Essays in Family Macroeconomics

Inauguraldissertation zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften der Universität Mannheim

Suzanne Bellue

Frühjahrs-/Sommersemester 2023

Abteilungssprecher: Prof. Klaus Adam, Ph.D.
Vorsitzender der Disputation: Prof. Dr. Antonio Ciccone
Referentin: Prof. Michèle Tertilt, Ph.D.
Koreferent: Prof. Minchul Yum, Ph.D.
Tag der Disputation: 30. Mai, 2023

Acknowledgments

First and foremost, I am deeply grateful to my main advisor, Michèle Tertilt, and my co-advisors, Klaus Adam, Antonio Ciccone, and Minchul Yum, for their outstanding guidance and support throughout my Ph.D. studies. Michèle's exceptional supervision, weekly meetings, and fruitful discussions were instrumental in helping me define, develop, and finish my dissertation. Her encouragement to apply to conferences and meet with external and internal researchers greatly enriched my doctoral experience. I am profoundly indebted to her for investing so much time and effort in my academic growth. I am also indebted to Minchul for his constant availability and generous sharing of his knowledge and expertise, which significantly contributed to the quality of my research projects. Finally, I am deeply grateful to Klaus and Antonio for providing a stimulating academic environment through their weekly colloquium and applied seminar. Klaus' attentive listening to my research project at different stages and his remarkable ability to grasp any topic instantly have pointed me in the correct research direction. I am profoundly grateful for Antonio's willingness to engage in long, thoughtful conversations that have challenged my thinking and enriched my understanding of the subject matter. I feel fortunate to have had such exceptional and approachable advisors whose guidance and support have been invaluable in my academic journey.

I have had the pleasure of working with Lukas Mahler on the second chapter of this dissertation, and I am grateful for our fruitful work together. Additionally, I have been in the company of a talented group of Ph.D. students, including Giovanni, Ursula, Jacopo, Tommaso, Pia, Tsung-Hsien, Alexander, Valentina, Laura, Michaela, Oliver, Andrés, Marco, Jonathan, and Jan, who made my Ph.D. studies a stimulating and enjoyable experience.

Moreover, I am grateful for the dynamic academic environment at Mannheim University. I want to thank Effrosyni Adamopoulou, Miren Azkarate-Askasua, Antoine Camous, Andreas Gulyas, Anne Hannusch, David Koll, Matthias Meier, Arthur Seibold, Konrad Stahl, Ernst-Ludwig von Thadden, and the seminar participants at the Mannheim Macro Seminar for their insightful comments and engaging discussions. Their feedback has been invaluable in shaping my dissertation.

I thank Corinna Jann–Grahovac, Ulrich Kehl, Golareh Khalilpour, Marion Lehnert, Caroline Mohr, Susanne Putz, and Natasa Stolze for their administrative support throughout my studies at Mannheim University. I am also thankful to Christoph Rothe, who helped me navigate the challenges of the job market process. I gratefully acknowledge financial support from the German Academic Exchange Service (DAAD) and the German Research Foundation (through the CRC-TR-224 project A03).

And finally, I want to thank my family: my parents, grandparents, and siblings for their presence and boundless kindness throughout my academic journey. I feel incredibly fortunate to have Patricio Cerda by my side. He is my rock and an everyday source of positivity. I owe much of my academic success to his unwavering support and belief in me.

Contents

Preface

This dissertation employs quantitative methods to study questions in family macroeconomics, with a focus on parental decisions and intergenerational mobility. It consists of two self-contained chapters.

Chapter 1 is titled "Why Don't Poor Families Move? A Spatial Equilibrium Analysis of Parental Decisions with Social Learning." In this first chapter, I propose a mechanism that rationalizes the observed heterogeneity in parental investment choices across socioeconomic groups in the United States. I develop a quantitative spatial and overlapping generations model in which parents decide on two parental inputs influencing their child's next period human capital: neighborhood quality and parental time. Importantly, I introduce imperfect information and social learning about one parameter of the child skill technology that governs the returns to the two parental inputs. Young agents learn about the unknown parameter using only information available at their neighborhood level. However, because of a selection neglect bias that leads to misinferences from non-representative samples, segregation generates information frictions that systematically distort parents' beliefs and behavior. Specifically, in equilibrium, parents raised in low-quality neighborhoods tend to underestimate the importance of parental inputs, and parents raised in high-quality neighborhoods tend to overestimate it.

I calibrate the model using several United States representative data sets. The calibrated model matches targeted and, more interestingly, non-targeted moments regarding parental behavior across socioeconomic groups. On the contrary, a model that matches the intergenerational persistence of income but assumes perfect information cannot rationalize the heterogeneity in parental inputs across socioeconomic groups without assuming highly heterogeneous preferences. I find that parents' beliefs about the importance of parental inputs increase the income Gini index by 3% and the intergenerational income rank coefficient by 12%. Finally, scaling up a housing voucher policy generates, in the long run, and in general equilibrium, a shift in parental beliefs that contributes to the reduction of inequality and to the improvement in intergenerational mobility.

Chapter 2 is titled "Efficiency and Equity of Education Tracking: A Quantitative Analysis" and is co-authored with Lukas Mahler. In this chapter, we investigate the question of early school tracking in Germany—the allocation of students to different types of schools. We develop a life-cycle overlapping generations model to evaluate the aggregate effects of a policy that would delay the school track decision by four years: from age ten to fourteen. Crucially, we incorporate linear classmate peer effects, non-linear school track instruction pace effects, and age-specific skill shocks into the child skill technology. We show analytically that this technology embeds this idea of learning gains through homogenous peer groups and that it can rationalize reduced-form empirical evidence on school tracking. The child skill technology parameters are one of the crucial elements that drive potential efficiency gains of early school tracking.

We then calibrate the model using multiple German representative data sets. Our calibrated model predicts that around 23% of lifetime earnings and 30% of lifetime wealth variation is already explained at age ten, the time of the school track choice. Conditioning on the initial school track choice alone accounts for 12% of lifetime earnings variation and 13% of lifetime wealth variation. Finally, we find that postponing the tracking age from ten to fourteen generates significant improvements in intergenerational mobility. However, these come at the cost of efficiency losses in aggregate economic output. The size of these losses depends on the design of the instruction levels in each school track.

Chapter 1

Why Don't Poor Families Move? A Spatial Equilibrium Analysis of Parental Decisions with Social Learning[1]

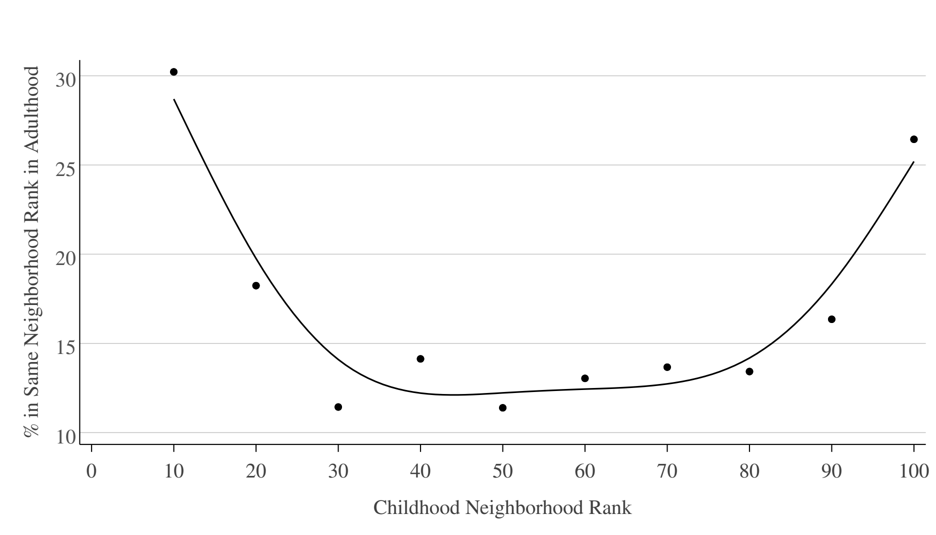
Abstract: In the United States, childhood neighborhood quality shapes adulthood economic opportunities. However, most children raised in bottom-quality neighborhoods still live in low-quality neighborhoods in adulthood. Could childhood neighborhood directly affect adulthood choices? I develop a quantitative spatial model of parental decisions that incorporates a novel mechanism: social learning about the technology of skill formation. Segregation generates information frictions that systematically distort parents' subjective beliefs and behavior. I calibrate the model using several United States representative data sets. The calibrated model matches targeted and non-targeted parental behavior and generates an endogenous distribution of subjective beliefs. I find a relatively modest level of delusion that increases the income Gini index by 3% and the intergenerational rank-rank slope by 12%. A housing voucher policy improves the neighborhood quality of eligible families, raising children's future earnings. When scaling up the policy, long-run and general equilibrium responses in subjective beliefs amplify the policy effects. Inequality reduces, and intergenerational mobility improves.

[1] I am indebted to Michèle Tertilt for her continued and invaluable support and guidance in this project. In addition, I am very grateful to Klaus Adam, Giovanni Ballarin, Antoine Camous, Antonio Ciccone, Anne Hannusch, Jamie Hentall MacCuish, Georgi Kocharkov, David Koll, Giacomo Ponzetto, José Víctor Ríos Rull, Jan Sun, Minchul Yum and seminar and conference participants at the XXV Workshop on Dynamic Macroeconomics 2022 (Vigo), the Summer School in Urban Economics 2022, the SMYE 2022, the CRC retreat 2022, the ENTER Jamboree 2022, UCL, the University of Mannheim for helpful discussions and suggestions. I gratefully acknowledge financial support from the German Academic Exchange Service (DAAD) and the German Research Foundation (through the CRC-TR-224 project A03).

1.1 Introduction

In the United States, childhood neighborhood quality shapes adulthood economic opportunities.[2] However, most children raised in bottom-quality neighborhoods still live in low-quality neighborhoods in adulthood. Figure 1.1 shows that while the probability that a child raised in a middle-range quality neighborhood also lives in a middle-range quality neighborhood in adulthood is less than 15%, this number rises to 30% for children raised in the bottom-decile of the neighborhood quality distribution. Given the benefits of escaping low-quality neighborhoods, why do families stay? Could childhood neighborhood quality directly affect adulthood choices?

Figure 1.1: Intergenerational Residential Mobility in the United States



Notes: This Figure shows the share of children who, in adulthood, choose to live in the same decile of the neighborhood quality distribution as their parents. The dots are data points; the solid line is a smooth interpolation—data source: Add Health; see Appendix 1.D.4 for details of data construction.

What if children raised in low-quality neighborhoods choose to stay because they systematically underestimate the returns to neighborhood quality for their children? I depart from perfect information and rational expectations assumptions and propose a new mechanism that endogenizes parental subjective beliefs about the child skill technology and generates systematic differences in subjective beliefs by childhood neighborhood. Suppose individuals' human capital depends on past parental inputs—, including neighborhood quality—and

[2] Chetty and Hendren (2018a) show that, in the United States, the neighborhoods in which children grow up shape their earnings, college attendance rates, and fertility and marriage patterns. I proxy neighborhood "quality" with neighborhoods' median household income that correlates with places' effects measured by Chetty and Hendren (2018b) but also with low poverty rates, low crime rates, and high-performing schools.

a random, idiosyncratic, and unobserved ability shock, which can be interpreted—to some extent—as luck. Assume people don't know the returns to parental inputs and must learn about them before making decisions. Young adults learn by observing their older neighbors' human capital and past parental inputs, i.e., through social learning. However, they only get an informative but bounded perception of their neighbors' abilities and cannot perfectly infer the returns. Crucially, I assume that people suffer from a selection neglect bias. They think of their neighbors as representative of the population.[3] This assumption is incorrect with income residential segregation—a form of spatial sorting partly based on ability shocks. Indeed, low-ability adults tend to be poor and live in low-rent and low-quality neighborhoods. Since young adults' perception of their neighbors' ability is bounded, in low-quality neighborhoods, where relatively many low-ability parents live, young adults overestimate the average local ability. They then implicitly attribute too much of the local average human capital to ability and underestimate the returns to parental inputs. The reverse happens in high-quality neighborhoods, where relatively many high-ability adults surround young adults. Combined with residential segregation, this mechanism leads to persistent delusion about the returns to parental inputs.

I incorporate this new mechanism into a quantitative spatial overlapping generations model with residential and parental time decisions. The model features heterogeneous agents, where parents choose the quality of their neighborhood and how much time to spend on their child's education. Residential segregation results from parents' location decisions and local rents, which are equilibrium objects. The child's next period human capital is a function of childhood neighborhood quality and parental time, parental human capital, and idiosyncratic ability shocks. Crucially, I depart from perfect information and introduce social learning with selection neglect. Agents are unaware of the returns to parental inputs: neighborhood quality and parental time. Children inherit subjective beliefs from their parents and update them by observing older neighbors' outcomes and history. By assumption, children only imperfectly see ability shocks and suffer from a selection neglect bias. In equilibrium, endogenous segregation generates systematic biases across socioeconomic groups and persistence in subjective beliefs within families. Children of low-subjective beliefs and poor parents are likely to live in low-quality neighborhoods—composed of relatively low-ability parents,— underestimate the returns to parental inputs, and become poor parents next period. The reverse happens to children of high-subjective beliefs and high-income parents. Agents differ in their human capital—primarily determined by their parents—and in their subjective beliefs—determined

[3] This cognitive bias called "selection neglect" or "assortativity neglect." Enke (2020) provides empirical evidence of it. Jehiel (2018) develops a theoretical model of over-optimism among entrepreneurs driven by selection neglect.

mainly by their parent and childhood neighborhood. Given the social learning mechanism, there are multiple critical equilibrium objects: the distribution of human capital, subjective beliefs, neighborhood choices, and local rents, which are endogenously determined as fixed points.

The model is solved numerically and calibrated to match a set of critical moments from several United States representative data sets. The model matches segregation and family earnings dispersion in the average commuting zone in the United States computed from the ACS 2000 and NHGIS 2000 data sets. In addition, it targets causal neighborhood effects on children's future incomes estimated by Chetty and Hendren (2018b) and social mobility measured by Chetty et al. (2014). Parents' decisions are disciplined by matching parental time in the ATUS survey and neighborhood quality choices from AddHealth. Even though the model does not feature preference heterogeneity, the calibrated model matches parental behavior across socioeconomic groups well. It provides a rationale for college parents spending more time with their children than non-college parents despite working more hours and matches non-targeted intergenerational residential mobility patterns well.[4]

In contrast, a perfect information version of the model cannot replicate the data without imposing sizable heterogeneous preferences across places of birth and education. To see this, I first re-calibrate the model shutting down the subjective beliefs channel. The calibrated model misses all the non-targeted residential mobility moments. I then introduce heterogeneous preferences across places of birth and add residential mobility moments as targets. I find that with this alternative model, agents born in bottom-quality neighborhoods must like their place of birth thirty times more than the others. This necessary feature is at odds with the empirical findings of Bergman et al. (2019) who compare low-income families who live in deprived neighborhoods and are randomly allocated between treatment and control groups. They find that parents in the treatment group, induced to move to higher-quality neighborhoods, are more satisfied and willing to stay in their neighborhood than those in the control group. Moreover, even if the alternative model matches residential mobility by construction, it fails to generate an education-parental time gap that is large enough.

I then use the calibrated model with social learning to understand parents' residential-quality choices and to conduct policy experiments. The first finding is that social learning and subjective beliefs explain a large share of the socioeconomic gap in parental decisions. Segregation generates information frictions that systematically distort parents' subjective beliefs concerning the technology of skill formation. Providing full information about the returns to parental inputs would improve low-income parents' subjective beliefs about the

[4] Parental time gap by education is a well-known puzzle described in Guryan et al. (2008). See Doepke et al. (2022) for a review.

returns by 17% and decrease those of high-income parents by 7%. Those numbers are of a reasonable order of magnitude compared to empirical estimates.[5] Importantly, I find that subjective beliefs have sizable effects on the economy. Children born in bottom-quality neighborhoods would be half as likely to remain in those neighborhoods in adulthood, and low-income parents would spend 31% more time with their children. Perfect information about the returns to parental inputs would decrease the rank-rank slope, which captures the degree to which children's incomes are determined by parents' incomes, by 12%, and the Gini index of income, a measure of inequality, by 3%.

