# Efficiency and Equity of Education Tracking

## A Quantitative Analysis of Delaying School Tracking Age in Germany

Suzanne Bellue Lukas Mahler*

April 2022

**Abstract**

We study the long-run aggregate and distributional effects of education tracking policies (allocating students to different types of schools) by incorporating secondary school track decisions into a general-equilibrium heterogeneous-agent overlapping-generations model. The key ingredient in the model is a child skill production function that features accelerated skill growth when children are closer to their classroom peers in terms of their skills. We calibrate the model to German data and simulate an education reform that delays the school tracking age from age 10 to age 14. This reform leads to a decrease in aggregate income, but raises intergenerational mobility.

**Keywords:**

**JEL codes:**

*Bellue: University of Mannheim (suzanne.bellue@gess.uni-mannheim.de); Mahler: University of Mannheim (lukas.mahler@mail.uni-mannheim.de); We thank Antonio Ciccone, Minchul Yum, Michele Tertilt, and participants at various conferences and seminars for useful comments and discussions. Financial support from the German Science Foundation through CRC TR 224 (Project A?) and the SFB 884 Political Economy of Reforms is gratefully acknowledged.

# 1. Introduction

Brief intro into what tracking is. Outline the pro and con arguments frequently put forth. Explain that it is hard/impossible to answer research question with reduced form evidence alone and why. Give overview of our approach.

Stress that much of the quantitative framework is more or less standard. Highlight the key novelties, i.e. the zooming in on secondary schooling and track choice, paired with the child skill formation technology that depends on peers.

Highlight key results and where they come from. Explain counterfactuals and possible policy implications.

**Related literature**.

The remainder of the paper is organized as follows.

# 2. Education Tracking in a Simple Model of Childhood Skill Formation

In this section, we study the implications of education tracking using a stylized model of child skill formation during secondary school. The model is designed to capture the key mechanisms at work in our quantitative setup, while being parsimonious enough to be analytically tractable.[1] Our goal is to illustrate that in theory, (i) an education tracking policy generally creates winners and loser, (ii) the timing of education tracking ... (iii) the relative forces of each mechanism depend on the initial distribution of child skills upon entry into secondary school. *list key theoretical findings*. The quantitative model in Section 3 then embeds the simple model of secondary schooling in a richer, more realistic life-cycle framework that allows us to evaluate the quantitative importance of these mechanisms after an education tracking reform in the case of Germany.

We consider an environment in which a cohort of children arrive at the end of (comprehensive) primary school with a univariate skill level, denoted by $\theta_0$, that follow the distribution $G(\theta_0)$, which is known to everyone.[2] Secondary school is characterized by an initial comprehensive phase, during which children of all skill levels are pooled in one representative school track $C$, followed by a tracking phase, during which children are separated into two different school tracks, $T = A$ (academic

[1] It complements more detailed theoretical analysis as by e.g. Brunello, Giannini, and Ariga (2007) and Epple, Newlon, and R. Romano (2002) but is tailored to our quantitative setup.

[2] Discussion of why focus on only one type of skill

track) and $T = V$ (vocational track). Thus, denoting by $\tau$ the fraction of time spent in comprehensive school, the skills at the end of secondary school, $\theta_1$, are given by

$$\theta_1 = \tau\theta_1^C + (1-\tau)\theta_1^T. \tag{1}$$

We assume that each track differs in the quality of teaching that is measured in terms of child skills. Similar to Epple, Newlon, and R. Romano (2002), we further make the assumption that the teaching level in each track is determined solely by the average abilities of children in that track. *More extensive discussion or footnote.*[3]

Given these assumptions, we consider the following skill formation technology for child $i$ in school track $T$:

$$\theta_{i,1}^T = \theta_{i,0} + \alpha\,\mathbb{E}[\theta_0^T] + \beta(\theta_{i,0} - \gamma\,\mathbb{E}[\theta_0^T])\,\mathbb{E}[\theta_0^T], \tag{2}$$

where $\mathbb{E}[\theta_0^T]$ denotes the expected value of child skills in track $T$. Intuitively, this technology captures the notion of peer effects in education:[4] Children gain in terms of their skill development when they are exposed to peers that are more skilled on average ($\alpha > 0$). However, the benefit from better peers also depends on a child's own skill level relative to her peers. More precisely, if $\beta > 0$, there is a positive complementarity between own skills and average peer skills. (which leads to positive assortative matching). Finally, $\gamma$, captures the degree to which ... *not really sure, will become clearer below.*[5]

Lastly, similar to Brunello, Giannini, and Ariga (2007), we define tracking as some threshold skill level $\theta_0^*$, such that that every child with a skill level $\theta_{i,0} \leq \theta_0^*$ is allocated into the vocational track, and every child with $\theta_{i,0} > \theta_0^*$ is allocated to the academic track.[6] In the subsequent analysis, we first contrast the case of full Comprehensive schooling ($\tau = 1$) with that of full tracking ($\tau = 0$), to which we also refer to as Early Tracking. In Section 2.2, we then compare an Early Tracking system with a Late Tracking system ($0 < \tau < 1$). *Maybe not necessary to do that*

[3]Discussion of how we can interpret this, i.e. as teacher targeting the average guy, what the introduction of exogenous quality differences would change and maybe what a teacher choosing her teaching level would do.

