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	2025-003D
Revision Date	October 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

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

D. Fuller† G. Vandenbroucke‡

October 2025

Abstract

We present a theory of intergenerational persistence in college attainment (a Bachelor's degree, BD) guided by two well-documented empirical regularities: (i) patient individuals are more likely to be college graduates and (ii) patience persists across generations because parents choose to transmit it to their children, among other non-cognitive traits. To the best of our knowledge, we are the first to formalize and evaluate these mechanisms in a general equilibrium model. In the model, (i) and (ii) are endogenous outcomes implying college persistence. "Standard mechanisms", i.e., credit constraints, parental investments, or exogenous persistence in cognitive ability are not necessary. Theoretically, we characterize the conditions for college persistence and show that patience persistence is necessary for college persistence. Empirically, the baseline model without ability persistence, calibrated to U.S. data, matches college persistence and cross sectional inequality moments while accounting for 30% of earnings persistence. An augmented model, with exogenous ability persistence, reveals that ability persistence accounts for 80% of earnings persistence and 10% of college persistence. The remainder of college persistence is accounted for by endogenous persistence in patience. Ability persistence in the augmented model is a reduced form for the combined effects of the standard mechanisms. So, our finding that the standard mechanisms matter for earnings persistence is in line with the literature, while our finding that they matter little for college persistence, once patience is modeled, is new. 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. Overall college attainment and its persistence increase. Specifically, the policy raises the likelihood of a BD for individuals whose parents have a BD and it lowers that likelihood for individuals whose parents do not have a BD. This result differs from that of the existing literature where college subsidies alleviate credit constraints and reduce persistence.

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

*We thank Domenico Ferraro, George-Levi Gayle, Lance Lochner, Mariacristina De Nardi, B. Ravikumar, and participants at the Saint Louis Fed brown bag seminar, Midwest Macro Meetings, 4th Annual Labor Markets and Macroeconomic Outcomes Conference, BSE 2025, SED 2025, and SAET 2025 for useful comments. 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, individuals born in the 1980s have a 64% probability of attaining a bachelor's degree (BD, hereafter) if at least one parent holds a BD, compared to just 24% if neither parent does (Ernst et al., 2024). In short, standards of living persist across generations through education. This persistence is a source of inequality and, as such, the subject of a vast literature. We propose a theory where college persistence arises from a mechanism which has been documented but neither formalized nor evaluated in a general equilibrium model: the intergenerational persistence of time preferences.

The existing literature attributes persistence to a combination of factors, including exogenous cognitive ability persistence, parental transfers and investments, and credit constraints (e.g., Becker and Tomes, 1979, 1986, Keane and Wolpin, 2001, Restuccia and Urrutia, 2004, Lee and Seshadri, 2019, Cordoba et al., 2016).[1] Quantitatively, exogenous cognitive ability persistence can account for a large fraction of observed earnings persistence: Restuccia and Urrutia (2004) find 50%, Lee and Seshadri (2019) find 66%, and Cordoba et al. (2016) find more than 100%.[2] This literature is often agnostic as to the source of the exogenous persistence, whether innate or resulting from prior parental investment. However, evidence of persistence between parents and their adopted children implies that there exist forces beyond exogenous persistence in innate traits (e.g., Sacerdote, 2000, 2007, 2011, Björklund et al., 2006), and, hence, further justifies the literature's emphasis on parental investments in the cognitive ability of their children, and the role credit constraints may play in distorting such investments.

There is also evidence that non-cognitive traits matter for educational attainment (e.g., Heckman et al., 2006, Almlund et al., 2011, Black and Devereux, 2011). In this regard, we note two important empirical findings supported by a large literature: First, patient individuals tend to achieve higher educational attainment. Second, patience persists across generations because parents transmit it, among other non-cognitive traits, to their children. Taken together, these observations indicate the existence of a channel for persistence that, to the best of our knowledge, we are the first to formalize and evaluate in the context of a standard neo-classical general equilibrium model.

In our model, the above empirical findings are equilibrium outcomes implying persistence. Time preferences are heterogeneous, i.e., there are patient and impatient individuals, and endogenous, i.e., parents can exert an effort to teach patience to their children. In addition to time preferences, individuals differ also by their cognitive abilities. We assume no persistence of cognitive ability in our baseline specification. We also assume perfect credit markets and do not allow parental transfers or investments in children's cognitive abilities. These assumptions serve to isolate the

[1] One notable exception is Gayle et al. (2022) where the differential returns between full-time and part-time jobs induces high income households to have fewer children while investing more per child.

[2] Cordoba et al. (2016) endogenous fertility framework introduces a force that mitigate persistence.

mechanism of our baseline specification from the existing literature.

Two assumptions, common in the literature, yield critical implications in the context of our model. First, we assume a utility cost of college (e.g., Heckman et al., 1998, 2003) to generate the positive association between patience and college attainment reported in the empirical literature. Second, we assume that parents value the future welfare of their children to generate the intergenerational persistence in patience reported in the empirical literature. On the first assumption, we note that the utility cost of college is critical for the college-patience association only because we assume perfect credit markets. Without a utility cost, credit constraints would also imply a positive association since impatient individuals would be more likely to be constrained out of attending college. It is precisely because of this interaction between credit constraints and patience that abstracting from the former is necessary to show the novel role of the latter. Hence our assumptions of perfect credit markets and a utility cost of college. The second assumption implies patience persistence because patient parents value their children's *future* welfare more than impatient parents. Thus, patient parents are more likely to exert the effort of teaching patience to their children. We do not assume that patient parents are more altruistic towards their children than others.

We make two additional remarks about our theory. First, the literature on credit constraints makes a critical distinction between credit constraints at early ages and at college-going ages (e.g., Keane and Wolpin, 2001, Lochner and Monge-Naranjo, 2012, Caucutt and Lochner, 2020). Relaxing the latter is found to have little impact on college attainment. Relaxing the former, however, is found to lead to increased parental investments in children prior to the college decision. We interpret these results as indications that modeling credit constraints at the college-going age is not of first-order importance for college attainment, and that modeling constraints at early ages is necessary for a theory of ability and its persistence across generations. We do not propose a theory of ability at the college-going age. Instead, we take ability to be exogenous and acknowledge, in Section 4.2, that when we consider persistence in ability, it is a reduced form for the effects of parental investment and early credit constraints.