The model features two key frictions that motivate a policy intervention. Parents cannot borrow against their children's future earnings, and due to the novel mechanism, segregation generates information frictions that distort parental decisions. Motivated by the evidence that housing vouchers improve the neighborhood quality of eligible families (Chetty et al., 2016), I use the calibrated model to study their effects on the United States economy. One could expect housing vouchers to decrease segregation, improve information, and dampen subjective beliefs' distorting effects. I consider a housing voucher that covers the difference between 30% of income and the rent up to a limit. Eligible households are parents from the bottom decile of the income distribution. In the first step, I study the partial equilibrium effects of the policy by simulating a randomized control trial within the model. Compared to the control group—eligible parents who do not receive housing vouchers,— eligible parents who receive the vouchers live in higher-quality neighborhoods, positively affecting their children's earnings. The effects on children's earnings are of a similar order of magnitude as the empirical estimates from Chetty et al. (2016). Subjective beliefs play a substantial role. If parents had perfect information about the returns to neighborhood quality, they would move to even higher quality neighborhoods, increasing their children's earnings at age 26 by an additional $132 per year.

The second finding is that, when scaling up the housing voucher policy, general equilibrium responses in local prices and, in particular, in subjective beliefs amplify the effects of the housing voucher policy. The voucher allows housing voucher holders to move to better neighborhoods, increasing the density in middle-range quality neighborhoods—especially at the rent limit—and forcing non-eligible households to move out. The housing market reaction creates winners and losers, with adverse effects in the aggregate; however, in the long run, information and subjective beliefs improve, particularly among low-income households, amplifying partial equilibrium effects on eligible families and generating positive aggregate

[5] Cunha et al. (2013) elicit maternal expectations about the technology of skill formation from a sample of socioeconomically disadvantaged African American women. The author's favorite estimates of the percentage difference between the actual and the believed returns to investment range between 4 and 39%.

effects. In the long run, intergenerational mobility improves by 3.8%, and inequality decreases by 0.8%.

Despite positive long-run effects, a housing voucher policy with a unique rent limit at the commuting zone has unintended consequences. It generates a bunching behavior of eligible households in middle-range quality neighborhoods, segregating the housing market and increasing the already overestimated subjective beliefs of rich families. I find that designing a housing voucher policy with rent limits set at a more granular level is a better tool for mitigating the distorting effects of parents' subjective beliefs. Under the alternative housing voucher policy, in the long run, intergenerational mobility improves by 8.5%, and inequality decreases by 1.5%.

Finally, I use the National Longitudinal Study of Adolescent to Adult Health to empirically verify two testable implications of the model. Since parental subjective beliefs are not measured in the data, controlling for all observable characteristics, I should still observe a positive correlation (a) between parental time and neighborhood quality and (b) between qualities of the childhood and adulthood neighborhood. I proxy parental time by the number of parent-child activities and neighborhood quality by the census tract median household income observed at ages 14 and 37. Both testable implications are empirically validated. Combined with the extensive literature on subjective beliefs and the great fit of the calibrated model, this suggestive evidence supports the plausibility of the proposed social learning mechanism.

Related Literature

This paper links several strands of the literature: the subjective beliefs literature, the quantitative family-macroeconomics literature, and the quantitative spatial economics literature.

First, this paper builds on empirical evidence from the parental subjective beliefs literature to carefully model endogenous parental subjective beliefs about the technology of skill formation. Since Cunha et al. (2013), a large body of research documents that actual technology of skill formation does not systematically differ by socioeconomic groups but that parents' subjective beliefs about the technology of skill formation differ, correlate with income, and influence their decisions (see for instance Jensen (2010); Attanasio and Kaufmann (2014); Kaufmann (2014); Caucutt et al. (2017); Boneva and Rauh (2016, 2018); Belfield et al. (2019); Dizon-Ross (2019); Wiswall and Zafar (2021)).[6] The idea that technology diffuses through social learning—by observing how older community members do—is consensual, and

[6] One exception is Attanasio et al. (2019) in the UK who does not find a socioeconomic gradient in subjective beliefs.

there is also evidence of it in education.[7] However, in education, because people sort based on the outcome—their human capital,—social learning effects often relate to selection. For instance, the fact that low-income people lack successful role models around them could explain the low level of motivation and effort (Nguyen, 2008; Alan et al., 2019; Breda et al., 2019; Algan et al., 2020; Guyon and Huillery, 2020). Social learning is one likely explanation for Chetty et al. (2022)'s recent findings; the share of high socioeconomic status friends in a ZIP code best predicts upward income mobility in the United States.

Second, this paper quantifies the macroeconomic effects of endogenous subjective beliefs by incorporating social learning into a quantitative spatial model of overlapping generations in which parents affect their child's human capital by choosing their neighborhood and parental time. By doing so, I contribute to the quantitative family macroeconomics literature. Human capital accumulation is modeled following recent macroeconomic papers such as Daruich (2018); Jang and Yum (2020); Kim et al. (2021); Yum (2022); Chyn and Daruich (2022). While in other sub-areas of macroeconomics, such as finance, subjective expectations are considered critical elements in explaining agents' investment behavior (see, for instance, Adam et al. (2017)), in family macroeconomics, heterogeneous subjective beliefs are usually ignored. Two major exceptions are Fogli and Veldkamp (2011) and Fernández (2013) who rationalize the change in female labor supply over time by a change in subjective beliefs. In this paper, the learning process builds on Fogli and Veldkamp (2011). A key difference is the introduction of residential choices which, combined with a selection neglect modeled as in Jehiel (2018), generates a bias, rationalizing the fact that low-income children living in poor-neighborhoods lack motivation due to a lack of successful role models. In equilibrium, the learning process generates a stable distribution of heterogeneous parental subjective beliefs that affects parental input choices. While the idea of neighborhood effects through social learning has been largely developed (see, for instance, Durlauf (2011)), very few papers have linked heterogeneous subjective beliefs with residential choices. Roemer and Wets (1994); Streufert (2000) are two exceptions. They provide different theoretical frameworks in which the selection induced by residential sorting could lead to systematic bias in subjective beliefs. However, this paper is the first to develop and calibrate a quantitative model with social learning.

Finally, the paper contributes to the quantitative spatial economics literature as it links segregation and inequality. I use methods from the quantitative spatial economics literature reviewed in Redding and Rossi-Hansberg (2017). Motivated by recent evidence of a causal relationship between exogenous neighborhoods and child's skills (Chyn, 2018; Chetty

[7] See for instance Conley and Udry (2010) who find that in rural Ghana, use of fertilizer by small farmers is boosted by the observation of surprisingly successful farmers.

and Hendren, 2018a,b; Hwang, 2022; Nakamura et al., 2022), the model displays exogenous amenities. As in Fogli and Guerrieri (2019); Eckert et al. (2019); Chyn and Daruich (2022); Gregory et al. (2022), neighborhoods' environment quality directly affects children's future human capital. By having endogenous valuations of neighborhood quality through subjective beliefs, I also contribute to the growing literature that considers endogenous neighborhood amenities (Ahlfeldt et al., 2015; Diamond, 2016; Bilal, 2021). A few papers have quantitatively linked segregation and inequality. Fogli and Guerrieri (2019) and Gregory et al. (2022) incorporate peer effects in spatial equilibrium, and Eckert et al. (2019); Chyn and Daruich (2022) model the effect of local taxes on neighborhood environment quality. This model provides and quantifies a new mechanism for the relationship between segregation and inequality. Segregation creates informational frictions: the more the economy is segregated, the steeper the socioeconomic gradient in parental investment.

The remainder of the paper proceeds as follows. Section 2 presents the spatial overlapping generations model. Section 3 explains the model calibration and presents some quantitative results. Section 4 uses the model for policy analysis. Section 5 empirically tests two of the model implications, and Section 6 concludes.

1.2 The Model

Consider one labor market composed of a finite number of neighborhoods populated by families of one parent and one child. Children's future earnings depend on childhood neighborhood quality, parental time, parental human capital, and an imperfectly observed idiosyncratic ability shock. Parents decide on two parental inputs of the child skill technology: in which neighborhood to raise their child and how much time to spend on their child's education. Sorting across neighborhoods within the labor market is only driven by families seeking better opportunities for their children. One of the key features of the model is imperfect information and social learning about the returns to parental inputs. Before making decisions, young agents learn about the returns using the information available at the neighborhood level.

The following sections describe the economic environment, the novel mechanism—social learning within neighborhoods—the parents' optimization problem, and the housing market. Then I give the equilibrium definition.

1.2.1 Economic Environment

Geography and Amenities: Consider one commuting zone with a finite number of ex-ante heterogeneous neighborhoods. Neighborhoods differ in quality m. Thus, a neighborhood is characterized by its quality m rather than its particular name. As wages do not vary across neighborhoods, neighborhood quality is the only exogenous amenity.

Families: The economy is populated by a continuum of families composed of one parent and one child. Time is discrete and each agent lives for two periods: childhood and parenthood. Parents choose in which neighborhood to raise their child and on how much time to spend on their child's education—parental time. Families are heterogeneous concerning four parental characteristics: accumulated human capital, college graduation status, the neighborhood of birth, and subjective beliefs about the technology of skill formation.

In the following, primed letters correspond to children's next period variable, lowercase letters to parents' variables, and uppercase letters to grand-parents' variables.

Technology of Skill Formation: Children's next period human capital h' mainly depends on their parent as it is a function of their childhood neighborhood quality m, parental time τ, parental human capital h, and ability shock a'. The functional form is as follows:

$$\begin{aligned} h' &= (i(m,\ \tau)\ +\ \underline{i})^{\alpha}\ h^{\ \beta}\ \exp(a') \\ i(m,\ \tau) &= \left(\gamma\ \left(\frac{\tau}{\overline{\tau}}\right)^{\varphi} + (1-\gamma)\, m^{\varphi}\right)^{\frac{1}{\varphi}}, \end{aligned} \tag{1.1}$$

where $\alpha,\ \beta,\ \gamma \in (0,1)$ and the child's ability shock a' is uncorrelated with parental characteristics and drawn from a normal distribution $\mathcal{N}(0, \sigma_a)$.[8] Crucially, ability shocks are imperfectly observed.

By assumption, parental human capital h enhances the productivity of the two parental inputs. This feature seeks to capture that high-human capital parents are better at building child skills and that environmental factors, such as in-utero experiences correlated with parental human capital, influence children's skills (Cunha and Heckman (2007, 2009); Cunha et al. (2010); Heckman and Mosso (2014)). Following the literature, I assume a nested CES function for the skill formation technology, with one of the elasticities set to unity which imposes a parsimonious Cobb-Douglas outer form.[9] Following Kim et al. (2021), the param-

[8] By assuming uncorrelated ability shocks across generations, I abstract from modeling genetics. If one added it, incentives to invest would be distorted as high human capital parents would already expect their child to have high ability shocks. However, the main results of the paper would go through. In particular, the relative importance of h in the technology of skill formation captures part of a genetic effect.

[9] For similar modeling assumptions, see for instance Fuchs-Schündeln et al. (2022); Jang and Yum (2020); Lee and Seshadri (2019); Yum (2022).

eter $\underline{i} > 0$ guarantees that every child has a minimal level of baseline human capital. This term captures, for instance, a uniform minimum level of public education across neighborhoods. Finally, parental τ is normalized by its baseline unconditional mean ($\overline{\tau}$) to achieve unit independence and computational stability following Jang and Yum (2020).

College Graduation Probability: To connect the model outputs to the data, I introduce college graduation status. Each child has a positive probability of entering and graduating from college. The college graduation probability $g(h', h, s)$ depends on the parent's schooling status, earnings, and the child's accumulated human capital. Agent's schooling status s is equal to one if the agent graduated from college, zero otherwise.

Parents' Earnings: Parents' earnings are a function of accumulated human capital h, college graduation status s and exogenous working time $\overline{\ell_s} > 0$. Exogenous working time depends on the college graduation status s and embeds differences in working hours resulting from non-modeled labor market frictions. Parents' earnings are:

$$y = w\ h\ (1 + \omega s)\ \overline{\ell_s},$$

with w the rental rate of human capital, which is exogenous and common across households and neighborhoods. The college premium is captured by ω.

Parents' Preferences: Parents consume and enjoy leisure. In addition, they are altruistic as their child's value enters their utility function. Preferences of a parent raised in childhood neighborhood quality m_0, and with subjective beliefs $\tilde{\alpha}$ can be summarized by the following utility function:

$$\log(c) + \frac{\left(1 - \overline{\ell_s} - \kappa\ \tau\right)^{1-\frac{1}{\psi}}}{1 - \frac{1}{\psi}} + \ \iota\ \mathbb{1}_{m_0 = m} + \ \nu \varepsilon_m + \ b\mathrm{E}\left[\mathcal{V}\left(h', s', m, \tilde{\alpha}\right) | \tilde{\alpha}\right],$$

where b is a strength of altruism. Parents derive utility from consumption c and enjoy leisure, defined as one unit of time minus working time and parental time: $1 - \overline{\ell_s} - \kappa\ \tau$. Parental time τ is weighted by κ, showing that parents might value time spent at work differently than with their child. The curvature of the utility function with respect to leisure is $-\frac{1}{\psi}$. Parents have a preference for their childhood neighborhood quality that is captured by ι. In addition, parents have an idiosyncratic taste shock over neighborhoods ε_m. This shock captures moving motives that are uncorrelated with the neighborhood quality. I assume this shock is i.i.d. across neighborhood qualities and over time and distributed according to a Type-I Extreme Value with parameters $(-\bar{\gamma}, 1)$, where $\bar{\gamma}$ is the Euler-Mascheroni constant which ensures that the distribution has zero mean. The dispersion of the idiosyncratic neighborhood taste shock is measured by ν. Finally, $\tilde{\alpha}$ stands for the parent's subjective beliefs regarding the value of parameter α in the technology of skill formation and $\mathrm{E}\left[\mathcal{V}\left(h', s', m, \tilde{\alpha}\right)|\tilde{\alpha}\right]$ is the expected child lifetime utility with respect to the child's ability shock and neighborhood taste shock conditional on the parent's subjective beliefs $\tilde{\alpha}$.

Housing Supply: There are perfectly competitive land developers who produce housing on a unit endowment of land in each neighborhood with an isoelastic production function: $\mathcal{H}_m = \zeta r_m^{\eta}$, where ζ and η are parameters, with η the price elasticity of housing supply. Equilibrium rents per housing unit $\{r_m\}$ are determined by rental market clearing, such that housing demand equals housing supply in each neighborhood.

Aggregate Rent Rebates: Rents are redistributed to all families with a non-distortionary flat earnings subsidy. Every household receives a rebate from aggregate rent payments equal to R, where R is the economy's average rent payments.

1.2.2 Social Learning

One key and novel feature of the model is the social learning process. I depart from rational expectations and perfect information. I assume agents know everything about the model but the returns to parental inputs; specifically, they do not know the value of α, and they cannot perfectly infer it because, by assumption, they only imperfectly see abilities (a). In addition, I assume agents get information from their neighborhood only.

Young agents must form expectations about the returns to parental inputs ($\tilde{\alpha}'$) before making parental decisions. Following Fogli and Veldkamp (2011), they have two sources of information: they inherit their parent's subjective beliefs ($\tilde{\alpha}$) and make their own inference by observing outcomes and history in their neighborhood m ($\hat{\alpha}_m$).

Crucially, I depart from rational expectations by assuming agents suffer a cognitive selection neglect bias. Agents know the functional form of the technology of skill formation—but α,—the randomness of the ability shock and the parent's problem. Still, they do not fully understand the spatial sorting process in the economy. In particular, young agents draw conclusions from the observation of their neighbors without correcting for the fact that their adult neighbors are not representative of the population with respect to abilities.[10] The mental model young agents use for an adult j in a given neighborhood m is:

$$h_j = \left(i(\mathrm{T}_j, M_j) + \underline{i}\right)^{\alpha} \; H_j^{\beta} \; \exp(a_j), \quad a_j \sim \mathcal{N}(0, \sigma_a), \quad j \in m,$$

where T and M represent past parental choices, and H represents past parental human

[10] Enke (2020) provide empirical evidence of the selection neglect cognitive bias. People tend to draw general conclusions from what they observe, ignoring that what they observe is selected. Frick et al. (2022) develop a theoretical framework in which the selection neglect persists.

capital. However, with spatial sorting, in a given neighborhood m, the true model is $h_j = (i(\mathrm{T}_j, M_j) + \underline{i})^\alpha \ H_j^\beta \ \ \exp(a_j), \quad a_j \not\sim \mathcal{N}(0, \sigma_a), \quad j \in m$.