[4]Talk about this literature here in more detail, summarized well in Epple and R. E. Romano (2011). Discuss briefly own empirical evidence?.

[5]Remark similarity of this functional form to variance.

[6]As we will see, such threshold level are the solution to both policymaker that seek to maximize the average child skill at the end of primary school and parents that seek to maximize their child's skill level. Discussion that in principle tracking need not be solely based on ability/skills.

## 2.1. Comprehensive Schooling versus Tracking

To simplify the analysis, we assume that the child skills, with which children arrive at the beginning of secondary school, are normally distributed around zero mean and with variance $\sigma^2 > 0$, i.e. $\theta_0 \sim \mathcal{N}(0, \sigma^2)$.[7] Then, given the child skill technology (2), we can establish the following results regarding skill formation in a fully comprehensive or early tracked secondary school system.

**Result 1** For any threshold level $\theta^*$, and $\alpha > 0, \beta > 0, \gamma \in (0, 1)$,

i. average end-of-school child skills in an early tracking system are larger than in a fully comprehensive system

$$\mathbb{E}[\theta_1^{ET}] = \beta(1-\gamma)(\sigma^2 - \mathbb{E}[Var(\theta_0|X)]) > 0 = \mathbb{E}[\theta_1^C], \tag{3}$$

where $X = 1$ if $\theta_0 > \theta^*$ and $X = 0$ otherwise.

ii. the variance of end-of-school child skills in an early tracking system is larger than in a fully comprehensive system. *needs to be shown analytically*

$$Var(\theta_1^{ET}) > \sigma^2 = Var(\theta_1^C) \tag{4}$$

Result 1.i. tells us that early tracking at any skill threshold $\theta^*$ (in the support of $\theta_0$) is always beneficial in terms of increasing average child skills at the end of secondary school relative to comprehensive schooling. Intuitively, this gain is achieved because tracking results in a lower conditional variance of skills within each track, $Var(\theta_0, X)$, compared to the unconditional variance $\sigma^2$. Thus, by creating more homogeneous peer groups, early tracking tracking improves overall learning.[8] At the same time, early tracking results in a more dispersed distribution of child skills at the end of secondary school.

*Pretty sure this result is sort of well known e.g. Epple, Newlon, and R. Romano (2002).*

Given this result, it becomes clear that a policy-maker, who is only concerned about maximizing average child skills will choose a skill threshold that minimizes the within-track variance, as is stated in the Result 2.

[7] Can we say this is without loss of generality? We can show that for math, language test score, this is a very reasonable assumption in the data.

[8] Maybe here (or later) a remark about what that would entail if there was the possibility for more than two tracks.

**Result 2** The threshold a skill-maximizing policy-maker would choose is $\theta^{*,G} = 0$.
Proof. Check Antonio's derviation of FOC:

$$\mathbb{E}[\theta_0|X=1] + \mathbb{E}[\theta_0|X=0] = f(\theta^{*,G}/\sigma) * \theta^{*,G}$$
$$\iff \sigma * (\frac{1}{1-F(\theta^{*,G}/\sigma)} - \frac{1}{F(\theta^{*,G}/\sigma)} = \theta^{*,G},$$

where $F(.)$ is the CDF of a standard normally distributed random variable. We know that the expression of the LHS is a monotonically increasing in $\theta^{*,G}$ and has a root at 0 (*not 100% sure about increasing but certainly around zero*).

That is, by exactly splitting the initial distribution of child skills at its mean, a policy-maker would create the most homogeneous peer groups possible. (provided there can only be two tracks).

*One possibility is to now talk about the winners and losers from such a tracking threshold relative to comprehensive schooling. But given that the results can be ambiguous (for kids in ET-A) I would probably prefer to include such a discussion only in the appendix and focus on incentive-compatible thresholds.*

While in some cases, education tracking policies take the form of such an exogenously mandated threshold, in most cases, and especially in Germany, typically the parents of a child can make a school track decision. To study tracking in such an environment, we again assume that parents also observe the distribution of child skills at the beginning of secondary school and know the child skill formation technology (2). Moreover, parents are solely interested in maximizing the skill of their child at the end-of secondary school. Result 3 then shows that the optimal threshold a policy-maker would choose does not necessarily coincide with an incentive-compatible threshold $\theta^{*,IC}$ at which parents would endogenously track.

**Result 3**

i. An incentive-compatible child skill threshold $\theta^{*,IC}$ solves

$$\gamma\sigma f(\theta^{*,IC}/\sigma) * (\frac{1}{1-F(\theta^{*,IC}/\sigma)} - \frac{1}{F(\theta^{*,IC}/\sigma)} = \theta^{*,IC} + \frac{\alpha}{\beta}, \tag{5}$$

where $F(.)$ is the CDF of a standard normal random variable and $f(.)$ is its density function.

ii. If $\alpha > 0$, $\beta > 0$, $\gamma \in (0,1)$, then the incentive-compatible is, if it exists, unique and is smaller than the one a policy-maker would choose.

$$\theta^{*,IC} < 0 = \theta^{*,G} \tag{6}$$

iii. If $\gamma > 0$, early tracking at threshold $\theta^{*,IC}$ entails a skill loss relative to comprehensive schooling for children around $\theta^{*,IC}$, where the mass of "losers" from tracking is given by $F(q_1^*) - F(q_2^*)$ with

$$q_1^* = \theta^{*,IC} - \theta\,\mathbb{E}[\theta_0|X = 0]$$
$$q_2^* = \theta^{*,IC} - \theta\,\mathbb{E}[\theta_0|X = 1].$$

The remaining mass of children gain from early tracking.

