b_i)F(da)\mathbb{I}_{da\subseteq[0,\min\{a_L^*,\hat{a}_i\}]}}^{b_L\text{-born, } b_i\text{-parents, not educated}}. \tag{21}$$

Note that (18)-(21) define $\Pr(b, \mathrm{x}, da)$ implicitly since $\Pr(b)$ is a marginal (for lack of a better term) of $\Pr(b, \mathrm{x}, da)$. In Appendix A.4 we derive the marginal distribution of endogenous state variables,

$$\Pr(b, \mathrm{x}) = \int \Pr(b, \mathrm{x}, da),$$

[5] In Equations (18)-(21) we use the convention $[x, y] \equiv \emptyset$ whenever $x > y$.

and show that it is stationary across generations (Lemma 9). Since the distribution of ability is stationary by assumption, it follows that $\Pr(b, \mathrm{x}, da)$ is stationary.

Let $c_j^{\mathrm{x}}(a,b)$ and $s_j^{\mathrm{x}}(a,b)$ denote the consumption and asset holdings of these adults at age j, respectively.[6] In any period, there are $J-S$ cohorts of working college graduates and J cohorts of working high school graduates. Thus, the labor market clearing conditions are

$$L^{\mathrm{co}} = (J-S)\sum_b \int a \Pr(b, \mathrm{co}, da) \quad \text{and} \quad L^{\mathrm{hs}} = J\sum_b \int a \Pr(b, \mathrm{hs}, da). \tag{22}$$

Capital is held by individuals, so the market clearing condition for savings is

$$\sum_{b,\mathrm{x},j} \int s_{j+1}^{\mathrm{x}}(a,b) \Pr(b, \mathrm{x}, da) = K. \tag{23}$$

Finally, the aggregate resource constraint is

$$\sum_{b,\mathrm{x},j} \int c_j^{\mathrm{x}}(a,b) \Pr(b, \mathrm{x}, da) + \delta K + e\sum_b \int \Pr(b, \mathrm{co}, da) = F(K, L^{\mathrm{co}}, L^{\mathrm{hs}}), \tag{24}$$

where the first element on the left-hand side is aggregate consumption, the second element is aggregate investment, and the third is aggregate educational spending. The right-hand side is gross domestic product.

Stationary equilibrium: An equilibrium consists of allocations for individuals, $\{c_j^{\mathrm{x}}(a,b), s_j^{\mathrm{x}}(a,b)\}$ and $\{a_i^*, \hat{a}_i\}$, allocations for the representative firm, $\{K, L^{\mathrm{co}}, L^{\mathrm{hs}}\}$, prices $\{r, w^{\mathrm{co}}, w^{\mathrm{hs}}\}$, and a measure $\Pr(b, \mathrm{x}, da)$, such that:

1. The firm optimizes given prices: $\{K, L^{\mathrm{co}}, L^{\mathrm{hs}}\}$ solve problem (1).
2. Individuals optimize given prices: $\{c_j^{\mathrm{x}}(a,b), s_j^{\mathrm{x}}(a,b)\}$ and $\{a_i^*, \hat{a}_i\}$ solve problem (2).
3. The measure $\Pr(b, \mathrm{x}, da)$ is consistent with individual choices: Equations (18)-(21) hold.
4. Markets are in equilibrium: Equations (22)-(24) hold.

[6] The sequence of period budget constraints associated with the lifetime budget constraint of Equation (4) are

$$\begin{aligned} \mathrm{x} = \mathrm{co} \quad &: \quad c_j^{\mathrm{co}}(a,b) + s_{j+1}^{\mathrm{co}}(a,b) + e\mathbb{I}_{j=1} = aw^{\mathrm{co}}\mathbb{I}_{j>S} + (1+r)s_j^{\mathrm{co}}(a,b), \\ \mathrm{x} = \mathrm{hs} \quad &: \quad c_j^{\mathrm{hs}}(a,b) + s_{j+1}^{\mathrm{hs}}(a,b) = aw^{\mathrm{hs}} + (1+r)s_j^{\mathrm{hs}}(a,b). \end{aligned}$$

2.5 Intergenerational persistence

We now discuss the persistence of discount factors and college attainment across generations. We adopt the following terminology. We use the term "upward mobility" to refer to the probability that adults are patient given that their parents are not, $\Pr(b'_H|b_L)$, and to refer to the probability that adults are college graduates given that their parents are not, $\Pr(\text{co}'|\text{hs})$. Similarly, we use the term "horizontal mobility" to refer to the probability that adults are patient given that their parents are also patient, $\Pr(b'_H|b_H)$, and to refer to the probability that adults are college graduates given that their parents also are, $\Pr(\text{co}'|\text{co})$. Persistence, in either variable, is the difference between horizontal and upward mobility.[7] So,

$$\begin{aligned}\text{Persistence in discount factors} &= \Pr(b'_H|b_H) - \Pr(b'_H|b_L),\\ \text{Persistence in college attainment} &= \Pr(\text{co}'|\text{co}) - \Pr(\text{co}'|\text{hs}).\end{aligned}$$

Recall that P_i is the probability that b_i-adults educate their children. It follows that

$$\Pr(b'_H|b_i) = 1-\mu+\mu P_i, \tag{25}$$
$$\Pr(b'_L|b_i) = \mu(1-P_i). \tag{26}$$

The first equation indicates that children of b_i-adults become b_H-adults if they are either b_H-born (probability $1-\mu$) or if they are b_L-born (probability μ) and receive education (probability P_i). The second equation indicates that children of b_i-adults become b_L-adults if they are b_L-born and not educated. Since $\hat{a}_H \leq \hat{a}_L$, it follows that $P_H \geq P_L$ and

$$\Pr(b'_H|b_H) - \Pr(b'_H|b_L) \geq 0, \tag{27}$$

Thus, patience is a persistent trait across generations.

Our main interest, however, is in the persistence of college attainment. We derive the transition matrix mapping the distribution of endogenous states for adults of a given generation, (b, x), into the distribution of endogenous states for adults of the next generation, (b', x'), which we denote $\Pr(b', \text{x}'|b, \text{x})$, in Equations (A.11)-(A.14) of Appendix A.4. The probability that children graduate college given their parents' states are

$$\Pr(\text{co}'|b, \text{x}) = \sum_{b'} \Pr(b', \text{co}'|b, \text{x}).$$

The horizontal mobility of college is then

$$\Pr(\text{co}'|\text{co}) = \frac{\Pr(\text{co}', \text{co})}{\Pr(\text{co})} = \frac{\sum_b \Pr(\text{co}'|b, \text{co}) \Pr(b, \text{co})}{\Pr(\text{co})} \tag{28}$$

[7] A proof of this statement is available upon request.

and the upward mobility of college is

$$\Pr(\text{co}'|\text{hs}) = \frac{\Pr(\text{co}', \text{hs})}{\Pr(\text{hs})} = \frac{\sum_b \Pr(\text{co}'|b, \text{hs}) \Pr(b, \text{hs})}{\Pr(\text{hs})}, \tag{29}$$

where $\Pr(\text{co}', \text{co})$ is the proportion of families where adults and their parents are both college graduates, $\Pr(\text{co}', \text{hs})$ is the proportion of families where adults are college graduates and their parents are not, and $\Pr(\text{x})$ is the proportion of x-graduate, that is $\Pr(\text{x}) = \sum_b \Pr(b, \text{x})$.

2.6 Discussion

Proposition 1 – Persistence of college attainment

$$\Pr(\text{co}'|\text{co}) - \Pr(\text{co}'|\text{hs}) = \underbrace{[F(\mathbb{B}_L) - F(\mathbb{B}_H)]}_{\text{negative or zero}} \times \text{negative number} \geq 0. \tag{30}$$

Proof – See Lemma 11.

Proposition 1 is our main result: The children of college graduates are at least as likely to graduate college as the children of high school graduates. In other words, college attainment is persistent across generations.

It is immediate, from Equations (17) and (25) that

$$P_L = P_H \;\Leftrightarrow F(\mathbb{B}_L) = F(\mathbb{B}_H) \text{ and } \Pr(b'_H|b_H) = \Pr(b'_H|b_L).$$

It follows, therefore, that

$$\Pr(\text{co}'|\text{co}) > \Pr(\text{co}'|\text{hs}) \;\Leftrightarrow\; \Pr(b'_H|b_H) > \Pr(b'_H|b_L). \tag{31}$$

That is, there is persistence in college attainment if and only if there is persistence in patience. This observation leads to the following conclusion: In data generated by our model, an observer (not observing patience) could conclude that family fixed effects matter for explaining college attainment. Such fixed effects would, in fact, be the family's persistent discount factor. We interpret this conclusion as a manifestation of Cameron and Heckman (1998)'s statement about the importance of family factors in college.

Two mechanisms are key to generate persistence in our model: First, more patient individuals are more likely to graduate college ($a^*_H \leq a^*_L$) and, second, more patient individuals are more likely to educate their children to be patient ($\hat{a}_H \leq \hat{a}_L$). The first mechanism arises because of the utility cost of college, ψ. Since utility cannot be traded intertemporally, the timing of utility flows matter and, hence, patience matters. The critical aspect of the theory, however, is not the utility cost itself; it is the impediment to intertemporal trade that it represents. Any other type of impediment

to intertemporal trade, such as borrowing constraints, would similarly imply that more patient individuals are more likely to graduate college. The second mechanism arises because parents are altruistic and value the future welfare of their children.

It is worth discussing the "special cases" in which there is no persistence, that is whenever $F(\mathbb{B}_L) = F(\mathbb{B}_H)$. There are four cases.