Young agents see average human capital and know about average past parental choices in their neighborhood only. Assuming agents would know all their neighbors' outcomes would be unrealistic, but by talking to neighbors and reading local news, I assume they have a good sense of local averages. In addition, they have information about average local abilities. Following Jehiel (2018) who develops a theoretical framework of selection neglect with bounded signals, I assume young agents receive informative but noisy and bounded signals about their adult neighbors' abilities.[11] The signal noisiness captures the fact that there is no real way to perfectly gauge ability—which can be interpreted as a combination of intelligence and luck—by simply observing people. The signal boundedness embeds the tendency to classify people's abilities within predefined categories.[12] Intelligence tests always have a scale with a predefined minimum and maximum level. The IQ test, for instance, classifies people's Intelligence Quotient between "*very superior*" and "*extremely low*." Note that the bounds of the signals are the same across neighborhoods. This common scale assumption is motivated by identical reference points regarding abilities. In every neighborhood, young agents interact with other young agents who are representative of the population with respect to abilities. They can all watch national media and gauge the speaker's ability. In a given neighborhood m, young agents' perception of the average local ability among adults is defined by:

$$\bar{\bar{a}}_m = \int_{-\infty}^{+\infty} \int_{-d}^{d} \tilde{a} \ f(\tilde{a}|a) \ l(a|m) \ d\tilde{a} \ da,$$

while the actual average local ability is:

$$\bar{a}_m = \int_{-\infty}^{+\infty} a \ l(a|m) \ da,$$

where conditional on the shock realization a, the signal realization $\tilde{a}$ is distributed according to the density $f(\cdot|a) = \mathcal{N}(a, \sigma_s)$ with full support in $[-d, d]$ with z, the bound, a real value. The precision of the signal is governed by the signal variance ${\sigma_s}^2$. The distribution of adults' ability in a given neighborhood m is denoted by $l(\cdot|m)$ and results from residential decisions in equilibrium. Note that, because of the bounds, despite the informativeness of the signal, the expected perceived ability is not always equal to the ability. The signal is upward biased

[11] Jehiel (2018) develops a theoretical framework of entrepreneurial decisions with bounded signals about the quality of businesses and selection based on success. He obtains over-optimism among entrepreneurs in equilibrium.

[12] The signal can be thought of as continuous or discrete. For computational reasons, in practice, every shock is discretized.

whenever the actual ability is below the average ability in the economy ($\forall\ a \leq 0$ then $\mathrm{E}\left[\tilde{a}|a\right] \geq a$), and downward biased whenever the actual ability is above the average ability in the economy ($\forall\ a \geq 0$ then $\mathrm{E}\left[\tilde{a}|a\right] \leq a$).[13] Intuitively, because the conditional signal is bounded, if the realization is closer to one of the bounds, many more high signals will be censored by this bound than by the other.[14] Consequently, in neighborhoods in which the average local ability is below the average ability in the economy, young agents overestimate average local ability (if $\overline{a}_m \leq 0$ then $\overline{a}_m \geq \overline{\tilde{a}}_m$), and conversely if the average local ability is above the average ability in the economy (if $\overline{a}_m \geq 0$ then $\overline{a}_m \leq \overline{\tilde{a}}_m$).[15]

Under the assumption of selection neglect, however, every agent in a given neighborhood m thinks of their perception of the average local ability as the truth ($\overline{\tilde{a}}_m = \overline{a}_m$).[16] She uses her observations of local averages to estimate the returns to parental inputs (α):[17]

$$\hat{\alpha}_m = \frac{\overline{\log(h)}_m \ - \ \beta\ \overline{\log H}_m - \overline{\tilde{a}}_m}{\overline{\log\left(i(\mathrm{T}, M) + \underline{i}\right)}_m}.$$

Young agents' estimate ($\hat{\alpha}_m$) is downward biased if they overestimate local abilities ($\overline{\tilde{a}}_m \geq \overline{a}_m$) and upward biased if they underestimate local abilities ($\overline{\tilde{a}}_m \leq \overline{a}_m$). Intuitively, when young agents overestimate abilities, they implicitly attribute too much of the local human capital to ability and too little to past parental inputs, underestimating the returns to parental inputs. Note that, because the signal variance ${\sigma_s}^2$ governs the precision of the perceived local ability, it also regulates the strength of the bias in the estimation.

Young agents then update their inherited subjective beliefs using a weighted average of the local estimate ($\hat{\alpha}_m$) and the inherited subjective beliefs ($\tilde{\alpha}$):

$$\tilde{\alpha}' = \mu\hat{\alpha}_m + (1-\mu)\tilde{\alpha}, \tag{1.2}$$

[13]See Appendix 1.C for a proof.

[14]See Appendix Figure 1.C.1 for an illustration.

[15]In the limit, if the signal variance (σ_s) tends to zero, or if the bounds (d) tend to infinity, agents' perception about their neighbors' ability shocks would always be equal to the truth and $\overline{\tilde{a}}_m = \overline{a}_m$. See Appendix 1.C for proof.

[16]This assumption would be correct if young agents were to observe everyone in the economy or without residential sorting ($l(\cdot|m) = \mathcal{N}(0, \sigma_a)$). Similar to Fogli and Veldkamp (2011), there would be convergence in subjective beliefs. Suppose young agents did not suffer from selection neglect and understood the spatial sorting process. Only if young agents knew the joint distribution between subjective beliefs, human capital, college graduation status, and place of birth could they compute the distribution of adults' ability in their neighborhood $l(\cdot|m)$ and then back out the actual average ability in their neighborhood.

[17]An alternative would be that agents observe individual outcomes of a given number of neighbors and run an OLS regression. To make this assumption realistic, one would have to draw, randomly or not, the number of neighbors each agent observes. This deviation would make the model richer but wouldn't change the paper's main results. In the baseline version, you agents in neighborhood m correctly see $\overline{\log(h)}_m$, $\overline{\log H}_m$ and $\overline{\log\left(i(\mathrm{T}, M) + \underline{i}\right)}_m$.

with $\mu \in (0, 1)$.

In equilibrium, this social learning process generates persistent delusion about the value of α across socioeconomic groups and within families. Children of poor and low-subjective beliefs parents are raised in low-quality neighborhoods, observe on average low-ability neighbors, overestimate local abilities, and are comforted in their inherited low subjective beliefs about the returns to parental inputs. Those children will likely become poor and low-subjective beliefs parents the next period. The opposite happens for children of high-income and high-subjective beliefs parents.

1.2.3 Parents' Problem

Parents are the only decision-makers in the economy. They make three decisions of which two that affect their child's next period human capital: neighborhood quality m and parental time τ. The timing is as follows: parents observe their idiosyncratic taste shocks for neighborhoods, take rents as given, and make decisions by maximizing their utility conditional on their subjective belief about the returns to parental inputs ($\tilde{\alpha}$). The maximization problem is the following:

$$\begin{aligned} \mathcal{V}\left(h, s, m_0, \tilde{\alpha}\right) = \max_{c,\tau,m} \{ \log\left(c\right) + \frac{\left(1 - \overline{\ell_s} - \kappa\, \tau\right)^{1-\frac{1}{\psi}}}{1 - \frac{1}{\psi}} \\ + \ \iota\, \mathbb{1}_{m_0 = m} + \ \nu \varepsilon_m + \ b \mathrm{E}\left[\mathcal{V}\left(h', s', m, \tilde{\alpha}\right)|\tilde{\alpha}\right] \} \end{aligned} \tag{1.3}$$

subject to:

$$\begin{aligned} & c + r_m = w\ h\ (1 + \omega s)\ \overline{\ell_s}\ + R \\ & \tau \in [0, 1 - \overline{\ell_s}] \\ & h' = f(\tau,\ m,\ h,\ a'|\tilde{\alpha}), \quad a' \sim \mathcal{N}(0, \sigma_a) \\ & p(s' = 1) = g(h', h, s), \end{aligned}$$

where r_m is the equilibrium rent of neighborhood m, $\mathrm{E}\left[\mathcal{V}\left(h', s', m, \tilde{\alpha}\right)|\tilde{\alpha}\right]$ is the expected child's utility conditional on the ability and the neighborhood taste shocks, and $f(.)$ is the technology of skill formation defined by (1.1). Parents decide how to allocate their income into consumption c and housing costs r_m, and one unit of time into leisure, exogenous working hours, and parental time.

1.2.4 Housing Market

Let $U = E(\mathcal{V})$ denote the expected lifetime utility of a representative parent with respect to the vector of idiosyncratic neighborhood taste shocks ε_m. Let $V\left(h, s, m_0, \tilde{\alpha}, m|r_m\right) = \log\left(c_m^*\right) + \frac{\left(1-\overline{\ell_s}-\kappa\,\tau_m^*\right)^{1-\frac{1}{\psi}}}{1-\frac{1}{\psi}} + \iota\, \mathbb{1}_{m_0=m} + b\mathrm{E}\left[\mathcal{V}\left(h', s', m, \tilde{\alpha}\right)|\tilde{\alpha}\right]$ the utility derived from living in neighborhood m abstracting from the neighborhood taste shock, with c_m^* and τ_m^* the optimal parent's choices given the neighborhood m and the rent price r_m.

Under the Type-I Extreme Value assumption, the expected lifetime utility of a parent is:

$$U\left(h, s, m_0, \tilde{\alpha}\right) = \nu \log \sum_m \exp\{\frac{1}{\nu}\, V\left(h, s, m_0, \tilde{\alpha}, m|r_m\right)\}.$$

The share of parents who choose to locate in neighborhood quality m among parents with human capital h, graduation status s, raised in neighborhood quality m_0 and with subjective beliefs $\tilde{\alpha}$ is :

$$\lambda_m(h, s, m_0, \tilde{\alpha}|r_m) = \frac{\exp\{\frac{1}{\nu}\, V\left(h, s, m_0, \tilde{\alpha}, m|r_m\right)\}}{\sum_n \exp\{\frac{1}{\nu}\, V\left(h, s, m_0, \tilde{\alpha}, n|r_n\right)\}}.$$

In equilibrium, rent prices are such that housing demand equals housing supply in each neighborhood m:

$$\sum_{m_0}\sum_s \int\int \lambda_m(h, s, m_0, \tilde{\alpha}|r_m)F(h, s, m_0, \tilde{\alpha})\; dh\; d\tilde{\alpha} = \zeta {r_m}^{\eta}, \tag{1.4}$$

with $F(h, s, m_0, \tilde{\alpha})$ the joint distribution of human capital, graduation status, neighborhood of birth, and subjective beliefs

1.2.5 Equilibrium

The key objects of the equilibrium in the steady state are the endogenous distribution of human capital, subjective beliefs, and rent prices. For a given initial human capital, graduation status, neighborhood of birth, and subjective belief distribution $F_0(h, s, m_0, \tilde{\alpha})$, an equilibrium is characterized by a sequence of residential and parental time choices, $\{m\}$ and $\{\tau\}$, a sequence of rents $\{r_m\}$ for each neighborhood, and a sequence of distributions $\{F(h, s, m_0, \tilde{\alpha})\}$, such that the following four conditions are satisfied:

1. agents solve (1.3).

2. housing market clearing: rent prices $\{r_m\}$ ensure housing demand equals supply in every neighborhood according to (1.4).

3. beliefs update: young agents update their inherited beliefs according to (1.2).

4. earnings, graduation status, place of birth, and subjective beliefs consistency: those are consistent with the parent's income, graduation status, subjective beliefs, and decisions.

Details on how I compute the steady state equilibrium are provided in Appendix 1.B.

1.3 Calibration

I numerically solve the model as detailed in Appendix 1.B and calibrate it to the average United States commuting zone in the 2000s. The calibration proceeds in multiple steps. First, some parameter values are directly estimated from the data. Second, others are chosen externally based on direct data analogs, the literature, or simple normalization, and finally, the remaining parameters are selected to match relevant data moments. To validate the model, I then compare the model's predictions to non-targeted moments.

1.3.1 Preliminaries

I let the discrete distribution for a to approximate a normal distribution $a \sim \mathcal{N}(0, \sigma_a)$ which I discretize using Tauchen (1986), with a ten-point grid. As standard in the literature, I set the grid bound to 2.5 times the standard deviation ($z = 2.5\ \sigma_a$). Finally, I assume the technology of skill formation has constant returns to scale, $\beta = 1 - \alpha$.

I use multiple data sources to compute the relevant moments to calibrate the model. Earnings distribution moments come from the American Community Surveys (ACS) from 2000. I use the information at the census tract level from the National Historical Geographic Information System (NHGIS) data set in 2000 (Manson et al., 2022) to construct neighborhood sociodemographic characteristics. I aggregate census tract information at the commuting zone level using a county-to-commuting zone crosswalk. Neighborhood choices and college graduation probabilities are estimated using data from the AddHealth survey described in Section 1.5.1.[18] Finally, parents' time use information is taken from the American Time Use Survey (ATUS) from 2003.[19]

[18]See Appendix Section 1.D.3 for more details on those statistics.

[19]Appendix Section 1.A.3 provides detailed information on how I compute parental time using ATUS.

Geography

The model is calibrated to the average commuting zone in the United States, matching neighborhoods' impact estimated by Chetty and Hendren (2018a) in the 100 biggest commuting zones. To be consistent, I only use the top 100 commuting zones in NHGIS 2000. In the model, I set the number of neighborhoods to ten, representing a synthetic decile census tract in the data. Specifically, in each of the 100 commuting zones of the NHGIS 2000, I sort census tracts by median household income and form ten synthetic neighborhoods from the deciles of this distribution.[20] I then create one average commuting zone by computing the population-weighted average characteristics of each of the ten neighborhoods.[21]

1.3.2 Empirically Estimated Parameters

Housing Markets: Rental prices are determined in equilibrium given the supply function: $\mathcal{H}_m = \zeta r_m^\eta$, where r_m is the equilibrium rent price in the neighborhood m, and η is the price elasticity of housing supply. η and ζ can be estimated directly from the synthetic neighborhood density and rents.[22,23]

College graduation probability: The college graduation probability depends on the parent's education and earnings and the child's accumulated human capital. Following Blandin and Herrington (2022) and using AddHealth, I estimate the following weighted logit regression of college completion:

$$g(h', y, s) = \frac{1}{1 + \exp(-(\gamma_1 + \gamma_2 rank_{h'} + \gamma_3 rank_y + \gamma_4 s))}$$

where $g(h', y, s)$ is the binary outcome of either graduating college or not in Wave V, $s = 1$ if the highest level of education of the mother is above or equal to a bachelor's degree, $rank_{h'} \in \{1, 2, ..., 10\}$ is the human capital rank proxied by the grade rank in Wave II of the interviewed adolescent in 1996, and $rank_y \in \{1, 2, ..., 10\}$ is the household income rank in 1994-1995.[24]

[20] I use median household income as proxy for neighborhood quality because it is one of the variables that correlates the most with place-based effects within commuting zones of Chetty and Hendren (2018a). Using their estimates, I find that growing up in one standard deviation higher median income county within a commuting zone increases a given child's income by 1.9%.

[21] Appendix Table 1.D.2 summarizes the ten neighborhoods' characteristics.

[22] Appendix Figure 1.D.1 summarizes the log-relationship between density (Column (5) Table 1.D.2) and rents across the ten synthetic neighborhoods.

[23] Note that in the literature, ζ is sometimes neighborhood specific (ζ_m). In this context, there is an almost linear log relationship between density and rent prices (see Appendix Figure 1.D.1); I choose to have the same ζ value across neighborhoods. Without loss of generality, the numeraire is the average household earnings in the economy.