Generally, a solution to (5) does not necessarily exist.[9] If $\frac{\alpha}{\beta} = 0$ then the incentive-compatible threshold equals the optimal threshold, from a policy makers point of view. However, with increasing $\frac{\alpha}{\beta}$, more parents will choose to send their children to the academic track than in an environment where the threshold is exogenously mandated. Figure illustrates this result.

The intuition behind this result is that, when the relative importance of the complementarity between own skills and average peer skills, governed by $\beta$, shrinks relative to the direct peer effect, governed by $\alpha$, parents have stronger incentives to send their child to the academic track to benefit directly from better peers. While doing so, they do not (fully) internalize the costs that their decision has as it decreases the average skill level in that track.

Finally, Results 3.iii. says that in a tracking system, where parents endogenously track their children at a threshold $\theta^{*,IC}$, not every child would necessarily be better off in terms of her skills at the end of secondary school, compared to a comprehensive schooling system. As illustrated in Figure, there is a mass of children whose skill levels are around the incentive-compatible threshold $\theta^{*,IC}$, that would prefer a comprehensive track. In fact, the marginal child $i$ that has a skill level of $\theta_{0,i} = \theta^{*,IC}$ would lose $-\beta\gamma(\sigma^2 - \mathbb{E}[var(\theta_0|X)]$. Thus, similar to the average child skill gain (2), this loss is proportional to the between-track variance in average child skills.

This notwithstanding, we know from Result 1.i. that the gain in skills experienced by those children to the right and left of $q_1^*$ and $q_2^*$, respectively, outweighs these losses, such that on average, tracking is still preferred to comprehensive schooling.

The key parameter that informs the range and magnitude of losers versus winners in this environment is $\gamma$. If $\gamma = 0$, everyone would always prefer tracking over

[9]Interestingly, a solution becomes less likely to exist if $\frac{\alpha}{\beta}$ increases. That is if the complementarity of own with peer effects become less important relative to past skills.

comprehensive schooling. With increasing $\gamma$, however, there mass of parents, who would prefer a comprehensive schooling system increases as well.[10]

## 2.2. Early versus Late Tracking

*Maybe introduce $\tau$ only here. Also maybe add a simple sketch of the timing of tracking*

In this section we analyse the implications of different timing of tracking. That is we allow the fraction of the total time that is initially spent in a comprehensive school track to be different from 0 and 1, so that $\tau \in (0, 1)$.

Given the simple additive nature of the skills obtained during the comprehensive part of secondary school and the skills obtained during the tracking part of secondary school, it is immediately clear that any $\tau < 1$, will result in higher average child skills at the end of secondary school compared to full comprehensive schooling, $\tau = 1$. Moreover, average child skills are maximized if $\tau = 0$. That means early tracking is preferred to late tracking.

*There is a trade-off when we consider the variance -> I still have to look at that.*

Furthermore, the optimal skill thresholds that a policy-maker, or parents endogenously, would choose that are described in Results 2 and 3 remain the same.

A popular argument that is often brought forward in support of late tracking policies is the idea that a child's "true" academic ability is inherently uncertain and becomes observable only with age (Pekkarinen, Uusitalo, and Kerr, 2009). Especially young children are therefore likely to experience unexpected persistent skill "shocks" during their schooling career, which may render an earlier tracking decision ex-post inefficient. To illustrate this point, we expand our simple child skill formation model as follows.

We assume that the "true" child skills with which children learn are given by

$$\tilde{\theta}_0 = \phi\theta_0 + \varepsilon. \tag{7}$$

Here, $\theta_0$ represent the child skills that are observable to everyone at the beginning of secondary school, $\phi$ is a persistency coefficient, and $\varepsilon \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ represents an iid skill shock. We assume that this shock is realized after a fraction $\tau$ of time spent in secondary school. We are interested in a comparison of an early tracking system, where children are allocated into the two tracks based on the observed skills $\theta_0$, and a late tracking system, where children are taught comprehensively for the initial fraction $\tau$ of time in secondary school, and then tracked into two tracks based on the

[10]It is interesting to speculate about the political economy consequences of this result.

skills $\tilde{\theta}_0$.[11] Importantly, we assume that in the early tracking system, no re-tracking is possible, i.e. once children are allocated into the two tracks they remain there until the end of secondary school.[12]

**Result 3** Given (7), Late Tracking after time fraction $\tau$ results in higher average child skills compared to Early Tracking iff

$$(1-\tau)Var(\mathbb{E}[\tilde{\theta}_0|\tilde{X}]) > Var(\mathbb{E}[\tilde{\theta}_0|X]), \tag{8}$$

where $\tilde{\theta}_0 = \phi\theta_0 + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Moreover, $X$ and $\tilde{X}$ are defined as:

$$X = \begin{cases} 1 & \theta_0 > \theta^* \\ 0 & else \end{cases} \qquad \tilde{X} = \begin{cases} 1 & \tilde{\theta}_0 > \tilde{\theta}^* \\ 0 & else, \end{cases}$$

and $\theta^*$ is the skill threshold in Early Tracking, while $\tilde{\theta}^*$, is the threshold in Late Tracking.