1. Suppose that there is either no utility cost of college, that is $\psi = 0$, or no heterogeneity in β_i, that is $\beta_L = \beta_H \equiv \beta$. Equation (12) implies $a_L^* = a_H^*$. It is then trivial, from Equation (16), that $\mathbb{B}_i = \emptyset$ and that $F(\mathbb{B}_L) - F(\mathbb{B}_H) = 0$.

   Patience is irrelevant to the college decision. Some parents may care more about their children than others, that is $\delta_H > \delta_L$. Such parents are more likely to educate their children, but this education has no bearing on the future educational attainment of the children. The b_L-born children receiving parental education enter adulthood with discount factors (β, δ_H) while the others enter adulthood with (β, δ_L). The former will care more about their own children (the 3rd generation in this discussion) than the latter, but both will make the same college decision (conditional on ability). Hence, the absence of persistence.

2. Suppose that there is no heterogeneity in δ_i, that is $\delta_L = \delta_H \equiv \delta$. Equation (15) implies $\hat{a}_L = \hat{a}_H$. Then Equation (16) implies $\mathbb{B}_L = \mathbb{B}_H$ and $F(\mathbb{B}_L) - F(\mathbb{B}_H) = 0$.

   All parents are equally altruistic toward their children, even though some parents may be more impatient than others about their own lifetime utils. Equal altruism toward children imply that the incentives to educate children to induce them to be college graduates are the same for college and high school graduate parents. Hence, the absence of persistence.

3. Suppose that $\hat{a}_L \leq a_H^*$. That is the case represented in Panel B of Figure 2 and in the fourth column of Table 3. It is immediate that $F(\mathbb{B}_L) - F(\mathbb{B}_H) = 0$.

   In this instance, the costs χ/δ_i are low enough that parents with the lowest discount factors educate even children that the know will not graduate college because of their low ability. The same is then necessarily true for parents with the high discount factors. It follows that the set of children that are induced to graduate college by their parents' education is the same regardless of the parents' type. Hence, the absence of persistence.

4. Suppose that $\hat{a}_H \geq a_L^*$. That is the case represented in the first column of Table 3. Here, $\mathbb{B}_i = \emptyset$ and $F(\mathbb{B}_L) - F(\mathbb{B}_H) = 0$.

   In this instance, the costs χ/δ_i of educating children are so high that only children who would

have graduate college regardless of their discount factors receive parental education. Hence, the absence of persistence.

The analysis to this point has characterized the key features determining intergenerational persistence in college attainment. Next, we investigate whether the model can be in line, quantitatively, with observed persistence as well as other pertinent moments such as the overall rate of college attainment and earnings dispersion within and between college and high school graduates.

3 Quantitative analysis

3.1 Calibration

We calibrate the model to U.S. data. Given the life-cycle and intergenerational features of the model, we use both longitudinal data and empirical estimates. Persistence in college attainment represents the key moment in our analysis, and our estimates of this are taken from Gonzalez and Langot (2022). In their work, the last available cohort corresponds to a 1982-1986 birth year. Thus, throughout the calibration we focus on a mid-1980s cohort. In addition, we only consider individuals who completed at least high school. We let the label "college graduate" denote the subset who completed at least a bachelor's degree and the label "high school graduate" denote the others.

We set a period to one year. We set childhood to last 20 years ($A = 20$), the working life to last 40 years ($J = 40$), and college to require four years ($S = 4$). Following Cooley and Prescott (1995), we set the rate of capital depreciation to 4.7% ($\delta = 0.047$) and the capital share to 40% ($\alpha = 0.40$). Following figures reported by Goldin and Katz (2007), we set the elasticity of substitution between college and high school labor to 1.5 ($\rho = 0.33$).

We assume that parents value their children's future utility as much as they value their own utility. That is,

$$\delta_i = \beta_i^{A-1}, \ i \in \{L, H\}.$$

This assumption simplifies the calibration by removing any potential heterogeneity in altruism (outside of the already existing heterogeneity in β.)

The ability distribution is assumed to be log-normal with mean 1:

$$\ln a \sim N(-\sigma_a^2/2, \sigma_a^2).$$

Following Becker and Mulligan (1997), we assume that all children are born impatient, i.e., with the low discount factor ($\mu = 1$). Becker and Mulligan (1997) assume that an individual's endowed discount factor—the birth discount factor in our model—is the lowest discount factor the individual

experiences over their life cycle. They argue that this lowest discount factor represents an imperfect ability to imagine the future, which is a typical trait of children.

This implies nine remaining parameters,

$$\theta \equiv \left[z^{\text{co}}, z^{\text{hs}}, e, \sigma_a, \psi, \chi, \sigma, \beta_L, \beta_H\right],$$

which are calibrated to match nine moments:

1. Pr(co): The proportion of college graduate workers for the mid-1980s cohort.
   We use census data from IPUMS.[8] Among employed individuals of age 28-32 in 2014 (born 1982-1986) that have at least a high school degree, 42% of them are college graduates.

2. Pr(co′|co): The horizontal mobility of college attainment for the mid-1980s cohort.
   We use 64% from Gonzalez and Langot (2022, Table 22).

3. Pr(co′|hs): The upward mobility of college attainment for the mid-1980s cohort.
   We use 24% from Gonzalez and Langot (2022, Table 22).

4. CP: The (lifetime) college premium

   $$\text{CP} = \frac{w^{\text{co}}}{w^{\text{hs}}} \times \frac{D^{\text{co}}}{D^{\text{hs}}} \times \frac{E(a|\text{co})}{E(a|\text{hs})},$$

   where D^{hs} and D^{co} are defined in Equations (7) and (8).
   Using cross-sectional census data from IPUMS, we construct lifetime earnings for employed workers ages 24 to 65 in 2004 (when the mid-1980s cohort is 20 years-old). Using $r = 3\%$ to compute the present value of earnings at 20 implies a college premium of 1.79.

5. CC: The cost of college relative to lifetime earnings is

   $$CC = \frac{e}{w^{\text{co}} \times D^{\text{co}} \times E(a|\text{co})}.$$

   Lifetime earnings are calculated using the earnings data from IPUMS. The cost of college is calculated assuming a tuition rate of $\$15,000$ per year for four years, taken from Digest of Education Statistics. We find that tuition represents 4.8% of the present value of college earnings.

6. CV^{co}: The coefficient of variation of earnings for college graduates.
   We use CPS data in the year 2014 for employed college graduates aged 25 to 35. We find a target figure of 0.60.

[8] IPUMS USA, University of Minnesota, www.ipums.org

7. CV^{hs}: The coefficient of variation of earnings for high school graduates.
   We use CPS data in the year 2014 for employed high school graduates aged 25 to 35. We find a target figure of 0.55.

8. $\bar{\beta}$: The average discount factor in the adult population

$$\bar{\beta} = \sum_i \Pr(b_i)\beta_i.$$

   We target a standard value $\bar{\beta} = 0.95$ (Cooley and Prescott, 1995).

9. KY: Capital-output ratio: We target a value of 3.3 (Cooley and Prescott, 1995).

Define

$$\Theta(\theta) = \begin{bmatrix} \Pr(\text{co})/0.42 \\ \Pr(\text{co}'|\text{co})/0.64 \\ \Pr(\text{co}'|\text{hs})/0.24 \\ \text{CP}/1.79 \\ \text{CC}/0.048 \\ \text{CV}^{\text{co}}/0.60 \\ \text{CV}^{\text{hs}}/0.55 \\ \bar{\beta}/0.95 \\ \text{KY}/3.3 \end{bmatrix}$$

and calibrate the model by solving

$$\min_{\theta} \ [\Theta(\theta) - 1]' [\Theta(\theta) - 1].$$

Even though the calibrated parameters are determined simultaneously, some targets are particularly informative for specific parameters. For instance, the variance of ability, σ_a^2, has a first order effect on the dispersion of income and, hence, it is informed by inequality-related moments such as the two coefficients of variation and the college premium. The college cost, e, is informed by the proportion of college graduates as well as the relative cost of college. The utility cost of college, ψ, is also informed by the proportion of college graduates. The level of discount factors, β_L and β_H, is informed by $\bar{\beta}$. This target ensures that our model is broadly consistent with the macroeconomics literature. It is worth noting that the figure $\bar{\beta} = 0.95$, used in Cooley and Prescott (1995), is determined in such a way that their representative agent model yields a capital-output ratio consistent with data, i.e., 3.30. Our model, however, is not a representative agent model, and the unique Euler equation used by Cooley and Prescott (1995) to calibrate their discount factor does not exist in our model. Instead, our model has as many Euler equations as there are combinations of age, ability, graduation status, and discount factor. It follows that using the same depreciation rates, capital share, and average discount factor as Cooley and Prescott (1995) does

not imply that our model's capita-to-output ratio should be consistent with data. It is for this reason that the capital-to-output ratio is added to the list of targets.

Finally, the differences between β_L and β_H, and the utility cost of education, χ, have first order effects on the intergenerational persistence in educational attainment. Specifically, we have shown (Equation 30) that when the discount factors are equal, the persistence in college attainment is zero. Thus, the stability in college attainment, $\Pr(\text{co}'|\text{co})$, and the upward mobility in college attainment $\Pr(\text{co}'|\text{hs})$, inform the discount factors and the utility cost of education (i.e. teaching patience).