[24] Appendix Section 1.D.3 provides details on the variable construction, and Appendix Table 1.D.3 shows the

1.3.3 Externally Calibrated Parameters

The parameter $\frac{1}{\psi}$ governs the curvature of the utility function with respect to leisure. I set the intertemporal elasticity of substitution ψ to 0.5, as is standard in the literature. I assume agents have an endowment of one unit of time corresponding to sixteen hours per day in the data.[25] The fraction of time allocated to market work by education comes from ATUS 2003, as described in Appendix Section 1.A.3. The wage rate w, the parental investment constant $\underline{i}$, and the average neighborhood quality is normalized to 1. I assume the number of neighborhoods N is equal to ten, and neighborhood quality is distributed according to $m \sim U(\underline{m}, \overline{m})$. Table 1.1 summarizes the parameters that are externally calibrated.

Table 1.1: Externally Calibrated Parameters

Parameter	Description	Value	Source
N	Number of neighborhoods	10	Deciles NHGIS
β	Returns on parental human capital	$1-\alpha$	Constant returns
$\overline{\ell_0}$	Non-college labor supply	0.275	ATUS 2003
$\overline{\ell_1}$	College labor supply	0.294	ATUS 2003
ψ	Intertemporal elasticity of substitution	0.5	Standard
w	Wage rate	1	Normalization
$\underline{\mathrm{i}}$	Parental investment constant	1	Normalization
μ_m	Average neighborhood quality	1	Normalization
ζ	Housing supply coefficient	-1.04	NHGIS
η	Price elasticity of housing supply	0.58	NHGIS
γ_1	College graduation probability coefficient	-3.83	Add Health
γ_2	College graduation probability coefficient	0.35	Add Health
γ_3	College graduation probability coefficient	0.15	Add Health
γ_4	College graduation probability coefficient	1.11	Add Health

Notes: The table shows all the externally calibrated parameters.

1.3.4 Internally Calibrated Parameters

The remaining twelve parameters to calibrate are listed in Table 1.2. I calibrate them by minimizing the sum of squared percentage differences between data and model moments. The data moments include two measures of household earnings dispersion within commuting zones computed from the ACS 2000: the Gini coefficient of household earnings and the income ratio of non-college and college households. Both are weighted population averages across the

weighted logit regression estimates.

[25] I remove eight hours of sleep needs, a standard assumption in the literature.

100 biggest commuting zones.[26] In addition, I ensure income and college graduation status relationship matches the data by incorporating the ratio of college-parents share in the first to the fourth quartile of the income distribution. I further include the rank-rank coefficient from Chetty et al. (2014), a coefficient that captures the income correlation between parents and children. It is an inverse measure of social mobility. To discipline the neighborhood quality distribution ($\{m\}$) that directly enters the child skill production function, I use the causal effect of a one standard deviation improvement in neighborhood quality for a child born in the 25th and 75th percentile of the household income distribution estimated by Chetty and Hendren (2018b). To match residential segregation, I add a Gini coefficient across the ten neighborhoods computed from the NHGIS 2000 data set. The place of birth preference parameter is calibrated by matching a moment labeled "residential mobility (D1)". It is defined as the fraction of children born in the first synthetic neighborhood who choose to live in the same synthetic neighborhood in adulthood. I include parental time by education (displayed in Appendix Table 1.D.1) and household income and education gradients to capture parental behaviors. In addition, to discipline subjective beliefs, I add the correlation between parents' and children's neighborhood choices.[27]

Table 1.2 reports calibrated parameters, corresponding moments in the data, and their model analogs. Even though every moment results from the combination of all parameters, certain moments are more sensitive to specific parameters. Understanding these intuitive links is informative about the underlying model mechanisms.

The first three parameters are preference parameters and govern parents' choices. In particular, childcare disutility weight κ is pinned down by the fraction of time allocated to childcare by non-college parents, and the preference for place of birth ι is pinned down by residential mobility in the first decile neighborhood. The college wage premium ω directly affects the earnings gap between college and non-college parents.

Neighborhood parameters govern the model geography. In particular, the standard deviation of neighborhood quality σ_m affects how much a child's earnings are affected by neighborhood choices. The relevant moment is the causal effect of neighborhoods measured by Chetty and Hendren (2018b). It determines how much would increase children's future income rank if they had been growing up in one standard deviation better neighborhoods. For a child born with a parent at the 25th percentile of the income distribution, the authors find a value of 6.2% of income at the county level within commuting zones. The taste shock vari-

[26] The Gini coefficient is lower than typically reported because it is a Gini over twenty years of household income. I compute a yearly Gini coefficient from the data and transform it into a twenty years Gini coefficient using Shorrocks mobility index estimated by Kopczuk et al. (2010) for 2002.

[27] Moments construction and data sources are detailed in Appendix Section 1.D.4.

Table 1.2: Internally Calibrated Parameters

Parameter	Description	Moment	Data	Model
Preferences and Labor Market				
$b = 0.5$	Altruism	Ratio share college parents Q1 to Q4	0.102	0.120
$\kappa = 0.6$	Parental time disutility	Parental time non-college parents	0.075	0.079
$\iota = 0.0001$	Place of birth preference	Residential mobility (D1)	0.302	0.274
$\omega = 0.005$	College wage premium	Earnings ratio non-college - college	0.554	0.556
Neighborhoods				
$\sigma_m = 0.26$	Neighborhood quality	Neighborhood effect (25th pct.)	0.062	0.056
$\nu = 0.01$	Taste shock variance	Census tract Gini	0.231	0.212
Skill Formation: $h' = \left(\left(\gamma \ \left(\frac{\tau}{\overline{T}}\right)^{\varphi} + (1-\gamma)\, m^{\varphi}\right)^{\frac{1}{\varphi}} + \underline{i}\right)^{\alpha} h^{1-\alpha} \ \exp(a')$ with $a' \sim \mathcal{N}(0, \sigma_a)$				
$\alpha = 0.77$	Elasticity of investments	Rank-rank coefficient	0.341	0.339
$\gamma = 0.55$	Parental time share	Neighborhood effect (75th pct.)	0.046	0.047
$\varphi = 0.5$	Substitutability	Income gradient in parental time	0.140	0.144
$\sigma_a = 0.55$	Ability shock variance	Income Gini	0.336	0.334
Belief Updating Process				
$\sigma_s = 2.5$	Cognitive bias strength	Neighborhood quality correlation	0.417	0.460
$\mu = 0.3$	Update weight	Education parental time gap	0.750	0.850

Notes: This table reports the internally calibrated parameters and the observed and simulated moments associated with the parameter estimates.

ance ν controls residential moves orthogonal to neighborhood quality and affects residential segregation measured by the Gini coefficient across neighborhoods.

The child's skill formation parameters are most relevant for this paper. In particular, the substitutability parameter φ is calibrated by matching the childcare time difference across income groups; here, I use the regression coefficient of parental time on household income quartiles.[28] With subjective beliefs and social learning, I find that neighborhood quality and parental time are substitute inputs in the child skill production function $\varphi > 0$ (in line with Agostinelli (2018) and Agostinelli et al. (2022)). Parental human capital share $1 - \alpha$ mechanically increases the income correlation between parents and children. Thus, the relevant moment is the rank-rank coefficient between parental and child earnings estimated by Chetty et al. (2014). As ability shock variance captures any income variation not explained by parental choices and human capital, it is calibrated by matching household earnings inequality.

Finally, the correlation between parental and children's subjective beliefs is governed by μ and affects the persistence of beliefs and hence, parental behavior within families. Since there is also persistence in earnings and schooling status within families, I discipline this parameter by matching parental time by schooling status. Specifically, I compute the parental time ratio by the college status of the parents. It is below one, meaning college parents spend more time with their children than non-college parents. This moment is labeled the education parental time gap. In the model, the parental time gradient in education result from two opposing forces. Delusion about the technology of skill formation, if correlated with earnings, increases the gap. However, the substitutability between parental time and neighborhood quality ($\varphi > 0$) and the differential in working hours decrease it. Finally, I calibrate the signal variance (σ_s) using the correlation between parents' and children's neighborhood choices. Parents' neighborhood choices affect children's future decisions through two channels: human capital formation and hence earnings and subjective beliefs. Parents influence on children's earnings is calibrated by matching the rank-rank coefficient and places effects. It is then essential to match this correlation to ensure that the subjective beliefs channel is not too strong.

1.3.5 Non-Targeted Moments

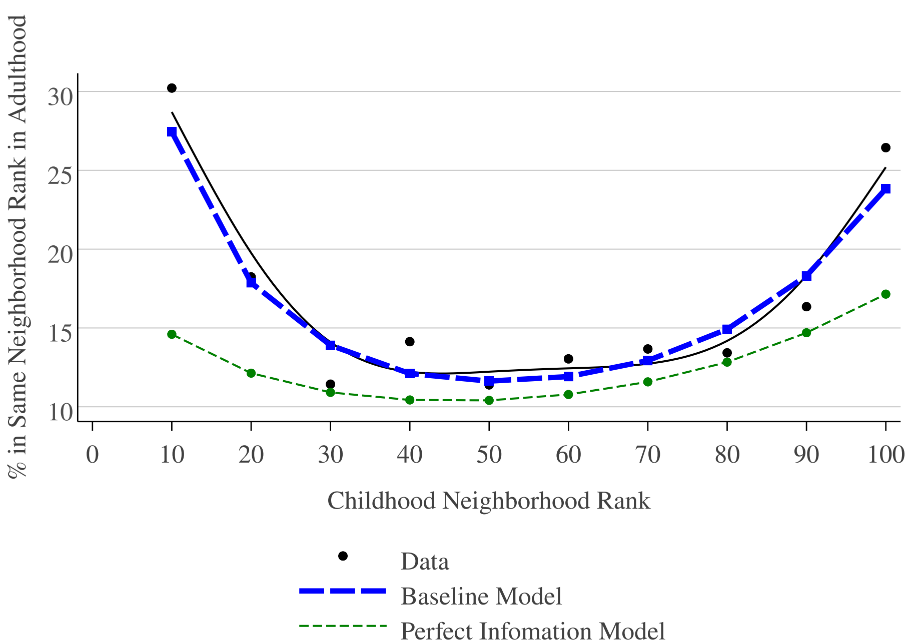
The calibrated model matches the targeted moments well and fits the non-targeted residential mobility patterns.[29] Figure 1.2 reproduces the motivating Figure 1.1 and presents the

[28] In the data, I control for the gender of the respondent and the age of the child.

[29] Appendix Section 1.D.5 shows non-targeted moments at the neighborhood level and income quintile matrix.

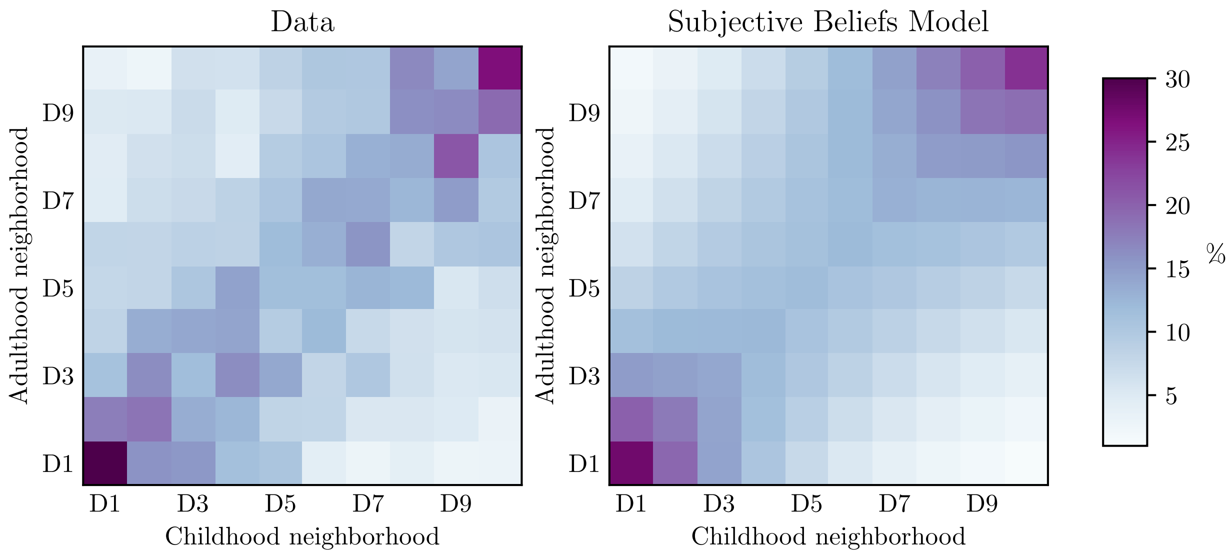
share of children who live in the same neighborhood in adulthood. The solid line is a smooth interpolation between the dots which are data moments. The blue dashed line represents model simulated analogs. While the first synthetic neighborhood statistic is a targeted moment, the others are not. The model generates a U-shape that is very close to the data. To go further, Figure 1.3 illustrates a frequency matrix of all possible intergenerational moves, each represented by a colored square. The darker a square is, the more likely a given move. For instance, a child born in neighborhood one is likely to live in neighborhood one or two when she becomes an adult, but she is very unlikely to live in neighborhood six or above. The calibrated model (right panel) matches the data patterns (left panel) remarkably well.

Figure 1.2: Non-targeted Moments: Residential Mobility



Notes: This Figure shows the share of children who, in adulthood, choose to live in the same decile of the neighborhood quality distribution as their parents. The dots and solid line are data moments; the baseline model simulated analogs in blue long-dashed-line, and the perfect information model in green dashed-line —data source: Add Health, see Appendix for details of data construction.

Figure 1.3: Non-targeted Moments: Detailed Residential Mobility



Notes: This Figure shows residential mobility between childhood and adulthood in the data (left panel), and their model simulates analogs (right panel). The x-axis contains the ten synthetic childhood neighborhoods, and the y-axis is the ten synthetic adulthood neighborhoods. The frequency of moves is represented by cell colors, with larger values in darker squares—data source: Add Health, see Appendix 1.D.4 for details of data construction.

1.3.6 The Role of Subjective Beliefs

The model is calibrated under the assumption of imperfect information and social learning at the neighborhood level. Natural questions are: What does the endogenous distribution of subjective beliefs do? And how well would the model match the moment under perfect information?

I first shut down the subjective beliefs channel to understand the role that subjective beliefs play in the economy by assuming parents know the returns to parental input (α). Table 1.3 presents the effects of subjective beliefs on the economy. In the bottom quartile of the income distribution, parents underestimate the returns to neighborhood quality and parental time by 17% while parents in the top quartile of the income distribution overestimate them by 7%. These numbers are of a reasonable order of magnitude. Cunha et al. (2013), who elicits disadvantaged African American mothers' subjective beliefs about the elasticity of child development with respect to parental investments, finds greater differences between the truth and their subjective beliefs. As a result, low-income parents spend too little time with their children, while high-income parents spend too much time. Providing information to parents would, in the long run, increase low-income households' parental time by 31% and decrease high-income households' one by 5%. Under perfect information, aggregate parental time increases by 7%. Parents' subjective beliefs decrease social mobility—negative rank-rank coefficient—and earnings by 12% and 3% respectively and increase inequality and poverty by 3% and 17%.[30] The Green dashed line in Figure 1.2 shows that subjective beliefs double the share of children born in bottom-quality neighborhoods who choose to remain in adulthood. These findings imply that a relatively modest level of delusion, coherent with micro-studies, has large effects on the economy.

Table 1.3: Effects of Subjective Beliefs

		Household Income quartile			
	All	1st	2nd	3rd	4th
Subjective Beliefs	-2.8%	-17.5%	-5.9%	+0.2%	+7.4%
Parental Time	-7.5%	-30.9%	-9.9%	-2.5%	+4.8%
Earnings	-2.5%	-6.4%	-5.8%	-4.7%	-2.2%
Rank-rank coefficient	+11.8%				
Census tract Gini (Segregation)	+0.6%				
Income Gini (Inequality)	+2.7%				
Absolute poverty	+16.9%				

Notes: This table displays percentage differences in model generated moments between the baseline calibrated model with and without subjective beliefs.