i. If $\sigma_\varepsilon^2 = (1-\phi^2)\sigma^2$, then $\tilde{\theta}^* = \theta^*$ and (8) reduces to:

$$\tau < \frac{\sigma_\varepsilon^2}{\sigma^2} = (1-\phi^2). \tag{9}$$

ii. If $\phi = 1$, and $\tilde{\theta}^* = \theta^* = \theta^{*,G} = 0$, (8) reduces to:

$$\tau < 1 - \frac{\sigma^2}{\sigma^2 + \sigma_\varepsilon^2}. \tag{10}$$

Result 3 provides a general condition under which average child skills at the end of secondary schooling can be larger in a late tracking system compared to an early tracking system.[13] Intuitively, (8) trades off a loss from early tracking during the initial time fraction $\tau$, during which average child skills remain at 0, with a potential gain during the remaining time fraction. The loss comes from the fact that, with late tracking, we forgo the benefits of having more homogeneous peer groups early on. The potential gain comes from the fact that, because later tracking allows us to take into account (more) realized uncertainty in child skills before the tracking decision,

[11]This is the most basic model that allows us to illustrate the trade-off between early and late tracking. A more comprehensive analysis would assume that there is remaining uncertainty over child skills even after the late tracking decision.

[12]This is a strong assumption, yet track switching during secondary school is in the data relatively uncommon.

[13]Imo, we can view this as a "snapshot", comparing some tracking age, with one that tracks a bit later, but where, in the meantime, you learned a bit better about the "true" child skills.

the peer groups in each track will be more homogeneous during the remaining time fraction.[14]

Result 3.i. considers a case when the persistence parameter, $\phi$, and the skill shock variance $\sigma_\varepsilon^2$ are such, that the overall variance of $\tilde{\theta}_0$ is exactly equal to that of $\theta_0$. Thus, the child skill shocks represent a "reshuffling" of child skills during secondary school (and their support does not pan out). In such a case, it is clear that the skill thresholds chosen by parents or the government in early and late tracking will be identical. Moreover, late tracking will result in higher average child skills at the end of secondary school, if the fraction of time spend in comprehensive schooling $\tau$ is smaller than the ratio of the shock variance to the variance of initially observed child skills.

Result 3.ii. describes another special case, where child skills are allowed to "expand" due to the shock, as $\phi = 1$. Equation (10) provides the condition for average child skills to be higher in a late tracking system, if a government would impose its optimal threshold ($\theta^{*,G} = 0$) in both systems. Without shocks to child skills ($\sigma_\varepsilon^2 = 0$), it can never be optimal to have a non-zero time fraction in comprehensive school. However, as $\sigma_\varepsilon^2$ increases, the time fraction $\tau$ that can be spent in comprehensive schooling increases and reaches 1 as $\sigma_\varepsilon^2 \to \infty$.

## 2.3. Extensions

Misallocation due to parental preferences

Parents take into account labor market prospects

something else?

[14] I say potential gain, because this need not be the case. It depends on the thresholds (and the variance).

A discussion of our assumptions on the child skill formation, its theoretical implications and supportive empirical evidence from the literature and our own.

Instead of laying out a formal theory of the optimal timing of school tracking, we defer to existing research such as Brunello  Giannini (2004) and restrict our analysis to an illustration of the main channels at work. Main argument here is that the effects on human capital (total) are uncertain and depend a lot on the distribution -> central task of our exercise.

Central arguments that should come out of simple theory are: Assuming policymaker has a choice between tracking early or later, cares (only) about mean (and variance) of child skills at the end of school, child skill formation technology features penalty from distance to exogenously given teaching levels in all tracks. Then, tracking early comes at the benefit that students learn for a longer time in homogeneous peer groups. Tracking later, however, creates a more equal group once tracking starts and hence makes learning in the later periods after tracking more efficient. This is partly because uncertainty shocks to the child skills are more resolved and partly because comprehensive schooling makes everyone more equal. Optimal timing should depend on initial distribution(?). Later tracking might lead to lower variance but lower mean as well.

How to move from simple model to quantitative model. Considerations that have not been done in simple model. Might be good to have both types of labor. There are possibilities of switching. Aggregate skills may not be the only target of a social planner etc.

# 3. Full Model

The model augments a general equilibrium framework of human capital formation across generations with education tracking during secondary schooling. Summary of what GE, tracking means?

We assume that time is discrete and infinite, and that one model period, $j$, corresponds to 4 years in real life. This frequency allows us to investigate meaningful variations in school tracking ages. The dynastic structure implies that there are 16 generations alive at every point in time. As in Lee and Seshadri (2019), we assume that there is a unit mass of individuals in each period. A life-cycle can be structured into several stages, as illustrated in Figure 1.

In period $j = 1$ each individual becomes the parent of a 2-year-old child.[15] For

[15] We choose this somewhat unorthodox timing, such that children are 10 years old when parents make the secondary school track decision, which resembles reality in Germany

$j = 1, ..., 4$, the parent makes all decisions for the household, including the education decisions of the child. These include the decision in which school track to enroll the child at the beginning of secondary school in $j = 3$. In line with the argument in Section 2, we assume that there are two representative school tracks in the model economy, a vocational school track and an academic school track.