Second, patience is not the only non-cognitive traits that may matter for education. Patience, however, presents a unique characteristic since it matters for both inter- and intra-generational decisions, a fact already pointed out by Becker and Barro (1988) who wrote "*Life cycle utilities (...) are discounted only by time preference, while generational utilities in the dynastic utility function are discounted* **also** *by the degree of altruism toward descendants*" (emphasis added). To see the implication of this dual role of patience, contemplate other factors that may affect college attendance. Parents could, for instance, teach their children a taste for effort which could manifest itself in a lower utility cost of college. In such setting, individuals with low utility costs would be the most likely to attend college. 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? The same argument can be applied to other non-cognitive skills and/or aspect of preferences. Maybe parent can foster curiosity in their children; but again, why would college-educated parent do it more than others? Both cases require additional assumptions on the transmission of the traits from parents to children. In contrast, patience accomplishes this without additional assumptions and is empirically connected to college attainment.

We embed our theory in a general equilibrium model with overlapping generations. Individuals are endowed at birth with an ability to be productive (a cognitive trait) and a discount factor (a non-cognitive trait). These traits are drawn from distributions that are independent of one another and independent of the parents' traits. Parents can change the discount factor of their children at a cost, before the children become adults, at which time the children's birth-ability and their (possibly changed) discount factor become their adult traits. Adults make the college decision before they, themselves, are endowed with a child. College increases lifetime wages, but it requires time (foregone wages) and entails a good cost (tuition) in addition to the utility cost.

At the individual level, the 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 also 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, the 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 find and analyze the stationary distribution associated with this process. We show that persistence may be zero because there are "regions of inaction:" When children have such low ability that college is suboptimal, even with patience; and when children have such high ability that college is optimal, even with impatience. In both cases, parents cannot affect their child's college decision and, hence, persistence cannot hold. For persistence to be positive, (i) parents must instill patience in their children, *and* (ii) such choices by the parent must change the child'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.

Another theoretical result describes the conditions under which patience is persistent, and shows that patience persistence is necessary for college persistence. This is an important result since it reveals the fundamental mechanism of the model and because, as we indicated earlier, the literature has already documented direct evidence of its existence. So, through the lenses of the model, the empirical findings of the literature on patience imply persistence in college attainment indeed.

The baseline model is calibrated to U.S. data from the mid-2010s. By that time, cohorts of individuals born in the mid-1980s would have completed college if they had chosen to attend. The calibration targets, in particular, moments related to the level of college attainment and its intergenerational persistence. Additional targets include moments capturing the dispersion of earnings both between and within educational groups, specifically the college premium and the coefficient of variation in earnings conditional on college attainment. The model replicates the persistence of college attainment across generations while being consistent with cross-sectional income inequality.

In the calibration, the baseline model implies 30% of the persistence in earnings—a non-targeted moment—measured in the U.S. data (5.3% v. 18%). To further explore this result, we follow the existing literature and augment the baseline model to allow for exogenous persistence in cognitive ability. The augmented model is then calibrated to earnings persistence in addition to the original targets. The model implies that ability persistence amounts to 80% of observed earnings persistence to account for the data. This "large" contribution is consistent with findings in the existing literature (e.g., Restuccia and Urrutia, 2004, Cordoba et al., 2016, Lee and Seshadri, 2019). To decompose the relative contribution of ability persistence for college and earnings persistence, we simulate the augmented model with ability persistence shut down. College persistence declined by 10%. We conclude from these experiments that ability persistence accounts for 10% of college persistence whereas it accounts for 80% of earnings persistence. We interpret the persistence in cognitive ability in the augmented model as a reduced form for the combined effects of innate persistence, parental investments, and credit constraints. The contributions of our quantitative work is to show that (i) these standard mechanisms contribute little to college persistence once patience heterogeneity is modeled, while they account for most of earnings persistence; and (ii) the persistence in patience accounts for most of the college persistence.

We also use the baseline model to simulate a policy experiment in which a college subsidy is financed via a proportional labor tax. The policy raises both college attainment and its persistence. The latter effect obtains because the policy distorts intertemporal tradeoffs via a general equilibrium effect: The subsidy (applied in the present period) changes prices that apply in the present and future periods in the labor market. The changes in future prices (relative to present) imply that the policy affects people differently depending on their patience. Specifically, the policy discourages impatient individuals to seek a BD. Since impatient individuals are likely to have impatient parents that do not have a BD, persistence is strengthened. Furthermore, the policy encourages more patient individuals to seek a BD. Since these individuals are likely to have patient parents that do have a BD, this second mechanism also strengthens persistence. Interestingly, our results differ from those in the literature where it is often found that education subsidies reduce intergenerational persistence by alleviating credit constraints.

2 Literature and direct evidence

The fundamental mechanism of our model is the endogenous persistence of patience which implies the persistence of college attainment. The literature has documented direct evidence of this mechanism via two sets of empirical findings: First, the finding that patient individuals tend to achieve higher educational attainment. Second, the finding that patience persists across generations because parents transmit it to their children. Below, we describe the relevant literature after a brief review of the literature on the measurement of patience.

The measurement of time preferences

The measurement of time preferences is the concern of a large and active literature. The interested reader may consult surveys by Frederick et al. (2002) and Cohen et al. (2020). Time preferences are often measured via surveys, laboratory, and/or field experiments. Falk et al. (2018), is a recent example, presenting the Global Preference Survey (GPS), which is a systematic attempt at eliciting time preferences (among other aspect of preferences) from 80,000 people across 76 countries. The GPS, as most surveys, elicits time preferences via questions of the form "*would you rather receive x today or y in n months?*" Depending on the survey, the values of x, y, and n differ. There are sometime more than one question of this form in a given survey to obtain a precise measure of patience. In the GPS, for instance, a sequence of 5 questions yields 32 degrees of patience. Other recent examples, aside from the GPS, rely on similar survey questions. Golsteyn et al. (2014) use Swedish data from surveys on children at age 13. De Paola and Gioia (2017) use a survey on students enrolling at the University of Calabria. Alan and Ertac (2018) assess time preferences of elementary school children in Istanbul.

Some authors rely on surveys where participants are incentivized to be truthful via monetary rewards, e.g., Harrison et al. (2002), Meier and Sprenger (2010), Tanaka et al. (2010), Falk et al. (2016), Andreoni and Sprenger (2012), Sutter et al. (2013), and Angerer et al. (2023). In these studies, participants are asked questions of the form described above, knowing that the monetary payoff will be effectively delivered according to their preferred timing.