How would a model do without subjective beliefs? I calibrate the same model shutting down subjective beliefs to see how well this alternative version of the model matches targeted and non-targeted moments. Appendix Table 1.D.5 shows the fit of the perfect information model version. As expected, to generate a positive correlation between income and parental time, time and neighborhood quality are complement inputs in the technology of skill formation ($\varphi < 0$). In addition, to have 30% of children born in the bottom-quality neighborhood who stay there in adulthood, the preference parameter ι needs to be much larger than in the subjective beliefs version of the model. However, this feature reduces mobility in all other neighborhoods.[31] Under perfect information, the model, misses all the intergenera-

[30] Poverty is measured by absolute level of poverty. The poverty threshold is defined at baseline by the tenth percentile of the household income distribution.

[31] Appendix Figures 1.D.3 and 1.D.4 show non-targeted simulated moments. In the mobility matrix under perfect information, not only is the diagonal off but all the squares off-diagonal are of similar color, which is at odds with the data.

tional residential mobility moments despite a good match of social mobility and inequality moments.

Since, the model fails to match important non-targeted moments without subjective beliefs, I augment it with heterogeneous preferences and moving costs. I now assume a quadratic moving cost function and preferences for childhood neighborhoods that vary by place of birth to capture the mobility matrix. Parents' preferences take the following forms:

$$\log(c) + \frac{\left(1 - \overline{\ell_s} - \kappa\, \tau\right)^{1-\frac{1}{\psi}}}{1 - \frac{1}{\psi}} + \iota_{m_0}\, \mathbb{1}_{m_0 = m} - \xi (m_0 - m)^2 + \nu \varepsilon_m + b\mathrm{E}\left[\mathcal{V}\left(h', s', m, \alpha\right)\right]$$

Appendix Table 1.D.6 shows the fit of the perfect information model version augmented with preference heterogeneity. To reproduce the U-shape patterns in residential mobility, the preference for place of birth needs to be about thirty times higher in the bottom-quality neighborhood than in the middle-quality ones. This feature is at odds with Bergman et al. (2019)'s empirical findings. The authors compare low-income families randomly allocated between treatment and control groups. Parents in the treatment group are induced to move to higher-quality neighborhoods. Those parents are more likely to move and to be satisfied and willing to stay in their neighborhood than those in the control group. The model with perfect information and heterogeneous preferences across places of birth fits residential mobility patterns by construction but misses parental time patterns across socioeconomic groups.[32] The data shows a steep parental time gradient in education. The model does not capture the income gradient well–it is 0.09 instead of 0.14 despite $\varphi = -5$ being very negative–and generates a too small education parental time gap –0.93 instead of 0.75. There is intuitive reasoning behind this result. Parental time and neighborhood quality are complements inputs of the technology of skill formation ($\varphi < 0$), which, combined with income segregation, generates a strong positive correlation between parental time and income and a smaller one with education. However, college parents work more hours than non-college parents. This feature decreases the correlation between parental time and education. College parents are also better at teaching their children than non-college parents. This feature has an ambiguous effect on the correlation between parental time and education. Overall, parental time is weakly correlated with education. Additional sources of heterogeneity are needed to fully match the parental behavior and understand why college parents spend more time with their children than non-college parents despite working more hours.[33]

[32] Appendix Figures 1.D.5 and 1.D.6 show the fit of residential mobility moments by this alternative version of the model.

[33] Differences by education in the altruism parameter b are not sufficient to fit the data.

While preference heterogeneity can help match the data, its origin is difficult to justify. Do parental time preferences systematically differ by education status? How does it transmit to children? Is a quadratic moving costs function credible? Why would children born in the bottom neighborhood be so much more attached to their neighborhood quality given all the negative features it has: high-crime rates, high-poverty rates, a low opportunity for children?[34] Discrimination or a homophily bias could motivate some of these modeling assumptions. However, in Table 1.9, I find that race is not the primary driver of intergenerational residential mobility once controlled for childhood neighborhood quality. In addition, Bergman et al. (2019) find higher satisfaction levels of low-income families who moved to higher-quality neighborhoods which suggests that if they face discrimination once installed, it does not make them want to move back to low-quality neighborhoods.

1.4 Housing Voucher Policies

The model displays two main frictions that motivate government involvement: parents cannot borrow against their children's future earnings and information frictions that result from segregation. Both lead to lower levels of parental inputs, in particular lower neighborhood quality, in low-income families compared to a perfect information world in which children could control inputs into their development.

In this section, I use the quantitative model to study the effects of housing vouchers. The Housing Choice Voucher program is the U.S. Department of Housing and Urban Development largest housing assistance program and its primary mechanism for promoting the mobility of low-income families.[35] The model provides a new rationale for this policy. In addition to addressing redistribution concerns, this policy can improve information about the child skill technology in the economy by reducing segregation.

In the model, housing vouchers are financed through property taxes, which adds two terms to the household budget constraint:

$$c + r^{HV}_{m,h,s}\,(1+\tau_r) = w\;h\;(1+\omega s)\;\overline{\ell_s}\; + R,$$

where τ_r is the tax rate and $r^{hv}_{m,h,s}$, is the rent faced by households once the housing voucher policy is implemented. The government budget constraint is balanced such that:

[34] Commuting costs are not explicitly modeled but captured by the idiosyncratic neighborhood taste shocks. However, they are unlikely to explain residential mobility patterns across neighborhood quality. Indeed Chetty and Hendren (2018a) find no correlation between short commute time and place effects on children's future earnings.

[35] https://www.huduser.gov/quarterly_update/update_June2015.html

$$\sum_m \sum_s \int (r_m - r^{HV}_{m,h,s}) F(h,s,m)\; dh = \sum_m \sum_s \int (r^{HV}_{m,h,s}\; \tau_r) F(h,s,m)\; dh.$$

1.4.1 Housing Voucher Policy

The Housing Choice Voucher program rule imposes that seventy-five percent of families who receive housing vouchers each year have "extremely low incomes," defined as incomes up to the poverty line. The others' income may not exceed 50% of the median income for the metropolitan area where the family chooses to live. The voucher generally covers the difference between 30 percent of the family income and the rent, up to a limit based on Housing and Urban Development's fair market rent estimates at the metropolitan area level.

I consider a housing voucher policy closely designed as the Housing Choice Voucher program. Housing vouchers are offered to young parents before they make their residential choice. The vouchers cover the difference between 30% of the family's income and the rent up to the rent ceiling, the 40th percentile rent in the commuting zone. Eligible households are those below the poverty threshold, defined as the income level at the tenth percentile of the income distribution. Let r_{m40} the 40th percentile rent in the commuting zone, then the rent price in the neighborhood m for a parent of income $y(h,s)$ who is eligible for the housing voucher is:

$$r^{HV}_{m,h,s} = \min(0.3 \times y(h,s),\; r_m) + \max(r_m - r_{m40},\; 0).$$

As a first step, I conduct a field experiment within the model to investigate the impact of housing vouchers and compare them to the empirical estimates by Chetty et al. (2016). These are partial equilibrium results as too few people are treated to change equilibrium forces. Then, I scale up the policy without changing the eligibility criterion and consider steady-state comparisons, which helps gauge the long-run implications.

Randomized Control Trial Within Model

Column 4 Table 1.4 shows the positive effects of housing vouchers on eligible households. For treated families, the housing voucher policy improves neighborhood rank by 1.5 points, improving children's earnings at age 26 by \$663. The predicted effect on children's earnings falls in the lower bound of Chetty et al. (2016)'s empirical estimate. This is most likely because treated individuals are poorer in the data than in the model. Indeed, Chetty et al. (2016) evaluate housing vouchers' effects on low-income households who already live in public housing in deprived neighborhoods. In the model, though, the voucher is offered to young parents with an income below the poverty rate, independently of their neighborhood choice.

Only 40% of them choose to live in the bottom decile neighborhood. In the data, the average family earnings at age 26 of children in the control group is $12,702, while in the model, it is $20,917. Nevertheless, the predicted effect on children's earnings falls within Chetty et al. (2016)'s estimated confidence interval ($1,452 with a standard error of 736). The policy induces an extra 62% of families to move out of bottom-quality neighborhoods in the data and the model.

Incorrect parental subjective beliefs are part of why not all low-income families move to high-quality neighborhoods. In Column 5 Table 1.4, I conduct the same experiment but provide information about the returns to neighborhood quality to treated families. The neighborhood rank of treated families improves by 1.9 points, 0.4 points higher than with only housing vouchers. As a result, perfect information further increases treated children's adulthood earnings by $132.

Table 1.4: Effects of a Housing Voucher Policy on Eligible Households

	Data		Model		
	Control	Housing Vouchers	Control	Housing Vouchers	Housing Vouchers + Info
	(1)	(2)	(3)	(4)	(5)
% in bottom-quality neighborhood	100%	[-62%; -70%]	40%	-62%	-76%
Children's future earnings	$12,702	[+$11;+$2,893]	$20,917	+$663	+$795
Neighborhood rank			2.14	+1.5	+1.9
Parental time (min./day)			57	+1	+1

Notes: This table shows the effects of housing vouchers on eligible families, from the data, and simulated by the calibrated baseline model. Data source: Chetty et al. (2016).

Scaling-up Housing Vouchers

In the next step, I scale up the policy to all families below the poverty threshold in the economy and compute the new steady state. The steady-state comparisons provide insights into the long-run implications of the policy.

General equilibrium responses in rents and subjective beliefs amplify housing voucher effects on eligible households. While Column 1 of Table 1.5 repeats the partial equilibrium effects of the housing voucher policy on eligible families, Column 2 shows the long-run effects of the same policy that is scaled up. Compared to partial equilibrium, long-run effects on eligible households' parental time and neighborhood rank are more prominent (+1.6 versus +1.5 and +9 versus +1, respectively), further increasing eligible children's earnings by +$1,008 per year ($1,671 versus $663).

In the long run, the housing voucher policy has overall positive effects. Column 3 of Table 1.5 shows the policy effects on all households. With housing vouchers, earnings at age 26 increase by $277, and inequality, poverty, and the rank-rank coefficient decrease. This suggests that housing vouchers benefit eligible households and can help combat segregation's adverse effects on inequality and intergenerational mobility.

Table 1.5: The Effects of Scaling-up Housing Vouchers

	Small Field	Large Scale and Long Run			
	Baseline model	Baseline model		Baseline model without change in social learning	
Households	Eligible (1)	Eligible (2)	All (3)	Eligible (4)	All (5)
% in neighborhood D1	-62%	-58%	-3%	-69%	-1%
Children's future earnings	+$663	+$1,671	+$277	+$870	-$57
Neighborhood rank	+1.5	+1.6	+0.0	+1.9	+0.0
Parental time (min./day)	+1	+9	+2	+0	-1
Inequality			-0.8%		+0.0%
Poverty			-6.3%		+1.3%
Earnings			+0.9%		-0.2%
Rank-rank coefficient			-3.8%		+0.5%

Notes: This table shows the effects of scaling-up housing vouchers within the calibrated baseline model.

I fix information frictions at its baseline level to understand whether the long-run effects of housing vouchers are due to the housing market response or a change in information frictions. The policy effects presented in Columns 4 and 5 of Table 1.5 can be thought of as short- or medium-run policy effects, as only the housing market response is considered. Partial equilibrium effects on eligible children's earnings are amplified, but only by $207 per year ($870 versus $663). This moderate amplification effect is driven by an increase in neighborhood rank due to the voucher and the housing market response but an absence of improvement in parental time. This suggests that most—about 80%—of the amplification effect on eligible households results from a change in information frictions.

Overall, ignoring information frictions leads to different conclusions concerning the policy desirability (Columns 5 of Table 1.5). If only the housing market reacts, inequality and poverty increase while earnings and social mobility decrease. Figure 1.4 and Table 1.6 help understand subjective beliefs' role in the long-run policy effects. In the long run, eligible households move to higher-rank neighborhoods, improving parents' subjective beliefs, particularly those of low-income parents. This leads to a change in parents' behavior. Low-income parents spend more time with their children despite an increase in neighborhood quality (remember that neighborhood quality and parental time are substitute inputs in the technology of skill formation). The change in subjective beliefs decreases the share of parents in the bottom-quality neighborhood and improves the share of parents in higher-quality neighborhoods (Figure 1.4). As a result, social mobility improves, and inequality decreases. However, all those positive effects are absent when ignoring the change in information frictions.

Table 1.6: The Effects of Scaling-up Housing Vouchers by Income Quartile

		Household Income quartile			
	All	1st	2nd	3rd	4th
Subjective Beliefs	+2.8%	+7.1%	+3.3%	+2.2%	+0.2%
Parental Time	+2.6%	+8.0%	+2.2%	+1.2%	+1.3%
Earnings	+0.9%	+0.7%	+2.3%	+3.5%	+2.3%

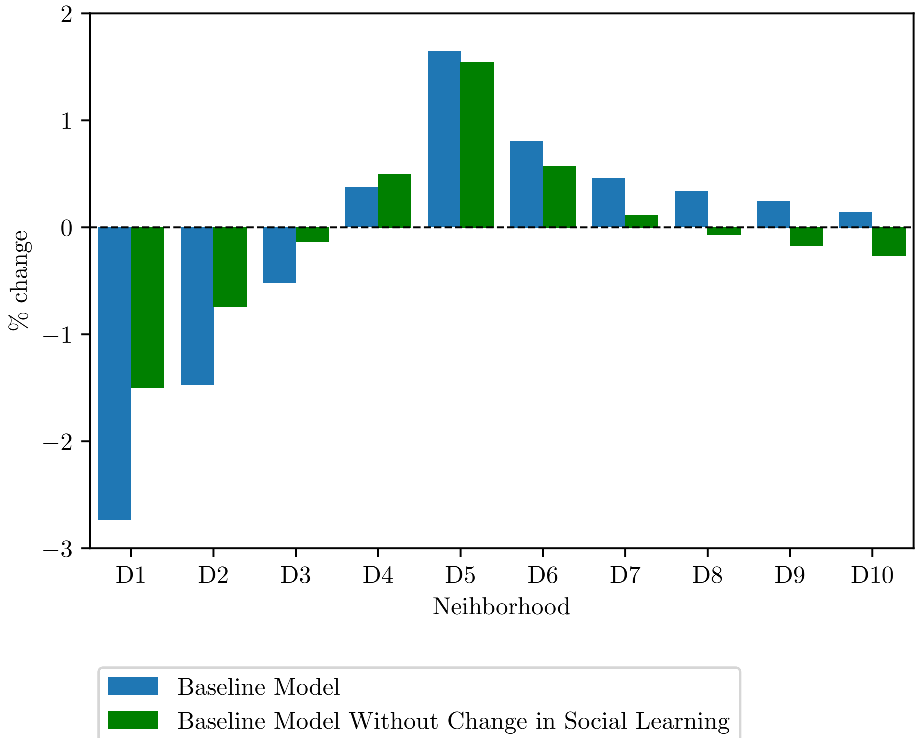
Notes: This table shows the effects of scaling-up housing vouchers by income group.

A surprising effect of the policy is the increase in high-income households' subjective beliefs (Column 5 Table 1.6). A decrease in segregation should improve information, decreasing the subjective beliefs of high-income parents who overestimate them. However, the policy has a unique rent ceiling for the commuting zone, which creates a bunching at the rent limit. This bunching is illustrated by the peak in neighborhood 5 in Figure 1.4 and observed in the data by Collinson and Ganong (2018). As a result, low-income households don't move to the highest-quality neighborhoods, and information does not improve in those neighborhoods. Since there is a debate concerning the unique rent feature of the Housing Choice Voucher program, I evaluate a housing voucher policy with rent limits set at the neighborhood level instead of at the commuting zone level.

1.4.2 Housing Voucher Policy with Rent Limits Indexed at the Neighborhood Level

Since Chetty et al. (2016) provided evidence that housing vouchers effectively improve adulthood earnings through improved neighborhood quality, the United States Department of

Figure 1.4: Effects of Housing Vouchers on Neighborhood Density



Notes: This Figure shows the long-run change in neighborhood density due to the scaled-up housing voucher policy.

Housing and Urban Development has put additional effort into promoting high-quality mobility. In particular, since 2019, the Department of Housing and Urban Development has allowed housing agencies to set voucher subsidies at local rents rather than at the metro area level.[36] This decision addresses an issue raised about unique rent ceilings: they do "not adequately help families access low-poverty neighborhoods."[37] Indeed, Collinson and Ganong (2018) finds that most rental units below the payment standard are in low-quality neighborhoods and that indexing rent limits to ZIP codes rather than to metropolitan areas improves the share of families who move into higher-quality neighborhoods.