3.2 Results

Table 1 shows the calibrated parameters and Table 2 shows the model's fit to the calibration target. The model's ability to capture the observed persistence in college attainment, $\Pr(\text{co}'|\text{co})-\Pr(\text{co}'|\text{hs})$, is of first order importance. Here the model does very well, and in fact generates mildly more persistence than the data. The ranking of ability thresholds in the calibrated equilibrium is

$$\hat{a}_H < a_H^* < a_L^* < \hat{a}_L,$$

which is represented in Figure 3, and corresponds to the sixth case in Table 3 (Appendix A.5). In this configuration, b_L-parents do not affect the college decision of their children: The ability thresholds at which they educated their children, $\hat{a}_L$, is higher than the ability thresholds at which the children, even with the low discount factor, would choose college, a_L^*. Thus, b_L-parents only educate children that have would gone to college regardless. Formally, $\mathbb{B}_L = \emptyset$. In contrast, b_H-parents affect the education decision of some of their children: Children with ability in the interval $[a_H^*, a_L^*]$ become college graduates only if their parents teach them patience. That is the case since $\hat{a}_H < a_H^*$. It follows that $\mathbb{B}_H = [a_H^*, a_L^*]$.

It is worth noting that persistence arises despite our assumption that there is no correlation between parent's ability and their children's. We made that assumption for two reasons. First, it allows us to illustrate the mechanism in our model. Second, it is in line with evidence reported in the literature. For instance, Cameron and Heckman (1998) measure the correlation between a child and their mother's college attainment after controlling for ability, family income, and the father's college attainment. They find that correlation to be significant, indicating that college persistence exists beyond the effect of transmitted abilities within families. Similarly, Sacerdote (2000) studies samples of adopted children where the correlation of innate ability does not exist by construction. He finds that *"children adopted by a mother with a college degree have a* 40 *percent higher chance of graduating college themselves compared to children adopted by a mother without a college degree."* We note that 40% is of similar magnitude as $\Pr(\text{co}'|\text{co}) - \Pr(\text{co}'|\text{hs})$ in the data we use, pertaining to the entire U.S. population (not just adopted children), from Gonzalez and Langot (2022).

While close in absolute value to the coefficient of variation targets, the model fails to generate

Table 1: Calibrated parameters

Parameters	Value	Description
σ	0.938	CRRA preference parameter
β_L	0.950	low discount factor
β_H	0.978	high discount factor
ψ	1.436	utility cost of college
χ	75.286	utility cost of education
e	1.757	good cost of college
z^{co}	1.515	college-biased productivity
z^{hs}	0.940	high school-biased productivity
σ_a^2	0.401	variance of ability

Table 2: Model fit

	Moment	Source	Data	Model	Notation
1	College graduates, %	a	42.0	42.0	Pr(co)
2	Horizontal mobility, %	b	64.0	67.8	Pr(co'|co)
3	Upward mobility, %	b	24.0	23.3	Pr(co'|hs)
4	College premium	a	1.79	1.82	CP
5	Cost of college, %	a & d	4.80	4.76	CC
6	Coef. of var.: col. earnings	c	0.60	0.55	CV^{co}
7	Coef. of var.: hs earnings	c	0.55	0.57	CV^{hs}
8	Average discount factor		0.95	0.97	$\bar{\beta}$
9	Capital-output ratio		3.30	3.06	KY

Source: (a) IPUMS USA; (b) Gonzalez and Langot (2022, Table 22); (c) CPS; (d) Digest of Education Statistics.

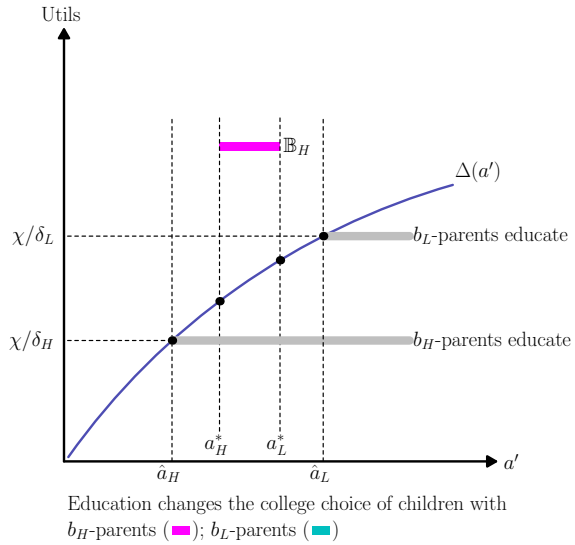


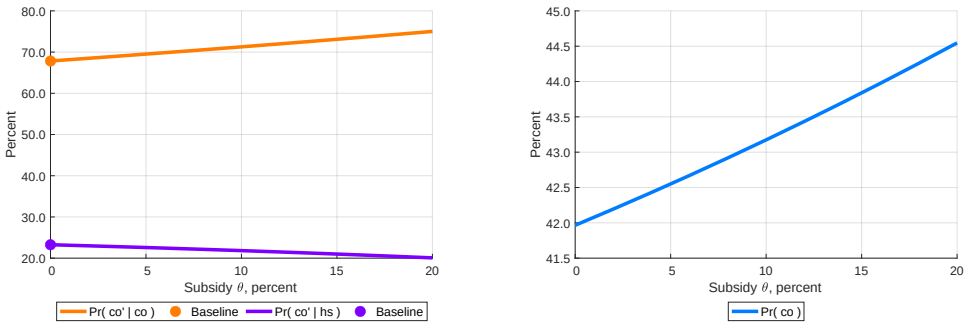
Figure 3: Ranking of ability thresholds in the calibrated model

enough relative difference in the wage dispersion among college graduates compared to high-school graduates. Fine-tuning wage dispersion is relatively difficult in our model, as there only exist two potential sources of earnings heterogeneity: (i) in ability via the distribution $F(a)$, and (ii) in college attainment. Despite this slight miss in targets, simultaneously hitting the college premium and generating the appropriate level of earnings variation shows the model's ability to match these features of the data. To further emphasize this point, we simulated the calibrated model and computed the implied intergenerational persistence in earnings, measured as the slope of the regression:

$$\ln(\text{earnings of children}) = \text{constant} + \text{slope} \times \ln(\text{earnings of parents}) + \text{error}$$

The resulting slope is 4.3%, which is significantly lower than the figures commonly reported in the literature, i.e., 40% (e.g., Restuccia and Urrutia, 2004, Lee and Seshadri, 2019). It is important to note, however, that we have assumed away persistence in innate ability.

Because our model is calibrated to match the cross-sectional variation in annual earnings, the low intergenerational correlation in earnings is perhaps not surprising. If the variance in the ability distribution, σ_a, were reduced, a larger proportion of earnings variation would result from the college decision (i.e. from the difference in w^{co} and w^{hs}). In this case, the intergenerational persistence in earnings would align more closely with intergenerational persistence in college attainment. Of course, such a change would lead the model to generate much less cross-sectional variation in earnings.



Persistence College attainment

Figure 4: The effect of college subsidies on college persistence and attainment

Overall, the model performs very well matching the data on intergenerational persistence while being consistent with observed cross-sectional inequality. Next we explore policy implications of the calibrated model.

3.3 Policy experiment

Our model is suited to study the implications of government policies toward college attainment. We analyze a college subsidy, θ, financed via a proportional tax, τ, levied on all wages. In this experiment college graduates pay only $(1-\theta)e$ instead of the full tuition e. We determine the tax rate such that, in equilibrium, the government's budget constraint holds:

$$\tau\left(w^{\text{co}}L^{\text{co}}+w^{\text{hs}}L^{\text{hs}}\right)=\theta e\sum_{b}\int \Pr(b,\text{co},da).$$

We compute equilibria for values of the college subsidy ranging from 0 (the baseline) to 20%. Figure 4 shows the resulting intergenerational mobility and college attainment. The main findings are the rise of college attainment for those whose parents are college graduates (Panel A), the reduction of college attainment for the others (Panel A), and the rise of overall college attainment (Panel B). In sum, intergenerational persistence measured by $\Pr(\text{co}'|\text{co})-\Pr(\text{co}'|\text{hs})$ increases.

To understand the increase of intergenerational persistence, recall, from Equation (30), that

$$\Pr(\text{co}'|\text{co})-\Pr(\text{co}'|\text{hs})=[F(\mathbb{B}_L)-F(\mathbb{B}_H)]\ \times\ \text{negative number}.$$

We find that the increase in $\Pr(\text{co}'|\text{co})-\Pr(\text{co}'|\text{hs})$ is entirely driven by the term $F(\mathbb{B}_L)-F(\mathbb{B}_H)$. So, we concentrate the remainder of the discussion on this term.[9] We also find that the ranking

[9] As the subsidy increases from 0 to 20%, the first term in $\Pr(\text{co}'|\text{co})-\Pr(\text{co}'|\text{hs})$ changes by 10% while the second term changes by 1% in the opposite direction. Thus, the changes in the second term is of negligible magnitude and

of the a_i^*s and $\hat{a}_i$s in the equilibria with subsidy remains the same as in the baseline (recall that it is $\hat{a}_H < a_H^* < a_L^* < \hat{a}_L$). That is, only patient parents affect the college decision of their children when these children have ability in the interval $[a_H^*, a_L^*]$. Thus, the increase in persistence results from the widening of this interval. In other words, as the subsidy increases, patient parents become more likely to have children that they can induce to graduate college. Since patient parents are more likely to be college graduates themselves, the rise in intergenerational persistence follows.

So, why does the interval $[a_H^*, a_L^*]$ widen? Let $\tilde{e} \equiv e(1-\theta)$ represent the college cost net of the subsidy, and let $\tilde{w}^{\mathrm{x}} \equiv w^{\mathrm{x}}(1-\tau)$ represent after-tax wages. As the subsidy increases, we find that college attainment increases as well. General equilibrium forces imply that (i) $\tilde{w}^{\mathrm{co}}$ decreases because w^{co} decreases with the rise in college attainment and because τ increases; (ii) $\tilde{w}^{\mathrm{hs}}$ increases because w^{hs} increases enough with the rise in college attainment to offset the effect of τ; (iii) the interest rate decreases negligibly.