Finally, some authors adopt a revealed-preference approach. DellaVigna and Paserman (2005) collect PSID and NLSY data on whether individuals smoke and/or drink, have a bank account, use contraception, have life-insurance, etc. They build a measure of patience from such information. Hanushek et al. (2023) train a machine-learning model to characterize the relationship between Facebook interests, revealed by Facebook customers via their activities on their accounts, and patience from the GPS.

The association between time preferences and educational outcome

Using data from public school districts in Georgia, Castillo et al. (2011) find that impatience is significantly and positively correlated with behaviors predictive of economic outcomes, e.g., disciplinary referrals. 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) provide evidence of an adverse relationship between discount rates and school performance, health, labor supply, and lifetime income in Swedish longitudinal data. Using NLSY data, Cadena and Keys (2015) find that impatient individuals are more likely to drop out of high school or college. De Paola and Gioia (2017) report a positive association of patience with academic performance for Italian undergraduate students. Furthermore, they document higher drop-out rates for the impatient students. Falk et al. (2018), Sunde et al. (2022), and Hanushek et al. (2022) find patience to be strongly associated with student achievement (measured via PISA scores) and educational attainment across the countries in the GPS sample. In particular, Falk et al. (2018) report a significantly positive association in more than 90% of the countries. Hanushek et al. (2023) find similar results across regions of Italy and the United States. They report that patience accounts for 2/3rd of achievement variation across Italian regions and for 1/3rd of achievement variation across U.S. states. Angerer et al. (2023) find patience to be significantly and positively associated with students choosing academic high school track instead of vocational school tracks.

Beyond educational attainment, DellaVigna and Paserman (2005) present evidence that time preferences matter for job-search behavior and outcomes. Impatience measures are negatively correlated with search effort and the unemployment exit as predicted by the hyperbolic-discounting model. Meier and Sprenger (2010) present evidence that impatient individuals are more likely to have credit card debts.

The persistence of time preferences

Albanese et al. (2016) use a representative survey of the Italian population, conducted every two years on a sample of 8,000 people. They report that values and traits, such as patience, are correlated with those received from parents and that they are often, in turn, transmitted to children. Their study emphasizes the "purposeful transmission" of traits and values by parents. Brenøe and Epper (2022) find a substantial transmission of patience from parents to children. Furthermore, their results suggest that parents' decisions of how and how often to interact with their children are determining factors in the transmission of time preferences. Brown and Van der Pol (2015) use Australian data on young adults (16-25) and their parents. They report a significant relationship between the time preferences of parents and that of their children. Chowdhury et al. (2022) conduct a large-scale experiment in rural Bangladesh and find significant persistence of time preferences, as well as risk preferences. Kosse and Pfeiffer (2012, 2013) find evidence of an associations between the time preferences of preschool children and their parents in the German Socioeconomic Panel, a large representative panel survey of households and their individuals. The former paper emphasizes the correlation between mother and children while the latter paper emphasizes short- versus long-run patience in a hyperbolic discounting setup. Gauly (2017) also uses the German Socioeconomic Panel and finds a significantly positive correlation between the patience of children and that of their mother and father. Furthermore, she indicates that the transmission of patience is influenced by parents and, hence, goes beyond genetic transmission. Coda Moscarola et al. (2024) use data on Italian households (parents with children 14-20). They find a positive correlation of patience between parents and children and, furthermore, find that this correlation is significantly strengthened when parents take actions to foster their children's patience.

Additional evidence

Our model's implications are consistent with other evidence indicating the role of parents in the development of their children's preferences and/or pointing out persistence of preferences, albeit not necessarily time preferences.

Dohmen et al. (2012) and Zumbuehl et al. (2021) document that attitudes such as risk preferences and trust are transmitted from parents to children. They indicate, in particular, the role of parents who affect the transmission process of non-cognitive traits by exerting effort. Grönqvist et al. (2017) study 38 cohorts of Swedish men and find that the intergenerational correlation in non-cognitive ability is close to that of cognitive ability. Furthermore, using a subset of adopted boys to avoid the confounding effect of genetics, they find that mothers are instrumental in passing down abilities. Li et al. (2024) find evidence of significant mother-child correlation of non-cognitive skills in Wuhan.

Doepke and Zilibotti (2017) provide a detailed review the evidence on the effect of parenting style on children's school achievement, either across countries, in the historical record, and in particular datasets such as the NLSY. Guryan et al. (2008) and Ramey and Ramey (2010) report evidence that parents with a college degree spend more time with their children than parents with a high school degree. Barnard (2004) report evidence that parents involvement, i.e., non-pecuniary investment, when children are in elementary school have positive effects on their children later schooling experience.

Endogenous preferences

Our paper also contributes to a larger literature where 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; in Fernández-Villaverde et al. (2014), parents' can shame children into avoiding premarital sex. Agostinelli et al. (2025) argue that parents are instrumental in the intergenerational transmission of cultural traits. In the spirit of this literature, ours is a model where parents influence their children's discount factor à la Becker and Mulligan (1997) or Bhatt and Ogaki (2012).

3 Model

3.1 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. Individuals live for L periods, with childhood lasting 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 is distributed i.i.d. across the population according to 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 $b_H \equiv (\beta_H,\delta_H)$. Newborns 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.[3]

We assume that all children complete high school and 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." When parents exert effort to change their children' discount factor, we say the children are "educated."

[3] We assume a utility cost of teaching patience for analytical convenience. If good resources (instead of χ) were involved in the teaching of patience the logic of our model would remain the same: We would conclude (see Eq. 14 below) that b_H-parents are more likely than b_L-parents to educate their children. If time was involved in the teaching of patience, however, high-ability b_H-parents may choose not to educate their children because their opportunity cost may be too high.

3.2 Production

The representative firm seeks to maximize profit:

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

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_K is the rate of depreciation of capital. The first-order optimality conditions are

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

We adopt the following functional form for the production function:

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

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 $1/(1-\rho)$. The parameters z^{co} and z^{hs} are skill-biased productivity terms.

3.3 Individuals

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

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

where

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

Equation (3) describes the value of graduating college (x = co) or not (x = hs). It comprises the utility $\mathcal{U}^{\mathrm{x}}(a,b)$ 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 $I^{x}(a)$ in Equation (4) denotes the net present value of labor income for high school and college graduates:

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

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

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). Positive utility implies that patient individuals are better off than impatient ones, given the same consumption sequence. Negative utility implies the opposite. This assumption is similar to that of models where endogenous survival probabilities act as discount factors (see Rosen, 1988).