Housing Voucher Policy with Multiple Rent Limits

I consider a housing voucher policy designed to mimic the Housing Choice Voucher program but with rent limits defined at the neighborhood level. Since the model does not feature heterogeneity in rents within neighborhoods, rent limits are determined by the median rent in each neighborhood. As before, eligible households are those below the poverty threshold, defined as the income level at the tenth percentile of the income distribution.

Housing vouchers cover the difference between a fraction of the family's income and a fraction of the median rent in each neighborhood. Those fractions are defined so that, (i) in partial equilibrium, the cost of the policy is the same as the cost of a housing voucher with a unique rent limit and that (ii) the rent faced by an average eligible household is the same across the two policies in the first neighborhood. Under this new housing voucher policy, housing vouchers cover the difference between 70.5% of the median rent and 20% of the family income in each neighborhood. Under this new housing voucher policy (NHV), the rent price in the neighborhood m for a parent of income $y(h, s)$ who is eligible for the housing voucher is:

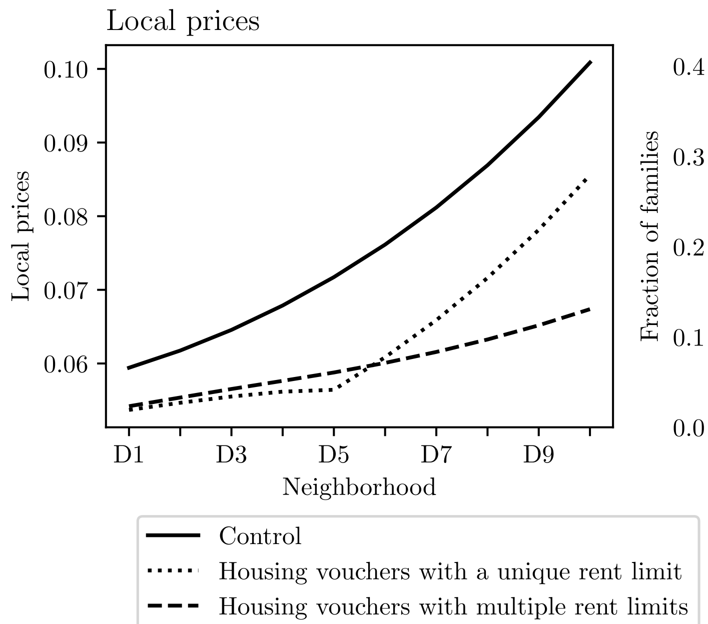
$$r^{NHV}_{m,h,s} = \min(0.2 \times y(h, s) + 0.295 * r_m,\ r_m).$$

The left panel of Figure 1.5 illustrates the rent schedule of this new policy for an average eligible household. While under a housing voucher policy with a unique rent limit, the rent schedule has a kink; it is smooth under a housing voucher policy with rent limits set at the neighborhood level.

[36] https://www.cbpp.org/research/housing/what-are-housing-mobility-programs-and-why-are-they-needed

[37] https://www.huduser.gov/portal/pdredge/pdr_edge_frm_asst_sec_061515.html

Figure 1.5: Design of Housing Voucher Policies - Eligible Households - Partial Equilibrium



Notes: The left panel of this Figure shows the rent schedule for an average eligible household under three scenarios: no policy (control), housing vouchers with a unique rent limit, and housing vouchers with multiple rent limits. The right panel of this Figure shows the residential choices of eligible households under the three scenarios.

Effects of a Housing Voucher Policy with Multiple Rent Limits

Partial Equilibrium Effects of the Policy. Consistent with Collinson and Ganong (2018)'s empirical findings, I find that a housing voucher policy with multiple rent limits further improves the neighborhood rank of eligible households even though more families choose to live in the bottom quality neighborhood (right panel of Figure 1.5 and Column 1 Table 1.7). As a result, the effect on children's future earnings is more substantial than under a unique rent limit housing voucher policy, with an increase of $809 per year at age 26.

Table 1.7: The Effects of Scaling-up Housing Vouchers with Multiple Rent Limits

	Small Field	Large Scale and Long Run	
	Baseline model	Baseline model	
Households	Eligible (1)	Eligible (2)	All (3)
% in bottom-quality neighborhood	-53%	-69%	-3%
Children's future earnings	+$809	+$2,477	-$1
Neighborhood rank	+1.8	+3.3	+0.0
Parental time (min./day)	+1	+10	-1
Inequality			-1.5%
Poverty			-6.2%
Earnings			-0.0%
Rank-rank coefficient			-8.5%

Notes: This table shows the effects of scaling-up housing vouchers with multiple rent limits within the calibrated baseline model.

Large Scale and Long-run Effects of the Policy. Column 3 of Table 1.8 shows that general equilibrium responses in local prices and subjective beliefs greatly amplify the effects of the policy on eligible households. The neighborhood rank is much higher than under a policy with a unique rent limit (3.3 versus 1.6), increasing children's future earnings by $2,477 per year. The difference in amplification effects between the two policies mainly results from improved housing market access. The attractiveness of the bottom-quality neighborhood decreases even further, as depicted in Figure 1.6, while the attractiveness of very high-quality neighborhoods increases. Ssubjective beliefs of low-income families increase under both policy regimes (Column 2 of Tables 1.6 and 1.8).

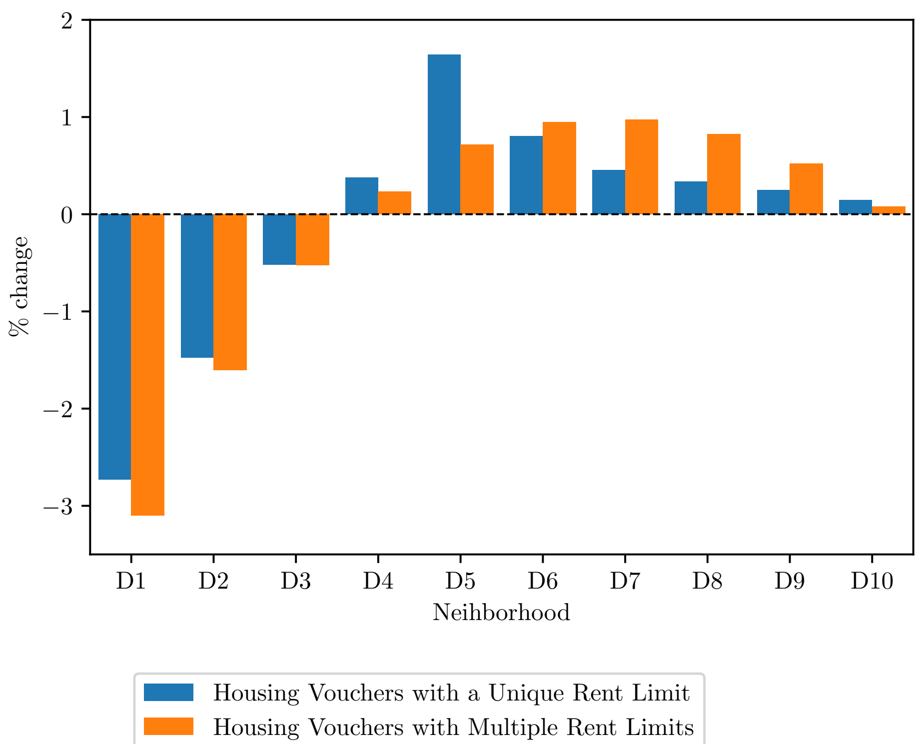
The housing voucher policy with multiple rent limits also positively affects the economy. Column 3 of Table 1.7 shows the policy effects on all households. This policy has more

Table 1.8: The Effects of Scaling-up Housing Vouchers with Multiple Rent Limits by Income Quartile

		Household Income quartile			
	All	1st	2nd	3rd	4th
Subjective Beliefs	-0.3%	+6.6%	+0.9%	-1.7%	-4.8%
Parental Time	-1.1%	+7.8%	-0.6%	-3.2%	-5.1%
Earnings	-0.0%	+0.7%	+2.2%	+1.9%	-0.6%

Notes: This table shows the effects of scaling-up housing vouchers with multiple rent limits by income group.

Figure 1.6: Effects of Housing Voucher Policies on Neighborhood Density



Notes: This Figure shows the long-run change in neighborhood density due to the scaled-up housing voucher policies.

substantial effects on inequality and social mobility (+1.5%, +8.5%, respectively). However, the effects on earnings and, consequently, on absolute poverty are weaker. Figure 1.6 and Table 1.8 help understand the underlying reasons. Under the housing voucher policy with multiple rent limits, eligible households have better access to all neighborhoods, decreasing segregation and improving information everywhere. In the long run, parents' subjective beliefs become closer to the truth, increasing the subjective beliefs of low-income parents but decreasing those of high-income parents. This leads to a change in parents' behavior across the entire income distribution. Low-income parents spend more time with their children (+8%), and high-income parents spend less time with their children (-5%). As a result, average parental time decreases (-1%), and earnings do not improve. However, the policy significantly affects social mobility and inequality as it mitigates the distorting effects of parents' subjective beliefs.

In sum, a housing voucher policy with multiple rent limits is a better tool to address redistribution concerns and improve information, mitigating the distorting effect of parental subjective beliefs than a housing voucher policy with a unique rent limit. However, this policy does not enhance aggregate income in the long run because it decreases high-income parents' subjective beliefs.

In the following section, to further support the paper's findings, I empirically test two of the model's implications.

1.5 Neighborhood Quality and Parental Time in the United States

In addition to the extensive subjective beliefs literature, the model builds on suggestive evidence from the National Longitudinal Study of Adolescent to Adult Health (Add Health). In this section I derive and test for two implications of the model. None of the implications are rejected which comforts the plausibility of the social learning mechanism.

1.5.1 Data

The Add Health survey is a nationally representative longitudinal survey of adolescents in the United States. In the academic year 1994-1995, about 20,000 students in grades 7-12 were sampled to complete an in-home interview. In 2016-2018, about 12,300 answered the last follow-up survey. The analysis focuses on the baseline survey in 1994 (Wave I), when interviewees were aged between 12 and 17, and the last follow-up survey in 2018 (Wave V), when interviewees were aged between 35 and 40.

The Add Health data includes detailed information on parents, children, and neighborhood characteristics which allows me to construct two parental inputs in the child skill production function: parental time and neighborhood quality. I proxy neighborhood quality by the household median income of the census tract in which the interviewee resided in 1994 and 2018.[38] Parental time is approximated to the number of parent-child activities over the past four weeks measured in 1994-1995.[39]

As a first step, I verify the two constructed variables correctly proxy for two parental inputs of the technology of skill formation. Columns (1) and (2) of Table 1.9 show that parental time and neighborhood quality positively correlate with later child skills. Conditional on other parents' socioeconomic status (SES), the number of parent-child activities and neighborhood quality co-move with the graduation probability measured twenty-two years later. All coefficients are positive and significant at a one percent level. Motivated by this evidence, in the following, I consider the parental time and neighborhood quality variables are good proxies for parental inputs of the technology of skill formation.[40]

1.5.2 Correlation between Time and Neighborhood

Suppose neighborhood quality and parental time are two inputs of the technology of skill formation. Assume parents' decisions are driven by their perceived value of the returns to both inputs – neighborhood quality and parental time. All else equal, parents with low (high) subjective beliefs will tend to live in worse (better) quality neighborhoods and spend less (more) time with their children. I expect to see a positive correlation between time and neighborhood quality in the data due to the omitted subjective beliefs variable. The data support this assumption.

Columns (3) and (4) of Table 1.9 display a positive and significant correlation between parental time measured by the number of parent-child activities done over four weeks and standardized neighborhood quality proxied by household median income.[41] Note that in both regressions of Column (3) and (4) of Table 1.9, I control for school ID fixed effects. A systematic difference in school characteristics is not what drives the correlation. Nevertheless, research finds that parental time increases with parents' education in the United States

[38] In Appendix Section 1.E I also use the share of residents above 25 with a college degree as a robustness check of the results (Diamond, 2016).

[39] Appendix Section 1.A.1 describes the data and variable construction in more details.

[40] Appendix Table 1.E.1 presents the OLS regression coefficients with a different definition of neighborhood quality and parental time variables. Results are robust to definition changes.

[41] The results are robust to the use of alternative proxies for parental inputs. See Appendix Table 1.E.1.

(see Doepke et al. (2022) for a review).[42] This observed correlation could be driven by a neighborhood composition effect. In Column (4) of Table 1.9, I control for three variables measuring households' socioeconomic status: parents' highest level of education, family income, and parent's marital status.[43] The sample size shrinks to 12,633 observations due to missing values in household characteristics variables, but the relationship remains positive and significant at a one percent level. This suggests that neighborhood composition effect does not drive all the correlation between the two parental inputs.

Table 1.9: Parental Time and Neighborhood Quality, Add-Health 1994-1995 and 2016-2018

	College Graduation 2016-2018		Parental Time 1994-1995		Neighborhood (std) 2016-2018	
	(1)	(2)	(3)	(4)	(5)	(6)
Neighborhood (std) 1994-1995	0.56*** (0.012)		0.212*** (0.03)	0.078*** (0.028)	0.146*** (0.038)	0.111** (0.043)
Parental Time 1994-1995		0.031*** (0.006)				
Race						0.08 (0.051)
Controls:						
Childhood SES	yes	yes	no	yes	yes	yes
Adulthood SES	no	no	no	no	no	yes
Same decile census tract	no	no	no	no	yes	yes
School ID FE	yes	yes	yes	yes	no	no
Commuting zone FE	no	no	no	no	yes	yes
No. of obs.	8,518	8,543	17,102	12,633	7,982	6,064
Clusters	101	101	120	105	97	90

Notes: This table provides between parental investment choices and college graduation (Columns (1) and (2)) correlations between parental investment choices (Columns (2) and (3)) and between childhood and adulthood neighborhood quality (Columns (5) and (6)). Socioeconomic status variables are family income, highest education level, and marital status. See Appendix Section 1.A.1 for details on variable construction. The dependent variables are indicated in the column title. All regressions control for age fixed effects. All observations are weighted by Wave specific sample weights provided by AddHealth. Standard errors displayed in brackets are clustered at the county level. *p<0.1; **p<0.05; ***p<0.01.

While this result is consistent with the assumption of parental beliefs being an omitted variable, it is also compatible with a human capital production function in which neighborhood quality and parental time are complementary inputs. There is no consensus in the

[42] In Appendix 1.D.1 I analyze the ATUS data set, and consistent with Guryan et al. (2008), I find a positive relationship between parental time and parental education.

[43] Not displayed in the table, the coefficient of parents' education is positive and significant, which is consistent with the educational gradient in childcare time observed in ATUS.

literature about this relationship.[44]

1.5.3 Childhood Neighborhood and Adulthood Choices

The second testable implication of the mechanism is that childhood neighborhood quality and later parental decisions, including neighborhood quality, are positively correlated due to social learning. In poor (wealthy) neighborhoods, I expect children to become pessimistic (over-optimistic) about parenting and later on to under-(over-)invest in their own child's human capital. In the data, childhood neighborhood quality should be positively correlated with later neighborhood choices.

Column (5) of Table 1.9 displays a positive and significant correlation between childhood (Wave I) and adulthood (Wave V) neighborhood quality. I restrict the sample to interviewees who do not live with their parents and control for commuting zone and age fixed effects. The correlation remains highly significant even after controlling for household income, marital status, and education status. Other factors might affect residential choices. Racial discrimination and wealth are two of them. In Column (6) of Table 1.9, to proxy for inherited wealth, I control for three variables measuring the parents' socioeconomic status: parents' highest level of education, family income, and parents' marital status. In addition, I control for the race of the interviewee.[45] The coefficient of relationships between neighborhood qualities barely decreases and remains highly significant. Interestingly, the race coefficient is insignificant, suggesting that race is not the main driver of residential quality choices after controlling for neighborhood quality.[46]

Neither of the two testable implications of the model is rejected. Combined with the extensive literature on subjective beliefs and the great fit of the calibrated model, this suggestive evidence supports the plausibility of the social learning mechanism.