The child finishes secondary school in the same track that it started in.[16] However, in period $j = 5$, when the child becomes an independent adult and finishes school, she can decide whether to switch tracks and pursue academic higher (college) education or start working in the vocational labor market. This allows for the possibility of "second chances" for high-skilled children to obtain an academic degree while at the same time accounting for the fact that not all graduates from academic school tracks go on to study in college.

The periods $j = 6, ..., 16$ are the working and parenting years of each individual. During that time, she makes consumption, savings and labor supply decisions. Moreover, when her own child turns 18 and becomes independent, the parent decides on an inter-vivos transfer. In the remaining model periods, the individual is retired and makes consumption and savings decisions while earnings retirement benefits.

Throughout their lives, there are a number of frictions that can affect the decisions of the model agents. First, there markets are incomplete as both adult human capital as well as child skills are subject to unisurable idiosyncratic uncertainty. Second, in each periods, individuals can only borrow an amount that is constrained by what they can fully repay in the next period using a government lump-sum transfer. Third, in line with the evidence presented in Section 2, parents have preferences to send their child to the same school track that they themselves went to. Finally, to possibility to obtain academic higher education is constrained by a child's school track, their human capital level (and luck).[17] In the following, we present the recursive formulation of the agents decisions in more detail.

Anything else? Talk about preferences here? The two labor markets and human capital growth and income (taxation)? Credit markets?

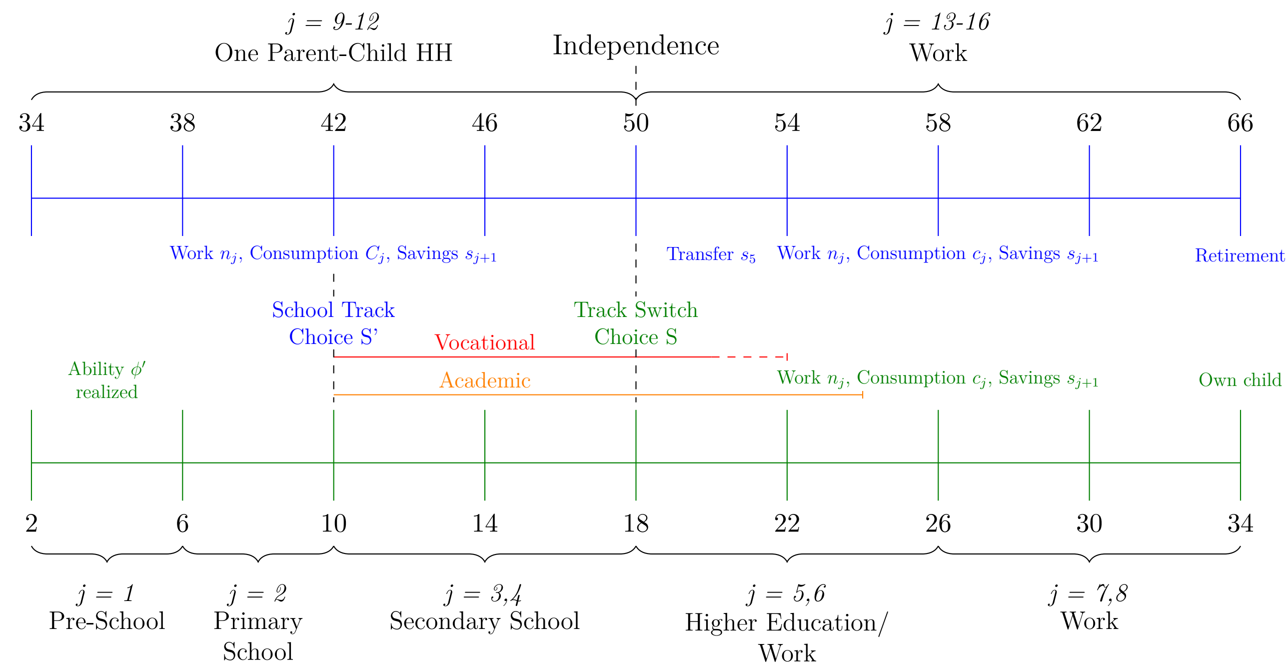
## 3.1. Individual Problems

**Period $j = 5$: Independence** At the beginning of the period, the state space of the newly independent adult comprises the secondary school track she graduated

[16] In reality, we do see a limited number of switches during secondary school. However,...

[17] In principle, only graduating from the academic school track allows for access to college education. However, in practice, there are a number of second chance options through which individuals with a vocational school track degree can gain access to academic higher education institutions.

**Figure 1:** Timeline of Life-cycle Events




from $S_c \in \{V, A\}$, where $V$ refers to the vocational school track and $A$ denotes the academic school track, her human capital $h_5$, and an initial savings level $s_5$ which she obtains as a financial transfer from her parent and takes as given. The last state, $\phi$, denotes innate ability of the individual which is kept as it will be transferred (imperfectly) to her child.