We differentiate β and δ to clarify the distinction between discounting and altruism. Suppose that individuals were endowed with an altruism parameter ω such that adults' preferences were

$$V_{\text{parents}}=\sum_{j=1}^{J}\beta^{j-1}U(c_j)+\overbrace{\beta^{A-1}\omega}^{\delta}V_{\text{child}}.$$

It is immediate that δ conflates discounting and altruism. Our theory is not about heterogeneity in altruism, however. Thus, we assume that ω is the same for all types and normalize it to 1. It follows that the discount factors β and δ are positively assorted in the pairs $b_L \equiv (\beta_L,\delta_L)$ and $b_H \equiv (\beta_H,\delta_H)$. In short, we do not assume patient individuals are more altruistic.

3.3.1 *College decision*

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

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

There exists one $\bar{a}>0$ if and only if $w^{\mathrm{co}}D^{\mathrm{co}} > w^{\mathrm{hs}}D^{\mathrm{hs}}$. 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

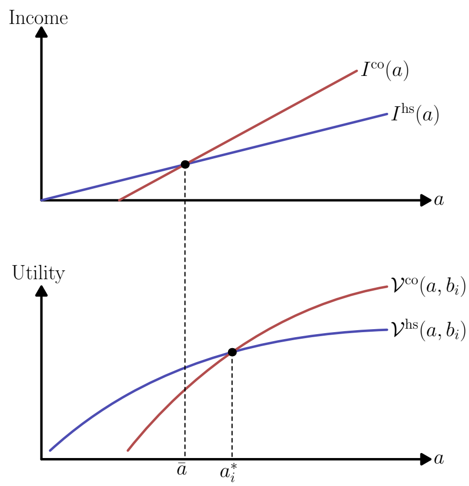


Figure 1: The college decision

$\bar{a} < 0$ 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.

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:

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

First, Lemma 4 (Appendix A) establishes that a_i^* exists, is unique, and that

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

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 becomes an income maximizing choice, since $I^{\mathrm{co}}(a) > I^{\mathrm{hs}}(a)$ 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, because the additional income from college does 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 because the expected utility of children is independent from a parent's college decision, as shown in 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 $\psi = 0$, individuals compare $I^{\text{co}}(a)$ and $I^{\text{hs}}(a)$ and their discount factors are irrelevant for college decisions. When $\psi > 0$, however, the NPV criterion does not apply. There are no markets from where to "borrow utils." Individuals must assess whether the future utils afforded by college are enough to offset the utility cost they bear at age 1. The higher their β, the more they do.

The utility cost of schooling is necessary to generate the observation that patient individuals are more likely to graduate college, *in the context of our model with prefect credit markets*. The utility cost of college is not necessary in general. If there was a credit constraint, impatient individuals, who seek to consume more in the early stages of their lives, would be more likely to be financially constrained and, hence, unable to afford the tuition cost e. Thus, in a model with a credit constraint, patient individuals would be more likely to graduate college. Just like they are in our model. It follows that the mechanism we describe here is not orthogonal to the mechanisms relying on imperfect credit markets in the intergenerational persistence literature. A corollary of these observations is that the emphasis on credit constraints in the existing literature could be viewed as an argument in favor of modeling heterogeneity in time preferences (beyond income and wealth) because time preferences matter for investment decisions whenever credit markets are imperfect.

3.3.2 *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). Three cases must be considered, depending on the college decision:

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, but 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

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

With this notation, parents educates their children when $\delta_i \Delta(a) \geq \chi$. Furthermore, Δ is continuous (Equation 10) and monotonically increasing (Lemma 8). Given this, 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

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

and

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

Equations (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 welfare being increasing in both ability and the discount factor. This implies that when parents educate their children they benefit from the higher future welfare of the children. The higher the ability of the children, the more likely it is the benefits to the parents 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, which implies that the ability threshold at which children are educated is lower when the parents are patient. Therefore, 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

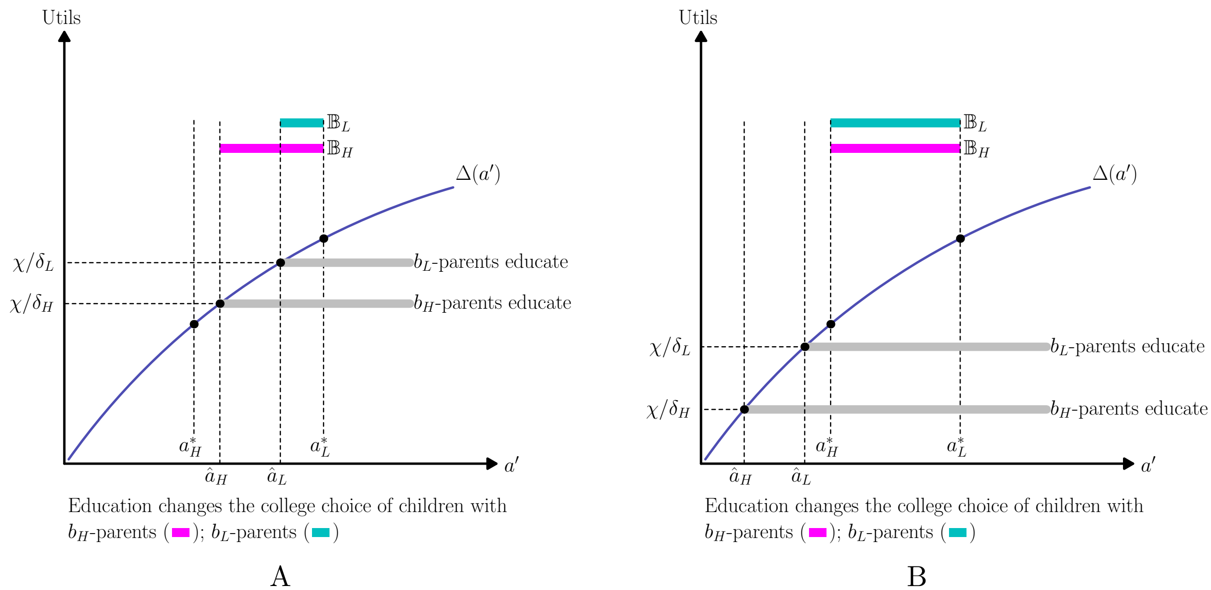


Figure 2: The education of (a', b'_L)-born children

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 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.