1.6 Conclusion

In this paper, I present a quantitative spatial model of residential and parental time decisions with social learning about the technology of skill formation. Introducing endogenous subjec-

[44] Chyn and Daruich (2022) find a complementarity between time and neighborhoods, but Agostinelli (2018) and Agostinelli et al. (2022)'s calibrated models imply that parental time and environment quality are substitute inputs in producing children's skills. To my knowledge, all empirical estimates of the relationship between parental time and environment quality suggest that they are substitutes (Kling et al., 2001; Pop-Eleches and Urquiola, 2013; Das et al., 2013; Kiessling, 2021).

[45] The variable is one if the race is white, zero otherwise.

[46] The results are robust to the use of the fraction of adults with a college degree for neighborhood quality. See Appendix Table 1.E.1.

tive beliefs helps understand parental behavior across socioeconomic groups. Once calibrated to the average commuting zone in the United States, the model predicts that segregation generates information frictions that shape the subjective beliefs distribution and distort parental investment choices. Low-income parents underestimate the returns to neighborhood quality and parental time, while high-income parents overestimate them. This model provides a rationale for two puzzling parental behaviors: children born in low-quality neighborhoods tend to raise their children in those neighborhoods, and college parents spend more time with their children than non-college parents despite working more hours.

I investigate the effects of a housing voucher policy that induces low-income households to move to higher-quality neighborhoods. Scaling up the policy amplifies the effects on eligible families and positively impacts the economy as it decreases inequality and poverty and increases social mobility in the long run. A change in subjective beliefs mainly drives this amplification effect. Ignoring this change would lead to misleading policy recommendations. Finally, I also find that a housing voucher policy with multiple rent limits within the commuting zone instead of one is a better tool to reduce the distorting effect that social learning and segregation introduce.

This paper shows that a relatively low level of delusion about the technology of skill formation, consistent with micro-studies, has a significant adverse effect on the economy. Estimating subjective beliefs' consequences at the aggregate level and proposing concrete policies that can dampen their negative impacts is an exciting avenue to pursue in future research.

Bibliography

Adam, K., Marcet, A., and Beutel, J. (2017). Stock price booms and expected capital gains. *American Economic Review*, 107(8):2352–2408.

Agostinelli, F. (2018). Investing in children's skills: An equilibrium analysis of social interactions and parental investments. *Unpublished Manuscript, University of Pennsylvania*.

Agostinelli, F., Doepke, M., Sorrenti, G., and Zilibotti, F. (2022). When the great equalizer shuts down: Schools, peers, and parents in pandemic times. *Journal of public economics*, 206:104574.

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The economics of density: Evidence from the berlin wall. *Econometrica*, 83(6):2127–2189.

Alan, S., Boneva, T., and Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*, 134(3):1121–1162.

Algan, Y., Bouguen, A., Chevallier, C., and Charpentier, A. (2020). The impact of a large-scale mindset intervention on school outcomes: Experimental evidence fromfrance. *Working paper*.

Attanasio, O., Boneva, T., and Rauh, C. (2019). Parental beliefs about returns to different types of investments in school children. Working Paper 25513, National Bureau of Economic Research.

Attanasio, O. P. and Kaufmann, K. M. (2014). Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, 109:203–216.

Belfield, C., Boneva, T., Rauh, C., and Shaw, J. (2019). What drives enrolment gaps in further education? the role of beliefs in sequential schooling decisions. *Economica*.

Bergman, P., Chetty, R., DeLuca, S., Hendren, N., Katz, L. F., and Palmer, C. (2019). Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. Working Paper 26164, National Bureau of Economic Research.

Bilal, A. (2021). The geography of unemployment. Working Paper 29269, National Bureau of Economic Research.

Blandin, A. and Herrington, C. (2022). Family heterogeneity, human capital investment, and college attainment. *American Economic Journal: Macroeconomics*, 14(4):438–78.

Boneva, T. and Rauh, C. (2016). Human capital production and parental beliefs. *Unpublished Manuscript, University College London*.

Boneva, T. and Rauh, C. (2018). Parental beliefs about returns to educational investments—the later the better? *Journal of the European Economic Association*, 16(6):1669–1711.

Breda, T., Grenet, J., Monnet, M., and van Effenterre, C. (2019). Female Role Models: are they effective at encouraging girls to study science? IPP Policy Brief n°45, Institut des politiques publiques.

Caucutt, E. M., Lochner, L., and Park, Y. (2017). Correlation, consumption, confusion, or constraints: Why do poor children perform so poorly? *The Scandinavian Journal of Economics*, 119(1):102–147.

Chetty, R. and Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R. and Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.

Chetty, R., Hendren, N., Kline, P., Saez, E., and Turner, N. (2014). Is the united states still a land of opportunity? recent trends in intergenerational mobility. *American Economic Review*, 104(5):141–47.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M., and Wernerfelt, N. (2022). Social capital i: measurement and associations with economic mobility. *Nature*.

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review*, 108(10):3028–56.

Chyn, E. and Daruich, D. (2022). An equilibrium analysis of the effects of neighborhood-based interventions on children. Working Paper 29927, National Bureau of Economic Research.

Collinson, R. and Ganong, P. (2018). How do changes in housing voucher design affect rent and neighborhood quality? *American Economic Journal: Economic Policy*, 10(2):62–89.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American economic review*, 100(1):35–69.

Cunha, F., Elo, I., and Culhane, J. (2013). Eliciting maternal expectations about the technology of cognitive skill formation. Working Paper 19144, National Bureau of Economic Research.

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

Cunha, F. and Heckman, J. J. (2009). The economics and psychology of inequality and human development. *Journal of the European Economic Association*, 7(2-3):320–364.

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

Daruich, D. (2018). The macroeconomic consequences of early childhood development policies. Working Paper 2018-29, FRB St. Louis.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K., and Sundararaman, V. (2013). School inputs, household substitution, and test scores. *American Economic Journal: Applied Economics*, 5(2):29–57.

Diamond, R. (2016). The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review*, 106(3):479–524.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review*, 109(8):2728–65.

Doepke, M., Hannusch, A., Kindermann, F., and Tertilt, M. (2022). The economics of fertility: A new era. Working Paper 29948, National Bureau of Economic Research.

Durlauf, S. N. (2011). *Chapter 6. Groups, Social Influences, and Inequality*, pages 141–175. Princeton University Press.

Eckert, F., Kleineberg, T., et al. (2019). Can we save the american dream? a dynamic general equilibrium analysis of the effects of school financing on local opportunities. In *2019 Meeting Papers*, number 1197. Society for Economic Dynamics.

Enke, B. (2020). What you see is all there is. *The Quarterly Journal of Economics*, 135(3):1363–1398.

Fernández, R. (2013). Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review*, 103(1):472–500.

Fogli, A. and Guerrieri, V. (2019). The end of the american dream? inequality and segregation in us cities. Working Paper 26143, National Bureau of Economic Research.

Fogli, A. and Veldkamp, L. (2011). Nature or nurture? learning and the geography of female labor force participation. *Econometrica*, 79(4):1103–1138.

Frick, M., Iijima, R., and Ishii, Y. (2022). Dispersed behavior and perceptions in assortative societies. *American Economic Review*, 112(9):3063–3105.

Fuchs-Schündeln, N., Krueger, D., Ludwig, A., and Popova, I. (2022). The long-term distributional and welfare effects of covid-19 school closures. *The Economic Journal*, 132(645):1647–1683.

Gregory, V., Kozlowski, J., and Rubinton, H. (2022). The impact of racial segregation on college attainment in spatial equilibrium. Working Paper 2022-36, FRB St. Louis Working Paper.

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

Guyon, N. and Huillery, E. (2020). Biased Aspirations and Social Inequality at School: Evidence from French Teenagers. *The Economic Journal*.

Heckman, J. J. and Mosso, S. (2014). The economics of human development and social mobility. Working Paper 19925, National Bureau of Economic Research.

Hofferth, S. L., Flood, S. M., Sobek, M., and Backman, D. (2020). American time use survey data extract builder: Version 2.8 [dataset]. *College Park, MD: University of Maryland and Minneapolis, MN: IPUMS.*

Hwang, Y. (2022). Neighborhood effects on intergenerational cultural transmission. Working Paper 4017952, SSRN.

Jang, Y. and Yum, M. (2020). Aggregate and Intergenerational Implications of School Closures: A Quantitative Assessment. Discussion Paper Series 234v1, CRC TR 224.

Jehiel, P. (2018). Investment strategy and selection bias: An equilibrium perspective on overoptimism. *American Economic Review*, 108(6):1582–97.

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2):515–548.

Kaufmann, K. M. (2014). Understanding the income gradient in college attendance in mexico: The role of heterogeneity in expected returns. *Quantitative Economics*, 5(3):583–630.

Kiessling, L. (2021). How do parents perceive the returns to parenting styles and neighborhoods? *European Economic Review*, 139:103906.

Kim, S., Tertilt, M., and Yum, M. (2021). Status externalities in education and low birth rates in korea. Discussion Paper No. 16271, CEPR Press.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2001). Bullets don't got no name: Consequences of fear in the ghetto. Working paper, JCPR.

Kopczuk, W., Saez, E., and Song, J. (2010). Earnings inequality and mobility in the united states: evidence from social security data since 1937. *The Quarterly Journal of Economics*, 125(1):91–128.

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

Manson, S., Schroeder, J., Van Riper, D., Kugler, T., and Ruggles, S. (2022). Ipums national historical geographic information system: Version 17.0 [dataset]. *Minneapolis, MN: IPUMS.*

Nakamura, E., Sigurdsson, J., and Steinsson, J. (2022). The gift of moving: Intergenerational consequences of a mobility shock. *The Review of Economic Studies*, 89(3):1557–1592.

Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from madagascar. *Unpublished manuscript*, 6.

Pop-Eleches, C. and Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. *American Economic Review*, 103(4):1289–1324.

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9:21–58.

Roemer, J.-E. and Wets, R.-J.-B. (1994). Neighborhood Effects on Belief Formation and the Distribution of Education and Income. Papers 94-02, California Davis - Institute of Governmental Affairs.

Streufert, P. (2000). The effect of underclass social isolation on schooling choice. *Journal of Public Economic Theory*, 2(4):461–482.

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics letters*, 20(2):177–181.

Wiswall, M. and Zafar, B. (2021). Human capital investments and expectations about career and family. *Journal of Political Economy*, 129(5):1361–1424.

Yum, M. (2022). Parental time investment and intergenerational mobility. *International Economic Review*.

Appendices

1.A Data

1.A.1 National Longitudinal Study of Adolescent to Adult Health (Add Health)

Description

The National Longitudinal Study of Adolescent to Adult Health (Add Health) survey is a nationally representative longitudinal survey of adolescents in the United States. In academic year 1994-1995, about 20,000 students in grades 7-12 were sampled to complete an in-home interview. They come from 132 schools and in 1994-1995, most of them are aged between 12 and 17 years old. In 2016-2018, about 12,300 of them have answered Wave V survey. At the date of the last survey wave, most of the interviewees are aged between 35 and 40 years old. The analysis focuses on the baseline survey in 1994 (Wave I) and the last follow up survey in 2018 (Wave V).

The data set includes detailed information on family background and a rich set of information on neighborhoods characteristics. In 1994, we observe the highest education level of the parents of about 17,000 interviewed adolescents, and we have information on residential neighborhood characteristics for about 18,100 adolescents. Information on neighborhood is available at the census tract level. In addition, Add Health contains questions on the frequency of a ten parent-child activities. Add-Health structure allows me to correlate the number of parent-child activities with neighborhood characteristics which can't be done using the more detailed American Time use survey (ATUS) (see Appendix Section 1.A.3).

Neighborhood and Parental Time

I proxy neighborhood quality by household median income of the census tract.[47] I create ten synthetic neighborhoods by ranking all census tracts by neighborhood quality and grouping them in ten groups of equal size. One synthetic neighborhood represent a decile of the census tracts distribution in the United States. Thanks to the panel form of the data set, I can observe in which synthetic neighborhood an adolescent lives in 1994-1995 and in which synthetic neighborhood she lives during adulthood, in 2016-2018.

Add Health survey contains information about ten parent-child activities in 1994-1995.[48] To proxy for time allocated in childcare, I construct a variable that counts the number of activities that happened over the past four week with the mother and the father of the child. I follow the United States Bureau of Labor Statistics (BLS) definition of "primary childcare activities" constructed and exclude two out the ten events: "shopping", "went to a religious service or church-related event" and.[49] The remaining eight activities are: "played a sport", "talked about someone you're dating, or a party you went to", "talked about a personal problem you were having", "had a serious argument about your behavior", "talked about your school work or grades", "worked on a project for school", "talked about other things you're doing in school", "went to a movie, play, museum, concert, or sports event".

1.A.2 National Historical Geographic Information System (NHGIS)

The National Historical Geographic Information System (NHGIS) provides access to summary tables and time series of population, housing, agriculture, and economic data, along with GIS-compatible boundary files, for years from 1790 through the present and for all levels of U.S. census geography, including states, counties, tracts, and blocks. Data is easily accessible on the IPUMS Website (Manson et al., 2022).

[47] I also use the share of residents above 25 with a college degree as a robustness check of the results (Diamond (2016)).

[48] The question of interest is: "Which of the things listed on this card have you done with [resident mother/father] in the past four weeks (check all that apply)".

[49] The definition of "primary childcare activities" that is provided by the United States Bureau of Labor Statistics (BLS) includes time spent providing physical care; playing with children; reading with children; assisting with homework; attending children's events; taking care of children's health needs; and dropping off, picking up, and waiting for children. It does not include activities with children without active childcare such as "watching television with my child". As a robustness check, I construct an alternative proxy without the activity"went to a movie, play, museum, concert, or sports event".

1.A.3 The American Time Use Survey (ATUS)

Description

The American Time Use Survey (ATUS) is a nationally representative survey of Americans aged 15 or over that provides extensive information on how and where Americans spend their time. From 2003 to 2020, almost 219,000 interviews were conducted, and all of those can be linked to data files from the Current Population Survey (CPS). I use already linked data sets provided by Hofferth et al. (2020) and available on the IPUMS website. As all the other datasets are from the year 2000, I use the earliest ATUS survey year, 2003.

The ATUS asked interviewees about their activities the day before the interview–the "diary day." For each activity, respondents are asked how long the activity lasted and, for most activities, where they were and who was with them. After completing the time diary, there are additional questions to identify work, volunteering, eldercare, and secondary childcare activities. Several activity categories are then defined by the United States Bureau of Labor Statistics (BLS), among which one is of particular interest for this research project: "primary childcare activities." This activity category includes time spent providing physical care, playing with children; reading with children; assisting with homework; attending children's events; taking care of children's health needs, and dropping off, picking up, and waiting for children. Passive childcare was included as a primary activity (such as "keeping an eye on my son while he swam in the pool"). However, a child's presence during the training is not enough to classify it as childcare; "watching television with my child" is coded as a leisure activity, not as childcare.

Parental Time

To measure parental time, I first restrict the sample to individuals in a two-parent household, defined as married individuals whose youngest child is under age 18. In 2003, 5,597 married parents were interviewed, among which 2,168 have a college degree.[50]. I use the BLS definition of childcare as "primary childcare activities" and respect the activity coding structure of ATUS to describe how parents from different socioeconomic groups allocate their time between childcare, leisure, and market work. Market work hours include all hours spent in work-related and education-related activities. The rest of the time is considered leisure time. For informational purposes, I divide this last group into two sub-groups: all personal leisure

[50] I removed 255 observations that have less than 23 hours of activity a day reported. 92% of the remaining observations have exactly 24 hours of activity a day reported.

activities and other types of activities.[51]

[51] Personal leisure is composed by eight activities: "eat and drink", "personal care", "telephone calls", "professional and personal care services", "religious and spiritual activities", "socializing, relaxing, and leisure", "sports, exercise, and recreation", "volunteer activities". I remove eight hours of sleep needs from "personal care" that includes sleep time, a standard assumption in the literature. Others are "household activities," "household services," "government services and civic obligations," "consumer purchases," "travel,"' and "caring for and helping non-household and [other] household members." Appendix Table 1.A.1 provides detailed information on each of those activities.