We define $\Gamma(a,b) \equiv \mathcal{V}^{\mathrm{co}}(a,b) - \mathcal{V}^{\mathrm{hs}}(a,b)$ such that a_i^* satisfies

$$\Gamma(a_i^*, b_i) = 0,$$

and analyze it in Lemma 3. We then approximate the effect of the policy as the direct effect of the subsidy (i.e., $d\tilde{e} < 0$) plus the equilibrium effects of wages (i.e., $d\tilde{w}^{\mathrm{co}} < 0$ and $d\tilde{w}^{\mathrm{hs}} > 0$), abstracting from the negligible effect of the interest rate (i.e., $dr \simeq 0$):

$$d\Gamma(a_i^*, b_i) \propto \underbrace{-I^{\mathrm{co}}(a_i^*)^{-\sigma} d\tilde{e}}_{\text{Direct subsidy effect} > 0} + a_i^* \times \underbrace{\left[I^{\mathrm{co}}(a_i^*)^{-\sigma} D^{\mathrm{co}} d\tilde{w}^{\mathrm{co}} - I^{\mathrm{hs}}(a_i^*)^{-\sigma} D^{\mathrm{hs}} d\tilde{w}^{\mathrm{hs}} \right]}_{\text{Equilibrium lifetime income effect} < 0}.$$

We make several observations. First, the subsidy effect is positive: It shifts the Γ schedule upward and, hence, tends to reduce a_i^* because Γ is monotonically increasing in a (Lemma 3). Second, the equilibrium effect is negative: It shifts the Γ schedule downward and, hence, tends to raise a_i^*. That is because, in equilibrium, the after-tax wage rate of college graduates decrease while that of high school graduate increase. So, the policy makes college cheaper on the one hand, but relatively less rewarding on the other hand. It follows that it may induce or deter individuals close the a_i^* threshold to become college graduates, depending on which one of the two effects dominate. Third, the subsidy effect is relatively negligible when a_i^* is large while the equilibrium effect is relatively negligible as a_i^* approaches zero. That is because the subsidy affects income in a manner unrelated to ability, while wages affect incomes in a manner proportional to ability. Put differently, the negative effect of the policy dominates at high enough ability while the positive effect dominates at low enough ability. Recall now that $a_H^* < a_L^*$ (Lemma 6) and so, it is possible (and indeed the case in our quantitative exercise) that $d\Gamma(a_H^*, b_H) > 0$ and $d\Gamma(a_L^*, b_L) < 0$, implying that the policy lowers a_H^* and raises a_L^* and, therefore, raises persistence.

acting in the direction opposite the increase of $\Pr(\mathrm{co}'|\mathrm{co}) - \Pr(\mathrm{co}'|\mathrm{hs})$.

The economics of these mechanisms can be summarized as follows. High-ability marginal individuals are impatient (i.e., a_L^* is "high"). As a result, the reduced lifetime income benefit of college is insufficient to offset the reduction in tuition, leading to a decline in the graduation rate of impatient individuals under the policy. Given the persistence of discount factors, these individuals are likely to have impatient parents, who are themselves often high school graduates. Consequently, upward mobility decreases. A similar analysis applies to patient individuals.

Conversely, low-ability marginal individuals are patient (i.e., a_H^* is "low"). For this group, the reduced tuition offsets the lower lifetime income benefit of college, making them better off under the policy and increasing their graduation rate. Since their parents are likely patient college graduates, horizontal mobility increases.

Our results differ from those reported in Restuccia and Urrutia (2004, Table 8) and Lee and Seshadri (2019, Figure 13). These authors find that subsidies can reduce intergenerational persistence by alleviating budget constraints. In our model, subsidies for college do increases college attainment in equilibrium; however, they also increase the persistence of college. Parents with a college degree are more likely to have children with a college degree and parents with only a high school diploma are more likely to have children with a high school diploma. Access to college increases, but only for those patient individuals. Thus, inequality in college attainment actually increases. This suggests that policies aimed at increasing incentives for parents to educate their children, which would increase the fraction of patient individuals in the economy, would be more effective at increasing access to college.

4 Conclusion

We present a novel theory of the persistence in college attainment. Despite the absence of persistence in innate ability, parental transfers, and borrowing constraints, persistence in college attainment arises from parents' desire to educate their children to be patient. The most patient parents are the most likely to have a college degree and they value their children's welfare the most. Hence, they are the most likely to induce them to attain a college degree by teaching them patience. Theoretically, persistence in college attainment may or may not obtain, depending on the differences in discount factors relative to the utility cost of educating patience for parents. Quantitatively, our model generates the observed persistence in college attainment and other moments pertaining to income inequality. Our model also provides a theory for a pervasive finding in the empirical literature examining the persistence in college attainment: The importance of long-run family factors in determining college attainment. Policy experiments show that a college subsidy to reduce the pecuniary cost of college does increase total college attainment, but it does so while increasing persistence and, hence, inequality at the same time. This finding differs from findings in the previous literature where college subsidies alleviate borrowing constraints and, hence, reduce college persistence. There is, thus, an interesting avenue for future research: Which effect dominates when

a subsidy is in place and there are both borrowing constraint to overcome and individuals with heterogeneous rates of time preferences?

We abstracted from persistence in innate ability, parental transfers, and borrowing constraints for illustrating the potential of our theory. Introducing such features into our model constitute an avenue for future research. We think that introducing a borrowing constraint into the model is of first-order interest. A borrowing constraint would interact in non-trivial ways with patience heterogeneity: Conditional on ability (and/or family wealth) the most impatient individuals are the most likely to be constrained. Such mechanism would not exist in models with borrowing constraint but no patience heterogeneity.

References

G. Albanese, G. De Blasio, and P. Sestito. My parents taught me. evidence on the family transmission of values. *Journal of Population Economics*, 29:571–592, 2016.

S. Angerer, J. Bolvashenkova, D. Glätzle-Rützler, P. Lergetporer, and M. Sutter. Children's patience and school-track choices several years later: Linking experimental and field data. *Journal of Public Economics*, 220:104837, 2023.

G. S. Becker and C. B. Mulligan. The endogenous determination of time preference. *Quarterly Journal of Economics*, 112(3):729–758, 1997.

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

G. S. Becker and N. Tomes. Human capital and the rise and fall of families. *Journal of labor economics*, 4(3, Part 2):S1–S39, 1986.

G. S. Becker, K. M. Murphy, and J. L. Spenkuch. The manipulation of children's preferences, old-age support, and investment in children's human capital. *Journal of Labor Economics*, 34 (S2):S3–S30, 2016.

A. Bisin and T. Verdier. "beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. *Quarterly Journal of Economics*, 115(3):955–988, 2000.

A. Bisin and T. Verdier. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319, 2001.

A. Bisin and T. Verdier. The economics of cultural transmission and socialization. In *Handbook of social economics*, volume 1, chapter 9, pages 339–416. Elsevier, 2011.

S. V. Cameron and J. J. Heckman. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy*, 106(2):262–333, 1998.

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

T. F. Cooley and E. C. Prescott. Economic growth and business cycles. In T. F. Cooley, editor, *Frontier of Business Cycle Research*, pages 1–38. Princeton University Press, Princeton New Jersey, 1995.

M. De Paola and F. Gioia. Impatience and academic performance. less effort and less ambitious goals. *Journal of Policy Modeling*, 39(3):443–460, 2017.

M. Doepke and F. Zilibotti. Occupational choice and the spirit of capitalism. *Quarterly Journal of Economics*, 123(2):747–793, 2008.

M. Doepke and F. Zilibotti. Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, 85(5):1331–1371, 2017.

R. Fernández. Does culture matter? In *Handbook of social economics*, volume 1, chapter 11, pages 181–510. Elsevier, 2011.

R. Fernández, A. Fogli, and C. Olivetti. Mothers and sons: Preference formation and female labor force dynamics. *Quarterly Journal of Economics*, 119(4):1249–1299, 2004.

J. Fernández-Villaverde, J. Greenwood, and N. Guner. From shame to game in one hundred years: An economic model of the rise in premarital sex and its de-stigmatization. *Journal of the European Economic Association*, 12(1):25–61, 2014.

G. Glomm and B. Ravikumar. Public versus private investment in human capital: endogenous growth and income inequality. *Journal of Political Economy*, 100(4):818–834, 1992.

C. Goldin and L. F. Katz. *The race between education and technology*. Belknap Press, Cambridge, MA, 2007.

B. H. Golsteyn, H. Grönqvist, and L. Lindahl. Adolescent time preferences predict lifetime outcomes. *The Economic Journal*, 124(580):F739–F761, 2014.

J. Gonzalez and F. Langot. Intergenerational Educational and Earnings Mobilities Trends in the U.S. Since the 60s, 2022. hal-03784447.

L. Guiso, P. Sapienza, and L. Zingales. Does culture affect economic outcomes? *Journal of Economic perspectives*, 20(2):23–48, 2006.

E. A. Hanushek, L. Kinne, P. Lergetporer, and L. Woessmann. Patience, risk-taking, and human capital investment across countries. *The Economic Journal*, 132(646):2290–2307, 2022.

E. A. Hanushek, L. Kinne, P. Sancassani, and L. Woessmann. Can patience account for subnational differences in student achievement? regional analysis with facebook interests. *NBER Working Paper Series*, (31690), 2023.