3.4 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^*.[4] 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

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

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

Note that (18)-(21) define $\Pr(b, \mathrm{x}, da)$ implicitly since $\Pr(b) = \sum_{\mathrm{x}} \int \Pr(d, \mathrm{x}, da)$ 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,

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

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.[5] 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

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

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.

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

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

3.5 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 is the difference between horizontal and upward mobility.[6] So,

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

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

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

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

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

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

and the upward mobility of college is

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

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

3.6 Discussion

Proposition 1 – College persistence

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

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.

Proposition 2 – Patience persistence is necessary for college persistence

$$\Pr(\mathrm{co}'|\mathrm{co}) > \Pr(\mathrm{co}'|\mathrm{hs}) \;\Rightarrow\; \Pr(b'_H|b_H) > \Pr(b'_H|b_L).$$

Proof – Equations (17) and (30) imply that $\Pr(\mathrm{co}'|\mathrm{co}) > \Pr(\mathrm{co}'|\mathrm{hs}) \Rightarrow P_H > P_L$. The result follows from Equation (27).

Thus, in data generated by our model, an observer of educational attainment (not observing patience) would note that college attainment is frequent in some families that cannot be distinguished, based on observable variables, from other families where college attainment is less frequent. The observer would then conclude that a family fixed effect matters to account for variations in college attainment. The fixed effect would, in fact, be the families' persistent discount factors. Thus, we interpret patience 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."*

A combination of two mechanisms yield 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 credit 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. This implies that patient parents have more incentives to educate their children to be patient. Critically, the mechanism is that parents are altruistic and *not* that patient parents are more altruistic than others.

In Section 2, we reviewed the literature documenting that patient individuals do indeed achieve higher levels of education, and that patience persists from parents to their children. Thus, there exist direct evidence of the mechanism that we emphasize in our model.

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 4. 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 they 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 4. 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.

Our analysis, to this point, has characterized patience and college persistence and established that the former is necessary for the latter. Next, we investigate the model's quantitative properties. We investigate whether the model can be in line 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.

4 Quantitative analysis

4.1 Calibration

We calibrate the model to U.S. data. Our estimates of horizontal and upward mobility are from Ernst et al. (2024). We focus our calibration on their last available cohort which was born in the mid 1980s. We only consider individuals with a high school degree. We label those who completed a bachelor's degree "college graduates," and the others "high school graduates."

We set a period to last one year, 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% ($\delta_K = 0.047$) 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 their own:

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

There is evidence that patience increases with age, and particularly during childhood (e.g., Harrison et al., 2002, Andreoni and Sprenger, 2012). We take this evidence to the limit for the sake of simplicity and assume that children are born impatient ($\mu = 1$). Becker and Mulligan (1997) make a similar assumption in their model of endogenous time preferences. Finally, the ability distribution is log-normal with mean 1:

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

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.[7] 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 Ernst et al. (2024, Table 19).

3. Pr(co′|hs): The upward mobility of college attainment for the mid-1980s cohort.
   We use 24% from Ernst et al. (2024, Table 19).

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.

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.

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

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 cross-sectional inequality-related moments such as the two coefficients of variation and the college premium. The variance of ability is also informed by intergenerational college persistence. Suppose there was no heterogeneity in ability. Then, the model would trivially imply 100% college persistence because all adults would make the same education and college decisions. The adult population would be composed either of only college graduates or only high school graduates. It follows that σ_a^2 is informed, simultaneously by cross-sectional earnings heterogeneity and intergenerational persistence.

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 capital-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 utility cost of education, χ, and the difference between β_L and β_H 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 horizontal and upward mobility in college attainment inform the discount factors and the utility cost of education.

4.2 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 generates mildly more persistence than in the data: 44.5% (i.e., $67.8 - 23.3$) v. 40% (i.e., $64 - 24$). 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 4 (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 would have 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 $\mathbb{B}_H = [a^*_H, a^*_L]$.

The literature on intergenerational persistence is often concerned with earnings persistence. We assess our model's implication for earnings persistence by computing the rank-rank correlation between the earnings of parents and children in model-simulated data.[8] We find persistence to be 5.3%. Ernst et al. (2024, Table 10), our data source for college persistence, report a persistence of 18% for the same mid-1980s cohort. Thus, our Baseline generates about 30% of the observed earnings persistence in the U.S. data.

[8] The traditional approach to measuring earnings persistence (e.g., Solon, 1992, 1999) consists in estimating the slope of a regression of (log) children's earnings on (log) parents' earnings. This slope, however, confounds intergenerational correlation and changes in the variability of earnings from one generation to the next (Chetty et al., 2014). The rank-rank correlation avoids this drawback.

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) Ernst et al. (2024, Table 19); (c) CPS; (d) Digest of Education Statistics.

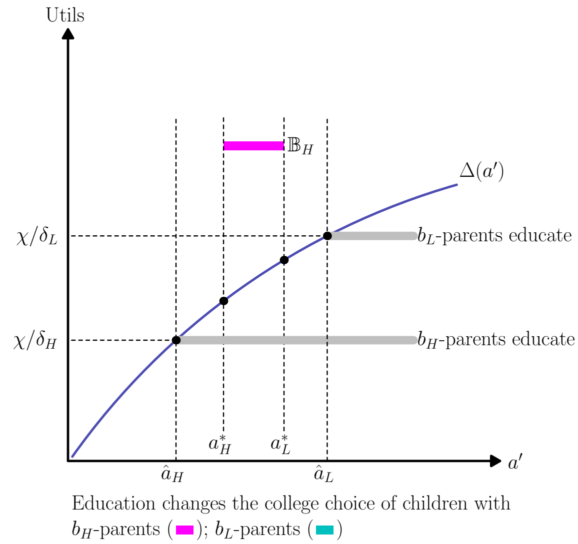


Figure 3: Ranking of ability thresholds in the calibrated model

Next, we calibrate a version of our model allowing for persistence in cognitive ability:

$$\begin{aligned} \ln a_{\text{child}} &= (\pi - 1)\frac{\sigma_a^2}{2} + \pi \ln a_{\text{parent}} + \epsilon\sqrt{\sigma_a^2(1-\pi^2)}, \\ \epsilon &\sim N(0,1). \end{aligned}$$

This process implies that the stationary distribution of ability is $\ln a \sim N(-\sigma_a^2/2, \sigma_a^2)$, i.e., ability has a mean of one and a variance dictated by σ_a^2. We add a target of 18% for earnings persistence to the targets in our Baseline and refer to this augmented model as Model B. Our results are displayed in Appendix B. We find π to be 14.5% and conclude that ability persistence accounts for 80% (i.e., 14.5/18) of earnings persistence; the remainder is due to the endogenous transmission of patience and the implied persistence in college. In the literature emphasizing parental transfers, investments, and borrowing limits, the contribution of ability persistence to earnings persistence is of first-order importance as well: Restuccia and Urrutia (2004) find that ability persistence accounts for half of earnings persistence, and Lee and Seshadri (2019) find that it accounts for two-third. Cordoba et al. (2016) require ability persistence to exceed that of earnings because their model features endogenous fertility and investment in children, a combination that dampens earnings persistence.