The first decision an independent adult makes after leaving secondary school, is whether to obtain academic higher (college) education $S_p = A$ or not $S_p = V$, where $W_5$ denotes the interim value for every higher education decision. The advantage of getting college education is that it gives an individual access to the labor market for college educated workers, which pays a different wage $w_A$ and also results in a different exogenous growth rate of human capital $\gamma_A$. The costs of going to college are twofold. First, it reduces the time available for work in $j = 5$ (at a wage $w_V$), as individuals spend part of their total time endowment of 1 studying, $\bar{x}(A) < 1$. Second, a non-pecuniary cost $\kappa_A$ incurs when going to college. This cost is dependent on the secondary school track $S_c$ and human capital $h_5$. The purpose of this cost is to parsimoniously model the observation that, while it is generally possible to obtain college education even after not graduating from an academic track secondary school, it is much more challenging.[18] Furthermore, we allow the college cost $\kappa$ to be

[18] In Germany, every graduate from an academic track secondary school gets an official qual-

dependent on the human capital level. The reason is that, in our model we assume that all college entrants also graduate from college and there is robust empirical evidence that graduating from college becomes easier the higher the human capital level you enter with.[19]

Note that, compared to studies that focus on the U.S. we do not model college costs as monetary costs. This is because most colleges in Germany are public and have very low tuition fees. Moreover, $\kappa_A$ directly captures preferences against college education for graduated from vocational schools (maybe due to skill complementarities etc.).

$$
\begin{aligned}
V_5(S_c, h_5, s_5, \phi) = \max_{S_p \in \{V, A\}} \{ & W_5(S_p = V, h_5, s_5, \phi) - \kappa_A(S_c, h_5), \\
& (\pi(S_c, (h_5)?\,) W_5(S_p = A, h_5, s_5, \phi) \\
& + (1 - \pi(S')) W_5(S = V, h_5, s_5, \phi) + \chi_{S'=A})
\end{aligned}
$$

$$
\begin{aligned}
W_5(S, h_5, s_5, \phi) = & \max_{\substack{c_5 > 0, s_6 \geq \underline{s}, \\ n_5 \in [0 \bar{x}(S)]}} \left\{ \underbrace{\frac{c_5^{1-\sigma}}{1-\sigma} - b \frac{n_5^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}}_{=u(c5, n5)} + \beta \int V_6(S, h_6, s_6, \phi) dF( \underbrace{\varepsilon_{j+1}}_{\text{iid luck shock}} ) \right\} \\
\text{s.t. } & c_5 + s_6 = y_5(w_{S_p} h_5 n_5; TT) + (1 + r) s_5 \\
& h_6 = \gamma_{5, S_p} h_5 \varepsilon_6
\end{aligned}
$$

Conditional on the higher education decision, the individual's problem in $j = 5$ is standard, as she chooses optimal consumption $c_5$ and savings $S_6$, subject to the life-cycle borrowing constraint $\underline{s}$, as well as hours worked $n_5$. The function $y_5$ gives an individual's labor income net of government taxes and transfers $TT$, while $r$ is the interest rate on savings. Finally, future values are discounted by $\beta$ and subject to uncertainty with respect to the realization of an idiosyncratic market luck shock to human capital $\varepsilon_6$ that follows the distribution $F$.[20] Thus, human capital in the next period $h_6$ is the product of this period's human capital, the exogenous,

ification that allows for access to academic higher education institutions, while graduates from vocational tracks do not. To go to college, these must either get a qualification through "evening schools", or may be allowed access to certain university degrees after having obtained a higher vocational degree or after having worked for a certain number of years.

[19]Explain evidence and that we also experimented with other specs but this is more parsimonious and we don't lose anything

[20]We assume that these are permanent shocks as in Lee and Seshadri (2019).

education-specific growth rate $\gamma_{Sp}$ and the market luck shock $\varepsilon_6$.

**Periods $j = 6, 7, 8$: Young adults without child** In this life stage, the model agents have completed all secondary and higher education but do not have own children yet. The choose optimal consumption, savings, and labor supply to solve the standard life-cycle problem.

$$V_j(S_p, h_j, s_j, \phi) = \max_{\substack{c_j > 0, s_{j+1} \geq \underline{s}, \\ n_j \in [0,1]}} \left\{ u(c_j) + \beta \int V_{j+1}(S_p, h_{j+1}, s_{j+1}, \phi) dF(\varepsilon_{j+1}) \right\}$$

$$\text{s.t. } c_j + s_j = y_j(w_{S_p} h_j n_j; TT) + (1 + r)s_j$$
$$h_{j+1} = \gamma_{j,S_p} h_j \varepsilon_{j+1}$$

In $j = 8$, the individuals know that they will become parents next period. For that reason, they take expectations over the learning ability of their future child $\phi'$, which are drawn according to $\phi' \sim G(\phi'|\phi)$, on top of the expectations over the market luck shocks.

**Period $j = 9$: Parent with newborn child** As the newborn child enters, a household is now comprised of a parent and a child. We denote by $q$ an adult consumption-equivalent scale so that $c_9$ remains the consumption of the parent (Yum, 2020). The child's learning ability $\phi'$ is realized and becomes the relevant state variable. Next period's child skills $\theta_2$ are a function of this learning ability and a constant $\eta(S_p)$, which depends on the parental education. We introduce this constant to allow for the fact that parents provide heterogeneous inputs into their child's skill evolution.[21] While we remain agnostic about the source of these different inputs and the specific channel through which they affect child skills, it is important to include such a factor to distinguish between innate differences in ability and post-birth differences.