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

J. J. Heckman, L. Lochner, and P. E. Todd. Fifty years of mincer earnings regressions. *NBER Working Paper Series*, (31690), 2003.

M. P. Keane and K. I. Wolpin. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, 42(4):1051–1103, 2001.

L. J. Kotlikoff and L. H. Summers. The role of intergenerational transfers in aggregate capital accumulation. *Journal of Political Economy*, 89(4):706–732, 1981.

S. Y. Lee and A. Seshadri. On the intergenerational transmission of economic status. *Journal of Political Economy*, 127(2):855–921, 2019.

L. Lochner and A. Monge-Naranjo. Credit constraints in education. *Annual Review of Econonmics*, 4:225–256, 2012.

A. L. Owen and D. N. Weil. Intergenerational earnings mobility, inequality and growth. *Journal of Monetary Economics*, 41(1):71–104, 1998.

D. Restuccia and C. Urrutia. Intergenerational persistence of earnings: The role of early and college education. *American Economic Review*, 94(5):1354–1378, 2004.

S. Rosen. The value of changes in life expectancy. *Journal of Risk and uncertainty*, 1(3):285–304, 1988.

B. Sacerdote. The nature and nurture of economic outcomes. *NBER Working Paper Series*, (7949), 2000.

B. Sacerdote. How large are the effects from changes in family environment? a study of korean american adoptees. *Quarterly Journal of Economics*, 122(1):119–157, 2007.

B. Sacerdote. Nature and nurture effects on children's outcomes: what have we learned from studies of twins and adoptees? In *Handbook of social economics*, volume 1, chapter 1, pages 1–30. Elsevier, 2011.

M. Saez-Marti and Y. Zenou. Cultural transmission and discrimination. *Journal of Urban Economics*, 72(2-3):137–146, 2012.

U. Sunde, T. Dohmen, B. Enke, A. Falk, D. Huffman, and G. Meyerheim. Patience and comparative development. *The Review of Economic Studies*, 89(5):2806–2840, 2022.

M. Sutter, M. G. Kocher, D. Glätzle-Rützler, and S. T. Trautmann. Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review*, 103(1): 510–531, 2013.

G. Vandenbroucke. The baby boomers and the productivity slowdown. *European Economic Review*, 132:103609, 2021.

A Model solution

A.1 Setup and utility maximization

Lemma 1. The ratio $I^{\text{co}}/I^{\text{hs}}$

Assume $e > 0$. Then,

$$\frac{I^{\text{co}}(a)}{I^{\text{hs}}(a)} < \frac{w^{\text{co}}/w^{\text{hs}}}{D^{\text{hs}}/D^{\text{co}}}$$

Proof Immediate from Equations (7) and (8):

$$\frac{I^{\text{co}}(a)}{I^{\text{hs}}(a)} = -\frac{e}{I^{\text{hs}}(a)} + \frac{aw^{\text{co}}D^{\text{co}}}{aw^{\text{hs}}D^{\text{hs}}} < \frac{w^{\text{co}}/w^{\text{hs}}}{D^{\text{hs}}/D^{\text{co}}}.$$

■

Standard calculations yield the following solution for problem (4)

$$\mathcal{U}^{\text{x}}(a,b) = \Lambda(\beta)I^{\text{x}}(a)^{1-\sigma} \quad \text{where} \quad \Lambda(\beta) = \frac{1}{1-\sigma}\left(\frac{1-\beta^{1/\sigma}\left(1+r\right)^{1/\sigma-1}}{1-\left[\beta^{1/\sigma}\left(1+r\right)^{1/\sigma-1}\right]^{J}}\right)^{-\sigma}. \tag{A.1}$$

Lemma 2. The functions Λ and $\mathcal{U}^{\text{x}}$

Assume $\sigma \in (0,1)$. Then,

$$\Lambda(\beta) > 0,\ \mathcal{U}^{\text{x}}(a,b) > 0,\ \partial\Lambda(\beta)/\partial\beta > 0,\ \text{and}\ \partial\mathcal{U}^{\text{x}}(a,b)/\partial\beta > 0.$$

Proof Since U is positive, $\mathcal{U}^{\text{x}}(a,b)$ is positive as well. Then, since income is positive, Equation (A.1) implies $\Lambda(\beta)$ is positive. Furthermore,

$$\frac{\partial\mathcal{U}^{\text{x}}\left(a,\beta\right)}{\partial\beta} = \sum_{j=2}^{J}\left(j-1\right)\beta^{j-2}U\left(c_j\right) = \frac{\partial\Lambda(\beta)}{\partial\beta}I^{\text{x}}(a)^{1-\sigma}$$

where the middle-term is positive. The first equality follows from the envelope theorem. The second equality follows from Equation (A.1). Since income is positive, the results follow. ■

A.2 College decision

The college decision, Equation (2), is characterized by a_i^* such that

$$\Gamma(a_i^*, b_i) = 0 \quad \text{where} \quad \Gamma(a,b) \equiv \mathcal{V}^{\text{co}}(a,b) - \mathcal{V}^{\text{hs}}(a,b) = 0.$$

Lemma 3. The function Γ

Assume $\sigma \in (0,1)$. Then,

$$\partial\Gamma(a,b)/\partial a > 0.$$

Proof

Using Equation (A.1)

$$\Gamma(a,b) = \Lambda(\beta) I^{\text{co}}(a)^{1-\sigma} - \Lambda(\beta) I^{\text{hs}}(a)^{1-\sigma} - \psi.$$

Note that

$$\begin{aligned}\frac{\partial\Gamma(a,b)}{\partial a} &= \Lambda(\beta)(1-\sigma)\left[I^{\text{co}}(a)^{-\sigma} w^{\text{co}} D^{\text{co}} - I^{\text{hs}}(a)^{-\sigma} w^{\text{hs}} D^{\text{hs}}\right], \\ &= \Lambda(\beta)(1-\sigma) I^{\text{hs}}(a)^{-\sigma} w^{\text{hs}} D^{\text{hs}} \left[\left(\frac{I^{\text{co}}(a)}{I^{\text{hs}}(a)}\right)^{-\sigma} \frac{w^{\text{co}}/w^{\text{hs}}}{D^{\text{hs}}/D^{\text{co}}} - 1\right],\end{aligned}$$

where the expression in brackets is decreasing in $I^{\text{co}}(a)/I^{\text{hs}}(a)$. Since the ratio $I^{\text{co}}(a)/I^{\text{hs}}(a)$ is bounded above by $(w^{\text{co}}/w^{\text{hs}})/(D^{\text{hs}}/D^{\text{co}})$ (Lemma 1), the expression in brackets attains a minimum

$$\left(\frac{w^{\text{co}}/w^{\text{hs}}}{D^{\text{hs}}/D^{\text{co}}}\right)^{1-\sigma} - 1 > 0$$

where the inequality follows from $w^{\text{co}} D^{\text{co}} > w^{\text{hs}} D^{\text{hs}}$. (An equilibrium that does not satisfy this condition would feature no college graduates.) Thus,

$$\left(\frac{I^{\text{co}}(a)}{I^{\text{hs}}(a)}\right)^{-\sigma} \frac{w^{\text{co}}/w^{\text{hs}}}{D^{\text{hs}}/D^{\text{co}}} - 1 > 0 \quad \text{for all } a.$$

The results follow from $\Lambda(\beta) > 0$ (Lemma 2). ■

Lemma 4. The decision rule for college

Assume $\sigma \in (0,1)$ and let a_i^* solve $\Gamma(a_i^*, b_i) = 0$. Then, a_i^* exists, it is unique and the college decision for an individual with ability a and discount factor β_i is

$$\begin{cases} \text{college if} & a \geq a_i^*, \\ \text{high school if} & a < a_i^*. \end{cases}$$

Proof Note that $\lim_{a\to 0}\Gamma(a,b) < 0$ because of the utility cost of college. Note also that

$$\lim_{a\to+\infty} I^{\text{co}}(a)^{1-\sigma} - I^{\text{hs}}(a)^{1-\sigma} = \lim_{a\to+\infty} I^{\text{co}}(a)^{1-\sigma}\left(1 - \left(\frac{I^{\text{hs}}(a)}{I^{\text{co}}(a)}\right)\right) = +\infty$$

because $I^{\text{hs}}(a)/I^{\text{co}}(a) < 1$ for a large enough. So, by the intermediate value theorem, there exists a value a_i^* such that $\Gamma(a_i^*, b_i) = 0$. Uniqueness follows from Γ's monotonicity (Lemma 3). Furthermore, since $\partial\Gamma(a,b)/\partial a > 0$, college is optimal for all $a \geq a_i^*$. ■

Lemma 5. The role of ψ

If $\psi = 0$ then $a_i^* = \bar{a}$. If $\psi > 0$ then $a_i^* > \bar{a}$.