Having established the role of ability persistence for earnings persistence, we turn to the role of ability persistence for college persistence. In Model B, the calibrated preference parameters differ from their Baseline values because they determine the part of college persistence not due to ability persistence, while the latter is disciplined by earnings persistence. To assess how large that part is, we compute Model C: An equilibrium using the parameters of Model B, but with $\pi = 0$, such that

Table 3: College persistence and its components across models

	Baseline	Model B	Model C
Horizontal mobility, %	67.8	70.0	67.1
Upward mobility, %	23.3	22.6	24.6
Persistence, %	44.5	47.4	42.5

Source: Authors' calculations.

a is iid across generations and drawn from the stationary distribution of Model B.

Table 3 reports persistence figures for the three models. It is worth noting that the stationary distribution of ability is the same in all three models by construction. Thus, they all feature similar cross-sectional inequality. Unsurprisingly, removing ability persistence implies a decline of college persistence from Model B (47.4%) to Model C (42.5%) due to a decline in horizontal mobility and an increase in upward mobility. Persistence in Model C is also less than in the Baseline: On the one hand the Baseline ascribes all college persistence to persistence in patience by design; on the other hand, Model C ascribes only the fraction of college persistence not due to ability persistence to persistence in patience.

The main lesson from Table 3 is the magnitude of the decline of college persistence from Model B to C: about 10%. In other words, through the lenses of our model, ability persistence accounts for 10% of college persistence whereas it accounts for 80% of earnings persistence. These figures do not add up to 100% since there is a positive interaction between the persistence of patience and that of ability: Under ability persistence, high ability parents are more likely to have high ability children whom, they are more likely to educate (see Equation 14) to become patient. These results indicate the quantitative relevance of our mechanism.

In model B, we are agnostic about the source of ability persistence. Even though we represent ability as an exogenous process, we do not interpret it as innate since age-1 adults are of college-going age. Ability persistence is better interpreted as a reduced-form representation of the combined effects of "standard mechanisms:" innate persistence, parental investments in early (i.e., pre-college) childhood, and the credit constraints that may impair such investments. Hence, our findings regarding the role of ability persistence for earnings persistence align with the literature emphasizing the standard mechanisms. Therefore, the contributions of our quantitative work are that (i) the standard mechanisms contribute little to college persistence once patience heterogeneity is modeled while, in line with the literature, they account for most of the earnings persistence; and (ii) patience persistence accounts for most of college persistence.

Modeling the standard mechanisms explicitly, instead of via a reduced-form, would reveal interactions with patience worth contemplating but out of the scope of this paper for the sake of simplicity and clarity. To see what these interactions would be, it is worth noting the distinction between credit constraints at early ages and at college-going ages (e.g., Lochner and Monge-Naranjo, 2012, Caucutt and Lochner, 2020). Consider early constraints first. That is, imagine a version of our model where parents can invest in their child's cognitive ability (not just their patience) prior to the child becoming an age-1 adult. These investments could be hiring tutors, sending children to quality preschool programs, enrolling them in quality private schools prior to college, moving to expensive neighborhoods with quality pubic schools, sending children abroad to learn foreign languages, etc. Imagine also that parents face a credit constraint. Conditional on ability, impatient parents are more likely to be constrained and thus unable to achieve the optimal investment in their child's cognitive ability. This mechanism differs from that emphasized in the literature where the low-income (i.e., low ability) parents are more likely to be constrained out of investing optimally in their child's cognitive ability. The interaction arises because low-income parents are simultaneously low-ability and impatient parents. Regarding late constraints, imagine a version of our model where young adults face a credit constraint at the college-decision stage. Conditional on ability, impatient young adults are more likely to be constrained out of college. Again, this mechanism differs from one where low-income families are constrained out of sending their children to college and the interaction arises because low-income families are also impatient families.

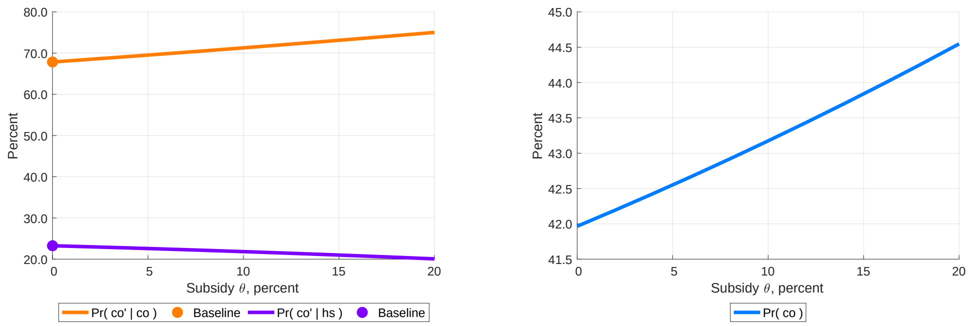
Because of the interactions we just described, the motivation for incorporating the standard mechanisms explicitly into our model cannot be to perform a decomposition of the form "the role of patience versus that of credit constraint." In the models we sketched above, removing the credit constraints would not yield an orthogonal decomposition since it would remove a channel via which patience operates. The motivation for explicitly modeling the standard mechanism could be, however, to evaluate policies alleviating credit constraints in order to promote college attainment and reduce its intergenerational persistence. Below, we propose a step in this direction. We return to our Baseline model without persistence in cognitive ability and evaluate the effects of a college subsidy financed via a flat labor income tax.

4.3 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



Persistence

College attainment

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

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, persistence 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 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.