[21] These comprise time and monetary inputs (CITE) but can also stem from different learning environments, interactions, etc. It is not clear whether these necessarily affect the time endowment, budget constraint or utility of a parent, it could just be differences

$$V_9(S_p, h_9, s_9, \phi') = \max_{\substack{c_9>0, s_{10}\geq \underline{s}, \\ n_9\in[0,1]}} \left\{ u(\frac{c_9}{q}, n_9) + \beta \int V_{10}(S_p, h_{10}, s_{10}; \theta_2, \phi') dF(\varepsilon_{10}) \right\}$$

$$\text{s.t. } c_9 + s_{10} = y_j(w_S h_9 n_9; TT) + (1+r)s_9$$
$$h_{10} = \gamma_{9,S_p} h_9 \varepsilon_{j10}$$
$$\theta_2 = \phi' \eta(S_p)$$

**Period $j = 10$: Parent with child in primary school** Every child goes to one period of comprehensive primary school. The state space now includes the child skill level $\theta_2$. Child skills evolve according to the function $f(\theta_2, \Theta_2, \eta(S_p))$. As discussed in Section 2, this function features increasing child skill gains when the distance to the teaching level in the school track is smaller. We assume that the teaching level, in terms of child skills, is equal to the average child skills in a school track.[22], such that in $j = 10$, the teaching level corresponds to the mean of the overall average child skill distribution, which we denote by $\Theta_2$. Given $\Theta_2$, parents form expectations over the evolution of the distribution of child skills according to .

Finally, future child skills are subject to an idiosyncratic shock $\nu$, which follows the distribution $H(\nu)$. (Daruich, 2020).

$$V_{10}(S_p, h_{10}, s_{10}; \theta_2, \phi') = \max_{\substack{c_{10}>0, s_{11}\geq \underline{s}, \\ n_{10}\in[0,1]}} \left\{ \begin{array}{c} u(\frac{c_{10}}{q}, n_{10}) + \\ \beta \int V_{11}(S_p, h_{11}, s_{11}; \theta_3, \phi', \Theta_3) dF(\varepsilon_{11}) dH(\nu_3) \end{array} \right\}$$

$$\text{s.t. } c_{10} + s_{11} = y_j(w_{S_p} h_{10} n_{10}; TT) + (1+r)s_{10}$$
$$h_{11} = \gamma_{10,S_p} h_{10} \varepsilon_{11}$$
$$\theta_3 = f(\theta_2, \Theta_2, \eta(S_p))\nu_3$$
$$\Theta_3 = \Gamma(\Theta_2)$$

**Period $j = 11$: Parent makes school track decision** In the beginning of the period, after the parent observes the realization of her child's skills, $\theta_3$, she makes the decision whether to send her child to the vocational or academic track school, $S_c \in \{V, A\}$. This decision is affected by the value of placing her child in each track, $W_{11}$, but also by a fixed preference shifter, $\xi(S_p, S_c)$, that depends on the child's but also on the parent's track. The purpose of this shifter is to model the bias in the

[22]Justify this assumption

school track decisions that we observe in the data (see before).

$$V_{11}(S_p, h_{11}, s_{11}; \theta_3, \phi', \Theta_3) = \max_{S_c \in \{V,A\}} \{W_{11}(S_p, h_{11}, s_{11}; S_c, \theta_3, \phi', \Theta_3) - \xi(S_p, S_c)\}$$

$$W_{11}(S_p, h_{11}, s_{11}; S_c, \theta_3, \phi', \Theta_3) = \max_{\substack{c_{11}>0, s_{12}\geq \underline{s}, \\ n_{11}\in[0,1]}} \left\{ \begin{array}{c} u(\frac{c_{11}}{q}, n_{11}) + \\ \beta \int V_{12}(S_p, h_{12}, s_{12}; S_c, \theta_4, \Theta_4, \phi') dF(\varepsilon_{12}) dH(\nu_4) \end{array} \right\}$$

$$\begin{aligned} \text{s.t. } c_{11} + s_{12} &= y_j(w_{S_p} h_{11} n_{11}; TT) + (1+r)s_{11} \\ h_{12} &= \gamma_{11,S_p} h_{11} \varepsilon_{12} \\ \theta_4 &= f(\theta_3, \Theta_3, \eta(S_p))\nu_4 \\ \Theta_4 &= \Gamma(\Theta_3) \end{aligned}$$

**Period $j = 12$: Parent with child in secondary school** The child remains in the school track chosen in the prior period.[23] There is no aggregate state for the child skill distribution in the next period, as the child then becomes independent and finishes secondary school.

$$V_{12}(S_p, h_{12}, s_{12}; S_c, \theta_4, \Theta_4, \phi') = \max_{\substack{C_{12}>0, s_{13}\geq \underline{s}, \\ n_{12}\in[0,1]}} \left\{ \begin{array}{c} u(\frac{c_{12}}{q}, n_{12}) + \\ \beta \int V_{13}(S, h_{13}, s_{13}; S', \theta_5, \phi') dF(\varepsilon_{13}) dH(\nu_{13}) \end{array} \right\}$$

$$\begin{aligned} \text{s.t. } c_{12} + s_{12} &= y_j(w_{S_p} h_{12} n_{12}; TT) + (1+r)s_{12} \\ h_{13} &= \gamma_{12,S_p} h_{12} \varepsilon_{13} \\ \theta_5 &= f(\theta_4, \Theta_4, \eta(S_p), \nu_5) \end{aligned}$$