Proof Recall that $\Gamma\left(a_i^*, b_i\right) = 0$ implies

$$I^{\text{co}}\left(a_i^*\right)^{1-\sigma} - I^{\text{hs}}\left(a_i^*\right)^{1-\sigma} = \psi/\Lambda\left(\beta_i\right).$$

Then, $I^{\text{co}}(a_i^*) = I^{\text{hs}}(a_i^*)$ whenever $\psi = 0$ and, hence, $a_i^* = \bar{a}$. When $\psi > 0$ the right-hand side is positive. Then, the left-hand side is positive too. Since the left-hand side is increasing in a (Lemma 3), $a_i^* > \bar{a}$ when $\psi > 0$ follows. ■

Lemma 6. Patient people are more likely to graduate college

Assume $\sigma \in (0, 1)$. Then,

$$a_H^* < a_L^* \text{ whenever } \psi > 0 \quad \text{and} \quad a_H^* = a_L^* \text{ whenever } \psi = 0.$$

Proof It is convenient to think of a^* and β as two continuous variables such that

$$\Gamma\left(a^*, b\right) = \Lambda(\beta)\left[I^{\text{co}}\left(a^*\right)^{1-\sigma} - I^{\text{hs}}\left(a^*\right)^{1-\sigma}\right] - \psi = 0.$$

Totally differentiating yields

$$\frac{da^*}{d\beta} = -\frac{\partial\Gamma\left(a^*, b\right)/\partial\beta}{\partial\Gamma\left(a^*, b\right)/\partial a}.$$

where

$$\frac{\partial\Gamma\left(a^*, b\right)}{\partial\beta} = \frac{\partial\Lambda(\beta)}{\partial\beta}\left[I^{\text{co}}\left(a^*\right)^{1-\sigma} - I^{\text{hs}}\left(a^*\right)^{1-\sigma}\right] = \frac{\partial\Lambda(\beta)}{\partial\beta}\frac{\psi}{\Lambda(\beta)}.$$

Thus,

$$\frac{da^*}{d\beta} = -\frac{\psi}{\Lambda(\beta)}\frac{\partial\Lambda(\beta)/\partial\beta}{\partial\Gamma\left(a^*, \beta\right)/\partial a}.$$

Since $\Lambda(\beta) > 0$ and $\partial\Lambda(\beta)/\partial\beta > 0$ (Lemma 2) and since $\partial\Gamma\left(a, \beta\right)/\partial a > 0$ (Lemma 3) it follows that $da^*/d\beta \leq 0$. Then, $\beta_L < \beta_H$ implies $a_H^* \leq a_L^*$ with strict equality whenever $\psi = 0$. ■

A.3 Education decision

Lemma 7. b_H-born children

Parents of b_H-born children do not educate them.

Proof From Equation (A.1)

$$\frac{\partial\mathcal{V}^{\text{x}}(a, b)}{\partial\beta} = \frac{\partial}{\partial\beta}\Lambda(\beta)I^{\text{x}}(a)^{1-\sigma} + \frac{\partial}{\partial\beta}E\left[\mathcal{W}\left(a', b'; b\right)\right]$$

where the first element is positive (Lemma 2). Furthermore, $V(a,b)$ is either $\mathcal{V}^{\text{co}}(a,b)$ or $\mathcal{V}^{\text{hs}}(a,b)$ and so, $\partial V(a,b)/\partial\beta > 0$. It follows that parents of b_H-born children do not change their children's discount factor to b_L. ■

The education decision is characterize by

$$\Delta(\hat{a}_i) = 0$$

where

$$\Delta(a) = \begin{cases} \mathcal{V}^{\text{hs}}(a, b_H) - \mathcal{V}^{\text{hs}}(a, b_L) & \text{when } a < a_H^*, \\ \mathcal{V}^{\text{co}}(a, b_H) - \mathcal{V}^{\text{hs}}(a, b_L) & \text{when } a_H^* \leq a \leq a_L^*, \\ \mathcal{V}^{\text{co}}(a, b_H) - \mathcal{V}^{\text{co}}(a, b_L) & \text{when } a > a_L^*. \end{cases}$$

Lemma 8. The function Δ

$$\partial\Delta(a)/\partial a > 0.$$

Proof From Equation (A.1)

$$\frac{\partial\Delta(a)}{\partial a}\frac{1}{1-\sigma} = \begin{cases} \left[\Lambda(\beta_H) - \Lambda(\beta_L)\right] I^{\text{hs}}(a)^{-\sigma} \frac{\partial I^{\text{hs}}(a)}{\partial a} & \text{when } a' < a_H^*, \\ \Lambda(\beta_H)\, I^{\text{co}}(a)^{-\sigma} \frac{\partial I^{\text{co}}(a)}{\partial a} - \Lambda(\beta_L)\, I^{\text{hs}}(a)^{-\sigma} \frac{\partial I^{\text{hs}}(a)}{\partial a} & \text{when } a_H^* \leq a' \leq a_L^* \\ \left[\Lambda(\beta_H) - \Lambda(\beta_L)\right] I^{\text{co}}(a)^{-\sigma} \frac{\partial I^{\text{co}}(a)}{\partial a} & \text{when } a' > a_L^*. \end{cases}$$

The first and last rows are positive because $\Lambda(\beta_H) \geq \Lambda(\beta_L)$ (Lemma 2). The middle row is

$$\begin{aligned} \frac{\partial\Delta(a)}{\partial a}\frac{1}{1-\sigma} &= \Lambda(\beta_H)\, I^{\text{co}}(a)^{-\sigma} w^{\text{co}} D^{\text{co}} - \Lambda(\beta_L)\, I^{\text{hs}}(a)^{-\sigma} w^{\text{hs}} D^{\text{hs}} \\ &= \Lambda(\beta_L)\, I^{\text{hs}}(a)^{-\sigma} w^{\text{hs}} D^{\text{hs}} \left[\frac{\Lambda(\beta_H)}{\Lambda(\beta_L)} \left(\frac{I^{\text{co}}(a)}{I^{\text{hs}}(a)}\right)^{-\sigma} \frac{w^{\text{co}}/w^{\text{hs}}}{D^{\text{hs}}/D^{\text{co}}} - 1\right] \end{aligned}$$

From the proof of Lemma 3 and the fact that $\Lambda(\beta_H) \geq \Lambda(\beta_L)$ (Lemma 2), the expression in brackets is positive. The result follows. ■

A.4 Stationary distribution

Define

$$E_i^{\text{co}} = F(\{a > \hat{a}_i\} \cap \{a > a_H^*\}) = \min(P_i, C_H), \tag{A.2}$$

$$E_i^{\text{hs}} = F(\{a > \hat{a}_i\} \cap \{a < a_H^*\}) = \max(0, P_i - C_H), \tag{A.3}$$

$$N_i^{\text{co}} = F(\{a < \hat{a}_i\} \cap \{a > a_L^*\}) = \max(0, C_L - P_i), \tag{A.4}$$

$$N_i^{\text{hs}} = F(\{a < \hat{a}_i\} \cap \{a < a_L^*\}) = 1 - \max(P_i, C_L). \tag{A.5}$$

Thus, E_i^{co} is the probability that the b_L-born children of b_i-adults are educated and graduate college. Similarly, N_i^{hs} is the probability that they are not educated and graduate high school, etc.

To simplify notations, it is convenient to refer to the proportion of b_L-adults as M:

$$M \equiv \Pr(b_L),$$

and to note that when M is stationary, it satisfies

$$M = \mu M(1 - P_L) + \mu(1 - M)(1 - P_H), \tag{A.6}$$

where the first element on the right-hand side is the proportion of b_L-born children not educated by b_L-parents. The second element is the proportion of b_L-born children not educated by b_H-parents. We derive the marginal distribution of the endogenous state variables (using Equations (A.2)-(A.5) throughout). From Equation (18),

$$\begin{aligned}
\Pr(b_H, \text{co}) &= \int \Pr(b_H, \text{co}, da), \\
&= (1-\mu) \int_{a_H^*} F(da) + \mu \left[(1-M) \int_{\max\{a_H^*, \hat{a}_H\}} F(da) + M \int_{\max\{a_H^*, \hat{a}_L\}} F(da) \right], \\
&= (1-\mu) C_H + \mu [M E_L^{\text{co}} + (1-M) E_H^{\text{co}}].
\end{aligned} \tag{A.7}$$

From Equation (19),

$$\begin{aligned}
\Pr(b_H, \text{hs}) &= \int \Pr(b_H, \text{hs}, da), \\
&= (1-\mu) \int^{a_H^*} F(da) + \mu \left[(1-M) \int_{\hat{a}_H}^{a_H^*} F(da) + M \int_{\hat{a}_L}^{a_H^*} F(da) \right], \\
&= (1-\mu)(1 - C_H) + \mu [M E_L^{\text{hs}} + (1-M) E_H^{\text{hs}}].
\end{aligned} \tag{A.8}$$

From Equation (20),

$$\begin{aligned}
\Pr(b_L, \text{co}) &= \int \Pr(b_L, \text{co}, da), \\
&= \mu(1-M) \int_{a_L^*}^{\hat{a}_H} F(da) + \mu M \int_{a_L^*}^{\hat{a}_L} F(da), \\
&= \mu [M N_L^{\text{co}} + (1-M) N_H^{\text{co}}].
\end{aligned} \tag{A.9}$$

Finally, from Equation (21),

$$\begin{aligned}
\Pr(b_L, \text{hs}) &= \int \Pr(b_L, \text{hs}, da), \\
&= \mu \left[(1-M) \int^{\min\{a_L^*, \hat{a}_H\}} F(da) + M \int^{\min\{a_L^*, \hat{a}_L\}} F(da) \right] \\
&= \mu [M N_L^{\text{hs}} + (1-M) N_H^{\text{hs}}].
\end{aligned} \tag{A.10}$$

Note that M is indeed the stationary proportion of b_L adults implied by $\Pr(b, \mathrm{x}, da)$:

$$\begin{aligned}\sum_{\mathrm{x}} \Pr(b_L, \mathrm{x}) &= \mu\left[MN_L^{\mathrm{co}} + (1-M)N_H^{\mathrm{co}}\right] + \mu\left[MN_L^{\mathrm{hs}} + (1-M)N_H^{\mathrm{hs}}\right], \\ &= \mu M(1-P_L) + \mu(1-M)(1-P_H), \\ &= M,\end{aligned}$$

where the second row follow from $N_i^{\mathrm{co}} + N_i^{\mathrm{hs}} = 1 - P_i$, as can be seen from Equations (A.4) and (A.5).