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 $\tilde{w}^{\text{x}} \equiv w^{\text{x}}(1-\tau)$ represent after-tax wages. As the subsidy increases, recall that college attainment increases as well. General equilibrium forces imply that (i) $\tilde{w}^{\text{co}}$ decreases because w^{co} decreases with the rise in college attainment and because τ increases; (ii) $\tilde{w}^{\text{hs}}$ 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 $\Gamma(a,b) \equiv \mathcal{V}^{\text{co}}(a,b) - \mathcal{V}^{\text{hs}}(a,b)$ such that a_i^* satisfies

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

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

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.

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 offsets the reduction in tuition, leading to a decline in the graduation rate of impatient individuals. 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. 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, 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 increase college attainment in equilibrium; however, they also increase college persistence. 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. Thus, inequality in college attainment increases.

The exercise above should not be interpreted as an assessment of actual policies since, as we indicated earlier, such assessment would require to explicitly incorporate credit constraint into the model. Instead, the value of the exercise is to point out a mechanism arising from patience heterogeneity, which may operate in a direction opposite to the goal of the policy.

5 Conclusion

We present a novel theory of the persistence in college attainment. Unlike in most of the existing literature, ours does not hinge on either exogenous cognitive ability persistence, parental investments, or credit constraints. Instead, we emphasize the role of non-cognitive traits, namely time preferences, and their transmission. Specifically, our theory is guided by two well-documented empirical regularities. The first is that patient individuals are more likely to be college graduates, The second is that patience persists across generations because parents choose to transmit it to their children, among other non-cognitive traits.

Ours is a standard neo-classical overlapping generations model with a college decision. We assume that there is a utility cost of college, that parents are altruistic toward their children, and that they have a technology to teach patience to their children at some cost. Under these assumptions, the aforementioned empirical regularities are endogenous outcomes implying college persistence.

Our main theoretical results describe the conditions under which patience is persistent, the conditions under which college is persistent, and establish that the former is necessary for the latter. This last result implies that the empirical literature (i.e., the empirical regularities we described) has already documented direct evidence of the fundamental mechanism of our model.

Quantitatively, our baseline model without cognitive ability persistence matches the observed persistence in college attainment and other moments pertaining to income inequality. It generates 30% of the observed earnings persistence, however. This finding is consistent with most of the literature (with the exception of the work of Gayle et al., 2022) where significant exogenous persistence in cognitive ability is needed to generate the observed earnings persistence.

An augmented version of the model, with exogenous persistence in cognitive ability, reveals that ability persistence accounts for 80% of earnings persistence and 10% of college persistence. The remainder of college persistence is accounted for by endogenous persistence in patience. We interpret ability persistence as a reduced-form representation of the combined effects of "standard mechanisms": innate persistence, parental investments in early childhood, and credit constraints. Thus, the importance of persistence in cognitive ability for earnings persistence is in line with the findings of the literature emphasizing the standard mechanisms. One of our contributions, therefore, is to indicate that, once patience heterogeneity is modeled, the standard mechanisms explain little of college persistence while patience persistence accounts for most of it.

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 credit constraints and, hence, reduce college persistence. The value of this exercise is to point out that, with patience heterogeneity, there exists a mechanism which may operate in a direction opposite to one of the stated goal of the college subsidy, i.e., that of improving social mobility.

There are interesting avenues for future research since our mechanism would interact in non-trivial ways with the standard mechanisms often considered in the literature: First, credit constraint would interact with heterogeneity in patience since impatient individuals would be more likely to be constrained than patient individuals. Second, patient parents would have higher incentives to invest in their children than impatient parents, and these incentives would be even higher if the children were patient themselves.

References

F. Agostinelli, M. Doepke, G. Sorrenti, and F. Zilibotti. A stairway to success: How parenting shapes culture and social stratification, 2025.

S. Alan and S. Ertac. Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy*, 126(5):1865–1911, 2018.

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.

M. Almlund, A. L. Duckworth, J. Heckman, and T. Kautz. Personality psychology and economics. In *Handbook of the economics of education*, volume 4, chapter 1, pages 1–181. Elsevier, 2011.

J. Andreoni and C. Sprenger. Estimating time preferences from convex budgets. *American Economic Review*, 102(7):3333–3356, 2012.

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.

W. M. Barnard. Parent involvement in elementary school and educational attainment. *Children and youth services review*, 26(1):39–62, 2004.

G. S. Becker and R. J. Barro. A reformulation of the economic theory of fertility. *Quarterly Journal of Economics*, 103(1):1–25, 1988.

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.

V. Bhatt and M. Ogaki. Tough love and intergenerational altruism. *International Economic Review*, 53(3):791–814, 2012.

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. Björklund, M. Lindahl, and E. Plug. The origins of intergenerational associations: Lessons from swedish adoption data. *The Quarterly Journal of Economics*, 121(3):999–1028, 2006.

S. E. Black and P. J. Devereux. Recent developments in intergenerational mobility. In *Handbook of labor economics*, volume 4, chapter 16, pages 1487–1541. Elsevier, 2011.

A. A. Brenøe and T. Epper. Parenting values and the intergenerational transmission of time preferences. *European Economic Review*, 148:104208, 2022.

H. Brown and M. Van der Pol. Intergenerational transfer of time and risk preferences. *Journal of Economic Psychology*, 49:187–204, 2015.

B. C. Cadena and B. J. Keys. Human capital and the lifetime costs of impatience. *American Economic Journal: Economic Policy*, 7(3):126–153, 2015.

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.

M. Castillo, P. J. Ferraro, J. L. Jordan, and R. Petrie. The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics*, 95(11-12):1377–1385, 2011.

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.

R. Chetty, N. Hendren, P. Kline, and E. Saez. Where is the land of opportunity? the geography of intergenerational mobility in the united states. *Quarterly Journal of Economics*, 129(4):1553–1623, 2014.

S. Chowdhury, M. Sutter, and K. F. Zimmermann. Economic preferences across generations and family clusters: A large-scale experiment in a developing country. *Journal of Political Economy*, 130(9):2361–2410, 2022.

F. Coda Moscarola, D. Del Boca, and G. Paladino. Intergenerational transmission of time preferences and saving attitude: the role of information sharing. *Review of Economics of the Household*, pages 1–43, 2024.

J. Cohen, K. M. Ericson, D. Laibson, and J. M. White. Measuring time preferences. *Journal of Economic Literature*, 58(2):299–347, 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.