**Period $j = 13$: Parent when child becomes independent** Just before her child reaches the age of 18 and becomes independent, the parent decides on a financial transfer that her child receives, $s_5$, while taking into account the child's future value. As in Daruich (2020), we model this as an interim decision problem and assume that the parent already knows the realization of all shocks. The strength of the parental

[23] In reality, switching between tracks during secondary school is possible. However, in the data we observe only x%, which is why we neglect that possibility.

altruism motive is governed by $\delta$.

$$V_{13}(S_p, h_{13}, s_{13}; S_c, \theta_5, \phi') = \max_{s_5 \geq 0} \left\{ \tilde{V}_{13}(S_p, h_{13}, s_{13} - s_5) + \delta V_5(S_c, s_5, \theta_5, \phi') \right\}.$$

The transfer cannot be negative, so that parents cannot borrow against the future income of their child.

$$\tilde{V}_{13}(S_p, h_{13}, s_{13}) = \max_{\substack{c_{13}>0, s_{14} \geq \underline{s}, \\ n_{13} \in [0,1]}} \left\{ u(c_{13}) + \beta \int V_{14}(S_p, h_{14}, s_{14}) dF(\varepsilon_{14}) \right\}$$

$$\text{s.t. } c_{13} + s_{14} + s_5 = y_j(w_{S_p} h_{13} n_{13}; TT) + (1+r)s_{13}$$

$$h_{14} = \gamma_{13,S_p} h_{13} \varepsilon_{14}$$

**Periods $j = 14, \ldots$ Remaining work life and retirement** During model periods $j = 14, 15, 16$, the agent continues to work until she retires at age 66.

$$V_j(S, h_j, s_j) = \max_{\substack{c_j > 0, s_{j+1} \geq \underline{s}, \\ n_j \in [0,1]}} \left\{ u(c_j) + \beta \int V_{j+1}(S, h_{j+1}, s_{j+1}) dF(\varepsilon_{j+1}) \right\}$$

$$\text{s.t. } c_j + s_{j+1} = y_j(w_S h_j n_j; TT) + (1+r)s_j$$

$$h_{j+1} = \gamma_{j,S} h_j \varepsilon_{j+1}$$

In the remainder of her life (model periods $j = 17, 18, 19, 20$), the agent no longer works but receives retirement benefits $\pi_j(h_j)$, which depend on the last human capital level before retirement.[24]

$$V_j(S, h_j, s_j) = \max_{c_j > 0, s_{j+1} \geq \underline{s}} \{ u(c_j) + \beta V(S, h_{j+1}, s_{j+1}) \}$$

$$\text{s.t. } c_j + s_{j+1} = \pi_j(h_j) + (1+r)s_j$$

$$h_{j+1} = h_j$$

## 3.2. Aggregate Production, Government, and Equilibrium

We assume that a representative firm produces output according to the Cobb-Douglas production function $Y = AK^{\alpha}H^{1-\alpha}$, where $A$ denotes total factor productivity, $K$ is the aggregate physical capital stock and $H$ is a CES aggregate of total

[24] Explain this simplifying assumption

labor supply, which is defined by:

$$H = [\omega H_V^\epsilon + (1-\omega) H_A^\epsilon]^{\frac{1}{\epsilon}}. \tag{11}$$

Here, $H_V$ is the aggregate labor supply in efficiency units of workers with vocational higher education, and $H_A$ that of workers with academic (college) higher education.

The government taxes only labor income progressively according to ... . It uses the tax revenue to finance retirement benefits $\pi_j$ as well as fixed lump-sum social welfare benefits $TT$. These may include child allowances, unemployment benefits, or contributions to health insurance.

# 4. Calibration

# 5. Quantitative Results

# 6. Conclusion

# References


Brunello, Giorgio, Massimo Giannini, and Kenn Ariga (2007). "The optimal timing of school tracking: a general model with calibration for Germany". In: *Schools and the equal opportunity problem*, pp. 129–156.

Daruich, Diego (2020). "The macroeconomic consequences of early childhood development policies". In: *FRB St. Louis Working Paper* 2018-29.

Epple, Dennis, Elizabeth Newlon, and Richard Romano (2002). "Ability tracking, school competition, and the distribution of educational benefits". In: *Journal of Public Economics* 83.1, pp. 1–48.

Epple, Dennis and Richard E Romano (2011). "Peer effects in education: A survey of the theory and evidence". In: *Handbook of social economics*. Vol. 1. Elsevier, pp. 1053–1163.

Lee, Sang Yoon and Ananth Seshadri (2019). "On the intergenerational transmission of economic status". In: *Journal of Political Economy* 127.2, pp. 855–921.

Pekkarinen, Tuomas, Roope Uusitalo, and Sari Kerr (2009). "School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform". In: *Journal of Public Economics* 93.7-8, pp. 965–973.

Yum, Minchul (2020). *Parental time investment and intergenerational mobility*. Tech. rep. Working Paper Series.

# A. Appendix