Let $\Pr(b, \mathrm{x})$, the distribution of the endogenous state variables for a given generation of parents, be represented by the 4×1 vector $\mathbf{S}$:

$$\mathbf{S} \equiv \left[\Pr\left(b_H, \mathrm{co}\right), \Pr\left(b_H, \mathrm{hs}\right), \Pr\left(b_L, \mathrm{co}\right), \Pr\left(b_L, \mathrm{hs}\right)\right]^{\top},$$

and let $\Pr(b', \mathrm{x}'|b, \mathrm{x})$ be represented by the 4×4 transition matrix $\mathbf{T}$, mapping the distribution of states for adults of a given generation into the distribution of states for adults of the next generation:

$$\mathbf{T} \equiv \begin{bmatrix} \Pr\left(b'_H, \mathrm{co}'|b_H, \mathrm{co}\right) & \Pr\left(b'_H, \mathrm{co}'|b_H, \mathrm{hs}\right) & \Pr\left(b'_H, \mathrm{co}'|b_L, \mathrm{co}\right) & \Pr\left(b'_H, \mathrm{co}'|b_L, \mathrm{hs}\right) \\ \Pr\left(b'_H, \mathrm{hs}'|b_H, \mathrm{co}\right) & \Pr\left(b'_H, \mathrm{hs}'|b_H, \mathrm{hs}\right) & \Pr\left(b'_H, \mathrm{hs}'|b_L, \mathrm{co}\right) & \Pr\left(b'_H, \mathrm{hs}'|b_L, \mathrm{hs}\right) \\ \Pr\left(b'_L, \mathrm{co}'|b_H, \mathrm{co}\right) & \Pr\left(b'_L, \mathrm{co}'|b_H, \mathrm{hs}\right) & \Pr\left(b'_L, \mathrm{co}'|b_L, \mathrm{co}\right) & \Pr\left(b'_L, \mathrm{co}'|b_L, \mathrm{hs}\right) \\ \Pr\left(b'_L, \mathrm{hs}'|b_H, \mathrm{co}\right) & \Pr\left(b'_L, \mathrm{hs}'|b_H, \mathrm{hs}\right) & \Pr\left(b'_L, \mathrm{hs}'|b_L, \mathrm{co}\right) & \Pr\left(b'_L, \mathrm{hs}'|b_L, \mathrm{hs}\right) \end{bmatrix}.$$

Using Equations (A.2)-(A.5), the elements of $\mathbf{T}$ are

$$\Pr\left(b'_H, \mathrm{co}'|b_i, \mathrm{x}\right) = \left(1-\mu\right) C_H + \mu E_i^{\mathrm{co}}, \tag{A.11}$$

$$\Pr\left(b'_H, \mathrm{hs}'|b_i, \mathrm{x}\right) = \left(1-\mu\right)\left(1-C_H\right) + \mu E_i^{\mathrm{hs}}, \tag{A.12}$$

$$\Pr\left(b'_L, \mathrm{co}'|b_i, \mathrm{x}\right) = \mu N_i^{\mathrm{co}}, \tag{A.13}$$

$$\Pr\left(b'_L, \mathrm{hs}'|b_i, \mathrm{x}\right) = \mu N_i^{\mathrm{hs}}, \tag{A.14}$$

for each $i \in \{L, H\}$ and each $\mathrm{x} \in \{\mathrm{co}, \mathrm{hs}\}$. Thus, $\Pr\left(b'_H, \mathrm{co}'|b_i, \mathrm{x}\right)$ is the probability that b_i-adults with graduation level x have children that become b_H-adults and graduate college: Either the children are born with b_H (probability $1-\mu$) and they graduate college (probability C_H), or they are born with b_L (probability μ) and they become educated and graduate college (probability E_i^{co}). Equations (A.12)-(A.14) have similar interpretations.

Lemma 9. The distribution of endogenous state variables is stationary

$$\mathbf{S} = \mathbf{TS}.$$

Proof Note, from (A.11)-(A.14), that $\Pr(b', \mathrm{x}'|b, \mathrm{x})$ is independent of x. It follows that

$$\sum_{b,\mathrm{x}} \Pr(b', \mathrm{x}'|b, \mathrm{x}) \Pr(b, \mathrm{x}) = \Pr(b', \mathrm{x}'|b_H, \mathrm{co})(1-M) + \Pr(b', \mathrm{x}'|b_L, \mathrm{co})M.$$

Then,

$$\begin{aligned}\sum_{b,\mathrm{x}} \Pr(b'_H, \mathrm{co}'|b, \mathrm{x}) \Pr(b, \mathrm{x}) &= \Pr(b'_H, \mathrm{co}'|b_H, \mathrm{co})(1-M) + \Pr(b'_H, \mathrm{co}'|b_L, \mathrm{co})M \\ &= \left[(1-\mu)C_H + \mu E_H^{\mathrm{co}}\right](1-M) + \left[(1-\mu)C_H + \mu E_L^{\mathrm{co}}\right]M \\ &= (1-\mu)C_H + \mu\left[ME_L^{\mathrm{co}} + (1-M)E_H^{\mathrm{co}}\right], \\ &= \Pr(b_H, \mathrm{co}),\end{aligned}$$

where the last equality follows from Equation (A.7).

$$\begin{aligned}\sum_{b,\mathrm{x}} \Pr(b'_H, \mathrm{hs}'|b, \mathrm{x}) \Pr(b, \mathrm{x}) &= \Pr(b'_H, \mathrm{hs}'|b_H, \mathrm{co})(1-M) + \Pr(b'_H, \mathrm{hs}'|b_L, \mathrm{co})M \\ &= \left[(1-\mu)(1-C_H) + \mu E_H^{\mathrm{hs}}\right](1-M) + \left[(1-\mu)(1-C_H) + \mu E_L^{\mathrm{hs}}\right]M \\ &= (1-\mu)(1-C_H) + \mu\left[ME_L^{\mathrm{hs}} + (1-M)E_H^{\mathrm{hs}}\right], \\ &= \Pr(b_H, \mathrm{hs}),\end{aligned}$$

where the last equality follows from Equation (A.8).

$$\begin{aligned}\sum_{b,\mathrm{x}} \Pr(b'_L, \mathrm{co}'|b, \mathrm{x}) \Pr(b, \mathrm{x}) &= \Pr(b'_L, \mathrm{co}'|b_H, \mathrm{co})(1-M) + \Pr(b'_L, \mathrm{co}'|b_L, \mathrm{co})M \\ &= \mu\left[MN_L^{\mathrm{co}} + (1-M)N_H^{\mathrm{co}}\right], \\ &= \Pr\left(b_L, \mathrm{co}\right),\end{aligned}$$

where the last equality follows from Equation (A.9).

$$\begin{aligned}\sum_{b,\mathrm{x}} \Pr(b'_L, \mathrm{hs}'|b, \mathrm{x}) \Pr(b, \mathrm{x}) &= \Pr(b'_L, \mathrm{hs}'|b_H, \mathrm{co})(1-M) + \Pr(b'_L, \mathrm{hs}'|b_L, \mathrm{co})M \\ &= \mu\left[MN_L^{\mathrm{hs}} + (1-M)N_H^{\mathrm{hs}}\right], \\ &= \Pr\left(b_L, \mathrm{hs}\right).\end{aligned}$$

where the last equality follows from Equation (A.10). ■

A.5 Persistence

Define

$$\bar{E}^{\mathrm{x}} = ME_L^{\mathrm{x}} + (1-M)E_H^{\mathrm{x}} \quad \text{and} \quad \bar{N}^{\mathrm{x}} = MN_L^{\mathrm{x}} + (1-M)N_H^{\mathrm{x}}. \tag{A.15}$$

Lemma 10. Conditional graduation probabilities

The probability of college-graduate children given college-graduate parents is

$$\Pr(\text{co}'|\text{co}) = (1-\mu)\,C_H + \mu^2\frac{[N_L^{\text{co}}+E_L^{\text{co}}]\,\bar{N}^{\text{co}}}{\Pr(\text{co})} + \mu\,(1-\mu)\,\frac{[N_H^{\text{co}}+E_H^{\text{co}}]\,C_H}{\Pr(\text{co})} + \mu^2\frac{[N_H^{\text{co}}+E_H^{\text{co}}]\,\bar{E}^{\text{co}}}{\Pr(\text{co})},$$

and the probability of college-graduate children given high school-graduate parents is

$$\Pr(\text{co}'|\text{hs}) = (1-\mu)\,C_H + \mu^2\frac{[N_L^{\text{co}}+E_L^{\text{co}}]\,\bar{N}^{\text{hs}}}{\Pr(\text{hs})} + \mu\,(1-\mu)\,\frac{[N_H^{\text{co}}+E_H^{\text{co}}]\,(1-C_H)}{\Pr(\text{hs})} + \mu^2\frac{[N_H^{\text{co}}+E_H^{\text{co}}]\,\bar{E}^{\text{hs}}}{\Pr(\text{hs})}.$$

Proof From Equations (A.7)-(A.10), derive

$$\begin{aligned}
\Pr(\text{co}) = \Pr(b_H,\text{co}) + \Pr(b_L,\text{co}) &= (1-\mu)\,C_H + \mu\left(\bar{E}^{\text{co}}+\bar{N}^{\text{co}}\right),\\
\Pr(\text{hs}) = \Pr(b_H,\text{hs}) + \Pr(b_L,\text{hs}) &= (1-\mu)\,(1-C_H) + \mu\left(\bar{E}^{\text{hs}}+\bar{N}^{\text{hs}}\right).
\end{aligned}$$