J. C. Cordoba, X. Liu, and M. Ripoll. Fertility, social mobility and long run inequality. *Journal of Monetary Economics*, 77:103–124, 2016.

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.

S. DellaVigna and M. D. Paserman. Job search and impatience. *Journal of Labor Economics*, 23 (3):527–588, 2005.

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.

T. Dohmen, A. Falk, D. Huffman, and U. Sunde. The intergenerational transmission of risk and trust attitudes. *Review of Economic Studies*, 79(2):645–677, 2012.

E. Ernst, F. Langot, R. Merola, and J. Gonzales. Intergenerational trends in educational and income mobility in the united states of america since the 1960s. *ILO Working Paper*, (111), 2024.

A. Falk, A. Becker, T. Dohmen, D. Huffman, and U. Sunde. The preference survey module: A validated instrument for measuring risk, time, and social preferences. *IZA Discussion Paper*, (9674), 2016.

A. Falk, A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde. Global evidence on economic preferences. *Quarterly Journal of Economics*, 133(4):1645–1692, 2018.

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.

S. Frederick, G. Loewenstein, and T. O'donoghue. Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401, 2002.

B. Gauly. The intergenerational transmission of attitudes: Analyzing time preferences and reciprocity. *Journal of Family and Economic Issues*, 38:293–312, 2017.

G.-L. Gayle, L. Golan, and M. A. Soytas. What is the source of the intergenerational correlation in earnings? *Journal of Monetary Economics*, 129:24–45, 2022.

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.

E. Grönqvist, B. Öckert, and J. Vlachos. The intergenerational transmission of cognitive and noncognitive abilities. *Journal of Human Resources*, 52(4):887–918, 2017.

J. Guryan, E. Hurst, and M. Kearney. Parental education and parental time with children. *Journal of Economic Perspectives*, 22(3):23–46, Summer 2008.

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.

G. W. Harrison, M. I. Lau, and M. B. Williams. Estimating individual discount rates in denmark: A field experiment. *American Economic Review*, 92(5):1606–1617, 2002.

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.

J. J. Heckman, J. Stixrud, and S. Urzua. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3):411–482, 2006.

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.

F. Kosse and F. Pfeiffer. Impatience among preschool children and their mothers. *Economics Letters*, 115(3):493–495, 2012.

F. Kosse and F. Pfeiffer. Quasi-hyperbolic time preferences and their intergenerational transmission. *Applied Economics Letters*, 20(10):983–986, 2013.

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

Y. Li, S. Zhao, Z. Shi, and J. Nie. Intergenerational transmission of non-cognitive abilities: Evidence from the wuhan birth cohort studies. *China Economic Review*, 87:102214, 2024.

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

S. Meier and C. Sprenger. Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, 2(1):193–210, 2010.

G. Ramey and V. A. Ramey. The rug rat race. *Brookings Papers on Economic Activity*, pages 129–199, Spring 2010.

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.

G. Solon. Intergenerational income mobility in the united states. *American Economic Review*, pages 393–408, 1992.

G. Solon. Intergenerational mobility in the labor market. In *Handbook of labor economics*, volume 3, chapter 29, pages 1761–1800. Elsevier, 1999.

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.

T. Tanaka, C. F. Camerer, and Q. Nguyen. Risk and time preferences: Linking experimental and household survey data from vietnam. *American Economic Review*, 100(1):557–571, 2010.

M. Zumbuehl, T. Dohmen, and G. Pfann. Parental involvement and the intergenerational transmission of economic preferences, attitudes and personality traits. *The Economic Journal*, 131 (638):2642–2670, 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

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

Lemma 8. The function Δ

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

Proof From Equation (A.1)

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

The first and last rows are positive because $\Lambda(\beta_H) \ge \Lambda(\beta_L)$ (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 $\Lambda(\beta_H) \ge \Lambda(\beta_L)$ (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),

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

From Equation (19),

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

From Equation (20),

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

Finally, from Equation (21),

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

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

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

where the second row follow from $N_i^{\text{co}} + N_i^{\text{hs}} = 1 - P_i$, 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

$$\begin{aligned} \Pr\left(b'_H, \mathrm{co}'|b_i, \mathrm{x}\right) &= (1-\mu)\, C_H + \mu E_i^{\mathrm{co}}, &\text{(A.11)} \\ \Pr\left(b'_H, \mathrm{hs}'|b_i, \mathrm{x}\right) &= (1-\mu)\,(1 - C_H) + \mu E_i^{\mathrm{hs}}, &\text{(A.12)} \\ \Pr\left(b'_L, \mathrm{co}'|b_i, \mathrm{x}\right) &= \mu N_i^{\mathrm{co}}, &\text{(A.13)} \\ \Pr\left(b'_L, \mathrm{hs}'|b_i, \mathrm{x}\right) &= \mu N_i^{\mathrm{hs}}, &\text{(A.14)} \end{aligned}$$

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,

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

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

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

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

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

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

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

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

A.5 Persistence

Define

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

Lemma 10. Conditional graduation probabilities The probability of college-graduate children given college-graduate parents is

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

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

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

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,

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

and

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

■

Lemma 11. Intergenerational persistence of educational attainment

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

Proof From Lemma 10

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

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 4. 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 4, 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 4 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 4: 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 $F(\mathbb{B}_i) = \max(0, \min(P_i, C_H) - C_L)$.

B Calibration with persistent ability

Table 5: Calibrated parameters with persistent ability

Parameters	Value	Description
σ	0.951	CRRA preference parameter
β_L	0.950	low discount factor
β_H	0.978	high discount factor
ψ	1.349	utility cost of college
χ	90.217	utility cost of education
e	2.519	good cost of college
z^{co}	1.732	college-biased productivity
z^{hs}	1.044	high school-biased productivity
σ_a^2	0.401	variance of ability
π	0.145	ability persistence

Table 6: Model fit with persistent ability

	Moment	Data	Model
1	College graduates, %	42.0	42.4
2	Horizontal mobility, %	64.0	70.0
3	Upward mobility, %	24.0	22.6
4	College premium	1.79	1.84
5	Cost of college, %	4.80	4.76
6	Coef. of var.: col. earnings	0.60	0.55
7	Coef. of var.: hs earnings	0.55	0.57
8	Average discount factor	0.95	0.97
9	Capital-output ratio	3.30	3.05
10	Earnings persistence, %	18.0	17.8

Source: See Table 2.