From Equations (A.11)-(A.14), derive

$$\begin{aligned}
\Pr\left(\text{co}'|b_L,\text{co}\right) &= \Pr\left(b_L',\text{co}'|b_L,\text{co}\right) + \Pr\left(b_H',\text{co}'|b_L,\text{co}\right),\\
&= \mu N_L^{\text{co}} + (1-\mu)\,C_H + \mu E_L^{\text{co}},\\
\Pr\left(\text{co}'|b_H,\text{co}\right) &= \Pr\left(b_L',\text{co}'|b_H,\text{co}\right) + \Pr\left(b_H',\text{co}'|b_H,\text{co}\right),\\
&= \mu N_H^{\text{co}} + (1-\mu)\,C_H + \mu E_H^{\text{co}},
\end{aligned}$$

and

$$\begin{aligned}
\Pr\left(\text{co}'|b_L,\text{hs}\right) &= \Pr\left(b_L',\text{co}'|b_L,\text{hs}\right) + \Pr\left(b_H',\text{co}'|b_L,\text{hs}\right),\\
&= \mu N_L^{\text{co}} + (1-\mu)\,C_H + \mu E_L^{\text{co}},\\
\Pr\left(\text{co}'|b_H,\text{hs}\right) &= \Pr\left(b_L',\text{co}'|b_H,\text{hs}\right) + \Pr\left(b_H',\text{co}'|b_H,\text{hs}\right),\\
&= \mu N_H^{\text{co}} + (1-\mu)\,C_H + \mu E_H^{\text{co}}.
\end{aligned}$$

Combining,

$$\begin{aligned}
\Pr\left(\text{co}'|\text{co}\right) &= \frac{\Pr(\text{co}'|b_L,\text{co})\Pr(b_L,\text{co})}{\Pr(\text{co})} + \frac{\Pr(\text{co}'|b_H,\text{co})\Pr(b_H,\text{co})}{\Pr(\text{co})},\\
&= \frac{[\mu N_L^{\text{co}}+(1-\mu)\,C_H+\mu E_L^{\text{co}}]\,\mu\bar{N}^{\text{co}}}{\Pr(\text{co})} + \frac{[\mu N_H^{\text{co}}+(1-\mu)\,C_H+\mu E_H^{\text{co}}]\,(1-\mu)\,C_H}{\Pr(\text{co})}\\
&+ \frac{[\mu N_H^{\text{co}}+(1-\mu)\,C_H+\mu E_H^{\text{co}}]\,\mu\bar{E}^{\text{co}}}{\Pr(\text{co})},\\
&= (1-\mu)\,C_H + \mu^2\frac{[N_L^{\text{co}}+E_L^{\text{co}}]\,\bar{N}^{\text{co}}}{\Pr(\text{co})} + \mu\,(1-\mu)\,\frac{[N_H^{\text{co}}+E_H^{\text{co}}]\,C_H}{\Pr(\text{co})}\\
&+ \mu^2\frac{[N_H^{\text{co}}+E_H^{\text{co}}]\,\bar{E}^{\text{co}}}{\Pr(\text{co})},
\end{aligned}$$

and

$$\begin{aligned}
\Pr\left(\text{co}'|\text{hs}\right) &= \frac{\Pr\left(\text{co}'|b_L,\text{hs}\right)\Pr\left(b_L,\text{hs}\right)}{\Pr\left(\text{hs}\right)} + \frac{\Pr\left(\text{co}'|b_H,\text{hs}\right)\Pr\left(b_H,\text{hs}\right)}{\Pr\left(\text{hs}\right)}, \\
&= \frac{\left[\mu N_L^{\text{co}} + (1-\mu)\,C_H + \mu E_L^{\text{co}}\right]\mu\bar{N}^{\text{hs}}}{\Pr\left(\text{hs}\right)} + \frac{\left[\mu N_H^{\text{co}} + (1-\mu)\,C_H + \mu E_H^{\text{co}}\right](1-\mu)\,(1-C_H)}{\Pr\left(\text{hs}\right)} \\
&+ \frac{\left[\mu N_H^{\text{co}} + (1-\mu)\,C_H + \mu E_H^{\text{co}}\right]\mu\bar{E}^{\text{hs}}}{\Pr\left(\text{hs}\right)}, \\
&= (1-\mu)\,C_H + \mu^2\frac{\left[N_L^{\text{co}} + E_L^{\text{co}}\right]\bar{N}^{\text{hs}}}{\Pr\left(\text{hs}\right)} + \mu\,(1-\mu)\,\frac{\left[N_H^{\text{co}} + E_H^{\text{co}}\right](1-C_H)}{\Pr\left(\text{hs}\right)} \\
&+ \mu^2\frac{\left[N_H^{\text{co}} + E_H^{\text{co}}\right]\bar{E}^{\text{hs}}}{\Pr\left(\text{hs}\right)}.
\end{aligned}$$

■

Lemma 11. Intergenerational persistence of educational attainment

$$\Pr(\text{co}'|\text{co}) \geq \Pr(\text{co}'|\text{hs})$$

Proof From Lemma 10

$$\Pr(\text{co}'|\text{co}) - \Pr(\text{co}'|\text{hs}) = \mu^2\underbrace{\left[(N_L^{\text{co}} - N_H^{\text{co}}) + (E_L^{\text{co}} - E_H^{\text{co}})\right]}_{A}\underbrace{\left(\frac{\bar{N}^{\text{co}}}{\Pr\left(\text{co}\right)} - \frac{\bar{N}^{\text{hs}}}{\Pr\left(\text{hs}\right)}\right)}_{B}.$$

Our goal is to show that $A \times B \geq 0$. There are 6 arrangements of the P_is and C_is respecting $P_H > P_L$ and $C_H > C_L$. These arrangements are indicated in Table 3. It is immediate that $A = F(\mathbb{B}_L) - F(\mathbb{B}_H) \leq 0$. To sign B, note that the numerator of B is

$$\text{numerator of } B = \bar{N}^{\text{co}}\Pr(\text{hs}) - \bar{N}^{\text{hs}}\Pr(\text{co}) = \bar{N}^{\text{co}} - M\Pr(\text{co})/\mu.$$

From Table 3, the highest possible value for $\bar{N}^{\text{co}}$ is in case 1:

$$\begin{aligned}
\bar{N}^{\text{co}}\Big|_{\text{case 1}} &= C_L - MP_L - (1-M)P_H \\
&= C_L - (1 - M(1-P_L) - (1-M)(1-P_H)) \\
&= C_L - (1 - M/\mu)
\end{aligned}$$

where the last row follows from Equation (A.6). Thus,

$$\bar{N}^{\text{co}} \leq C_L - (1 - M/\mu).$$

Table 3 also indicates that $C_H \geq \bar{E}^{\text{co}} + \bar{N}^{\text{co}} \geq C_L$. Recall that $\Pr(\text{co}) = (1-\mu)C_H + \mu(\bar{E}^{\text{co}} + \bar{N}^{\text{co}})$ from Equations (A.7) and (A.9). Thus, $\Pr(\text{co}) \geq C_L$ and hence

$$-M\Pr(\text{co})/\mu \leq -MC_L/\mu.$$

Adding these inequalities yields

$$\text{numerator of } B = \bar{N}^{\text{co}} - M\,\Pr(\text{co})/\mu \leq (C_L - 1)(1 - M/\mu) < 0$$

where the last inequality follows from $C_L < 1$ and $M/\mu < 1$. Hence, the numerator of B is negative and, therefore, $A \times B \geq 0$. ■

Table 3: Arrangements of P_is and C_is

	C_is and P_is ranked by decreasing order					
	C_H, C_L, P_H, P_L	C_H, P_H, C_L, P_L	P_H, C_H, P_L, C_L	P_H, P_L, C_H, C_L	C_H, P_H, P_L, C_L	P_H, C_H, C_L, P_L
E_L^{co}	P_L	P_L	P_L	C_H	P_L	P_L
E_H^{co}	P_H	P_H	C_H	C_H	P_H	C_H
$\bar{E}^{\text{co}}$	$MP_L + (1-M)P_H$	$MP_L + (1-M)P_H$	$MP_L + (1-M)C_H$	C_H	$MP_L + (1-M)P_H$	$MP_L + (1-M)C_H$
N_L^{co}	$C_L - P_L$	$C_L - P_L$	0	0	0	$C_L - P_L$
N_H^{co}	$C_L - P_H$	0	0	0	0	0
$\bar{N}^{\text{co}}$	$C_L - MP_L - (1-M)P_H$	$M(C_L - P_L)$	0	0	0	$M(C_L - P_L)$
A	0	$C_L - P_H$	$P_L - C_H$	0	$P_L - P_H$	$C_L - C_H$
$\bar{E}^{\text{co}} + \bar{N}^{\text{co}}$	C_L	$MC_L + (1-M)P_H$	$MP_L + (1-M)C_H$	C_H	$MP_L + (1-M)P_H$	$MC_L + (1-M)C_H$
$F(\mathbb{B}_L)$	0	0	$P_L - C_L$	$C_H - C_L$	$P_L - C_L$	0
$F(\mathbb{B}_H)$	0	$P_H - C_L$	$C_H - C_L$	$C_H - C_L$	$P_H - C_L$	$C_H - C_L$

Note: Recall $E_i^{\text{co}} = \min(P_i, C_H)$, $N_i^{\text{co}} = \max(0, C_L - P_i)$, and $F(\mathbb{B}_i) = \max(0, \min(P_i, C_H) - C_L)$.