Table 1.A.1: ATUS Activity Coding Structure, 2003

Label	Description	Non-college parents	College parents
		Hours per day	
Childcare		**1.17**	**1.58**
Caring for and helping household children	Time spent in caring for or helping household children	1.17	1.58
Work and Education		**4.36**	**4.66**
Working and Work-related Activities	Time spent in work activities such as working, doing activities as part of one's job, engaging in income-generating activities (not as part of one's job), and looking for jobs and interviewing.	4.27	4.56
Educational activities	Time spent in non-work education activities, such as taking classes, conducting research and homework, administrative tasks, and extracurricular activities except sports.	0.10	0.10
Personal leisure		**14.39**	**13.64**
Personal care	Time spent in personal care activities such as sleeping, grooming, and health-related self care.	9.05	8.66
Eat and drinking	Time spent in activities such as eating and drinking not done as work or a volunteer activity, whether the respondent was alone, with others, at home, at a place of purchase, in transit, or somewhere else.	1.02	1.19
Socializing, relaxing, and leisure	Time spent in personal interest or leisure activities such as communicating with others and attending parties and meetings; and leisure activities such as relaxing, playing (passive) games (unless playing with children only), watching television, playing or listening to music, reading, writing, and all hobbies.	3.69	2.88
Sports, exercise, and recreation	Time spent in sports, exercise and recreational activities such as pleasure boating, throwing a Frisbee, kite flying, or ballooning, and active, participatory outdoor games or activities, such as horseshoes, croquet, and paintball.	0.25	0.33

Religious and spiritual activities	Time spent in work activities such as working, doing activities as part of one's job, engaging in income-generating activities (not as part of one's job), and looking for jobs and interviewing.	0.14	0.17
Volunteer activities	Time spent in volunteer (unpaid) activities done by the respondent for individuals or institutions through formal organizations, such as unpaid performance arts activities, socializing related to volunteering and attending church service as a volunteer activity.	0.13	0.25
Telephone calls	Time spent in telephone communication activities such as talking on phone, waiting for a phone call or Skyping (2011+).	0.05	0.07
Professional and personal care services	Time spent in activities such as obtaining, receiving, and/or purchasing professional and personal care services provided by someone else. Professional services include child care, financial, legal, medical and other adult care, real estate, and veterinary. Personal care service activities include massages, haircuts, manicures, and tanning at salons.	0.07	0.09
Other		**4.04**	**4.07**
Household activities	Time spent in household activities such as maintaining their household, household management and organizational activities.	2.13	1.94
Caring for and helping household members (except household children)	Time spent in caring for or helping any adult in the respondent's household, regardless of relationship, age, or physical or mental health status.	0.04	0.03
Caring for and helping non-household members	Time spent in caring for or helping any child or adult who is not part of the respondent's household, regardless of relationship, age, or physical or mental health status.	0.10	0.08

Household services	Time spent in activities such as obtaining and purchasing household services provided by someone else. Household services include yard and house cleaning, cooking, pet care, tailoring and laundering services, and vehicle and home repairs, maintenance, and construction. Watching someone else perform paid household activities (cooking, cleaning, repairing, etc.) should be coded here, provided "watching" was the respondent's primary activity.	0.01	0.02
Government services and civic obligations	Time spent in activities such as using government services (police, fire, social services), purchasing government-required licenses or paying fines or fees, fulfilling government-required duties (jury duty, parole meetings, court appearances), and participating in activities that assist or impact government processes (voting, town hall meetings).	0.00	0.00
Consumer purchases	Time spent in activities such as purchases and rentals of consumer goods, regardless of mode or place of purchase or rental (in person, via telephone, over the internet, at home, or in a store).	0.43	0.49
Travel	Time spent in travel or transportation activities such as commuting, walking someplace or waiting for the bus or train.	1.33	1.51
Total		**24.0**	**24.0**

Notes: This table provides information on time allocation by married couples with a child below 18 in the household in the United States. Data source: ATUS 2003.

1.B Equilibrium Computation

The below algorithm uses an iterative method to find the steady state.

1. Make an initial guess for the distribution $F(h, s, m_0, \tilde{\alpha})$ and value function $U(h, s, m_0, \tilde{\alpha})$.

2. Given $U(h, s, m_0, \tilde{\alpha})$, compute the policy function $\tau(h, s, m_0, \tilde{\alpha}, m)$ and the corresponding $V(h, s, m_0, , \tilde{\alpha}, m)$

3. Make an initial guess for rent prices $\{r_m\}_{m \in \mathbb{M}}$

4. Given $V(h, s, m_0, \tilde{\alpha}, m)$, $\tau(h, s, m_0, \tilde{\alpha}, m)$ and $\{r_m\}_{m \in \mathbb{M}}$, compute the share of families $(h, s, m_0, \tilde{\alpha})$ in each neighborhood ($\lambda_m(h, s, m_0, \tilde{\alpha})$ for every m).

5. Compute $\{r_m\}_{m\in\mathbb{M}}$ given the share of families in each neighborhood.

6. Iterate 3 to 5 until $\{r_m\}_{m\in\mathbb{M}}$ converges.

7. Given $\{r_m\}_{m\in\mathbb{M}}$, compute the expected value function $U(h,s,m_0,\tilde{\alpha})$ and based on it, obtain the policy function for time investment $\tau(h,s,m_0,\tilde{\alpha})$.

8. Compute the distribution $G(H,\mathrm{T},m_0,h,m)$ given λ_m, and obtain updated subjective beliefs in each neighborhood $\tilde{\alpha}(\tilde{\alpha},m)$.

9. Compute the time invariant distribution$F(h,s,m_0,\tilde{\alpha})$, based on the initial guess, the policy functions for neighborhoods $\lambda_m(h,s,m_0,\tilde{\alpha})$, and $\tau(h,s,m_0,\tilde{\alpha},m)$, and on beliefs updating $\tilde{\alpha}(\tilde{\alpha},m)$ obtained above.

10. Iterate from 1 to 9 until $F(h,s,m_0,\tilde{\alpha})$ converges.

1.C Theory Appendix

Assume $x \sim \mathcal{N}(\overline{\mu},\sigma_s^2)$ and define the function $\text{bound}(x,d) := x\mathbb{1}_{\{-d\le x\le d\}} + d\mathbb{1}_{\{x>d\}} - d\mathbb{1}_{\{x<d\}}$.

Let $d>0$ constant and $y := \text{bound}(x,d)$. Let Φ the CDF of the standard normal distribution and ϕ the PDF of the standard normal distribution.

Then it holds:

1. If $\overline{\mu}\ge 0$ then $\mathbb{E}(y)\ge 0$ and if $\overline{\mu}\le 0$ then $\mathbb{E}(y)\le 0$

2. If $\overline{\mu}\ge 0$ then $\mathbb{E}(y)\le\overline{\mu}$ and if $\overline{\mu}\le 0$ then $\mathbb{E}(y)\ge\overline{\mu}$

3. $\lim_{\sigma_s\to 0}\mathbb{E}(y)=\overline{\mu}$ and $\lim_{d\to+\infty}\mathbb{E}(y)=\overline{\mu}$

4. $\lim_{\sigma_s\to+\infty}\mathbb{E}(y)=0$

Figure 1.C.1 illustrates a conditional signal distribution of y given $d>\overline{\mu}>0$.

Preliminary common results for (1)-(4):
Assume $\overline{\mu}\ge 0$

(a) Let

$$\mathbb{E}(y)=\int_{-d}^{d}\frac{u}{\sigma_s}\,\phi\left(\frac{u-\overline{\mu}}{\sigma_s}\right)\,du+d\,\left(1-\Phi\left(\frac{d-\overline{\mu}}{\sigma_s}\right)\right)-d\,\Phi\left(\frac{-d-\overline{\mu}}{\sigma_s}\right)$$

(b) Let $u = v + \overline{\mu}$. By properties of the Gaussian distribution, $\forall u \geq 0$ $\phi(v) = \phi(-v)$ and $\phi(v) \geq \phi(-v - 2\overline{\mu})$ or $\phi(u - \overline{\mu}) = \phi(-u + \overline{\mu})$ and $\phi(u - \overline{\mu}) \geq \phi(-u - \overline{\mu})$

(c) Let

$$\begin{aligned}
\mathbb{E}(x) &= \int_{-\infty}^{\infty} \frac{u}{\sigma_s}\, \phi\left(\frac{u-\overline{\mu}}{\sigma_s}\right)\, du \\
&= \int_{-d}^{d+2\overline{\mu}} \frac{u}{\sigma_s}\, \phi\left(\frac{u-\overline{\mu}}{\sigma_s}\right)\, du \;+\; (d+2\overline{\mu})\, \left(1 - \Phi\left(\frac{d+\overline{\mu}}{\sigma_s}\right)\right) - d\;\Phi\left(\frac{-d-\overline{\mu}}{\sigma_s}\right) \\
&= \overline{\mu}
\end{aligned}$$

Proof. (1) Assume $\overline{\mu} \geq 0$. By (b),

$$\int_0^d \frac{1}{\sigma_s}\, u\, \phi\left(\frac{u-\overline{x}}{\sigma_s}\right)\, du \;\geq\; \left|\int_0^d \frac{1}{\sigma_s}\, (-u)\, \phi\left(\frac{-u-\overline{x}}{\sigma_s}\right)\, du\right| \;=\; \left|\int_{-d}^0 \frac{1}{\sigma_s}\, u\, \phi\left(\frac{u-\overline{x}}{\sigma_s}\right)\, du\right|$$

and

$$\left(1 - \Phi\left(\frac{d-\overline{x}}{\sigma_s}\right)\right) \;\geq\; \Phi\left(\frac{-d-\overline{x}}{\sigma_s}\right) \geq 0$$

Hence, by (a), $\mathbb{E}(y) \geq 0$.

By the symmetry of the Gaussian distribution, if $\overline{\mu} \leq 0 \;\Rightarrow\; \mathbb{E}(y) \leq 0$. □

(2) Assume $\overline{\mu} \geq 0$. By (c),

$$\begin{aligned}
\overline{\mu} = &\int_{-d}^{d} \frac{u}{\sigma_s}\, \phi\left(\frac{u-\overline{\mu}}{\sigma_s}\right)\, du \;+\; \int_{d}^{d+2\overline{\mu}} \frac{u}{\sigma_s}\, \phi\left(\frac{u-\overline{\mu}}{\sigma_s}\right)\, du \\
&- \; d\;\Phi\left(\frac{-d-\overline{\mu}}{\sigma_s}\right) \;+\; (d+2\overline{\mu})\, \left(1 - \Phi\left(\frac{d+\overline{\mu}}{\sigma_s}\right)\right) \\
\geq &\int_{-d}^{d} \frac{u}{\sigma_s}\, \phi\left(\frac{u-\overline{\mu}}{\sigma_s}\right)\, du \;+\; d\, \left(1 - \Phi\left(\frac{d-\overline{\mu}}{\sigma_s}\right) - \left(1 - \Phi\left(\frac{d+\overline{\mu}}{\sigma_s}\right)\right)\right) \\
&- \; d\;\Phi\left(\frac{-d-\overline{\mu}}{\sigma_s}\right) \;+\; (d+2\overline{\mu})\, \left(1 - \Phi\left(\frac{d+\overline{\mu}}{\sigma_s}\right)\right) \\
= &\int_{-d}^{d} \frac{u}{\sigma_s}\, \phi\left(\frac{u-\overline{\mu}}{\sigma_s}\right)\, du \;+\; d\, \left(1 - \Phi\left(\frac{d-\overline{\mu}}{\sigma_s}\right)\right) \\
&- \; d\;\Phi\left(\frac{-d-\overline{\mu}}{\sigma_s}\right) \;+\; 2\overline{\mu}\, \left(1 - \Phi\left(\frac{d+\overline{\mu}}{\sigma_s}\right)\right) \\
= &\;\mathbb{E}(y) + \; 2\overline{\mu}\, \left(1 - \Phi\left(\frac{d+\overline{\mu}}{\sigma_s}\right)\right) \qquad \text{(by (a))} \\
\geq &\;\mathbb{E}(y)
\end{aligned}$$

Hence, $\mathbb{E}(y) \leq \overline{\mu}$.

By the symmetry of the Gaussian distribution, if $\overline{\mu} \leq 0 \;\Rightarrow\; \mathbb{E}(y) \geq \overline{\mu}$. $\square$

(3) Note $\lim_{x \to +\infty} \Phi(x) = 1$ and $\lim_{x \to +\infty} \Phi(-x) = 0$. By (a)

$$\lim_{d \to +\infty} \mathbb{E}(y) = \overline{\mu}$$

Trivially,

$$\lim_{\sigma_s \to 0} \mathbb{E}(y) = \overline{\mu}$$

$\square$

(4) Note $\lim_{\sigma_s \to +\infty} \int_{-d}^{d} \frac{u}{\sigma_s} \, \phi\left(\frac{u-\overline{\mu}}{\sigma_s}\right) \, du = 0$ and $lim_{\sigma_s \to +\infty} \Phi\left(\frac{d-\overline{\mu}}{\sigma_s}\right) = 0.5$.

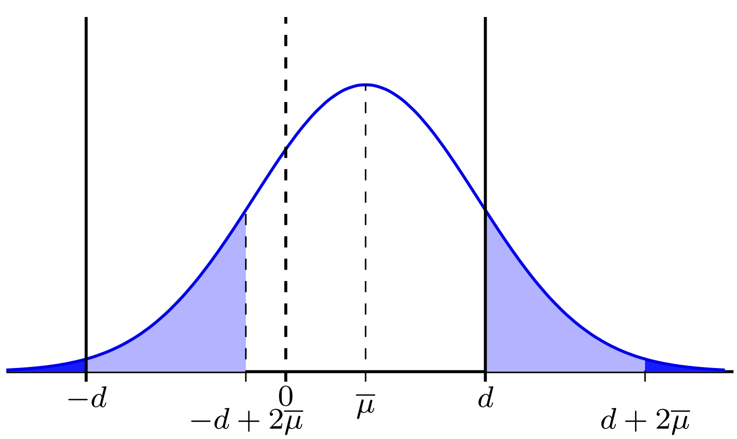
By (a)

$$\lim_{\sigma_s \to +\infty} \mathbb{E}(y) = 0$$

.

$\square$

Figure 1.C.1: Signal Illustration



Notes: This Figure illustrates a conditional signal distribution of y given $d > \overline{\mu} > 0$.

1.D Additional Information on the Calibration

1.D.1 Parental Time

Table 1.D.1 summarizes parents' time use in the United States by education. I first restrict the sample to individuals in a two-parent household, defined as married individuals whose youngest child is under age 18.[52] In 2003, 5,597 married parents were interviewed, among which 2,168 have a college degree.[53,54] As standard in the literature, I assume 16 hours of disposable hours and eight hours of sleep needs per day.

Both parental time and market work time increase in education. Parents with a college degree spend about 1.6 hours per day in childcare activities and 4.7 hours in market work activities, while parents without a college degree spend 1.2 hours in childcare and 4.4 hours in market work.[55] Leisure time mechanically decreases in education. It is interesting to note, however, that this decrease is entirely driven by college parents spending less time on personal leisure activities compared to non-college parents. Time spent in other types of activities is relatively constant across educational groups.[56] In the following, I consider an endowment of 16 disposable hours per day and normalize it to one. Parental time patterns are moments to match.

1.D.2 Synthetic Neighborhoods Characteristics

Table 1.D.2 summarizes the characteristics of these neighborhoods. By construction, median household income increases with neighborhood quality. As expected, so does the fraction of individuals above 25 with a college degree (Column (2)).[57] Note that housing expenditure shares decrease with neighborhood quality (Column (4)) which suggests and motivates non-homothetic preferences.

---

[52] I only keep two-parent households for consistency with the model that does not consider marital status. 78% of parents in the sample are married.

[53] Parent's education is defined as the highest level of education of the respondent. Using the highest education level of both parents or of the mother doesn't change the results in Table 1.D.1.

[54] As, for Table 1.A.1, I removed 255 observations that have less than 23 hours of activity a day reported. 92% of the remaining observations have exactly 24 hours of activity a day reported.

[55] The number of children is close to two for both groups. Time per child in a two-parent household is very similar to childcare time, 1.1 hours for non-college and 1.6 hours for college-graduated households.

[56] Appendix Table 1.A.1 describes how parents allocate their time spent in each of the ATUS activities.

[57] See Appendix Section 1.A.2 for more details information on each of the variables used to calibrate the model.

Table 1.D.1: Parents' Time Allocation by Education, ATUS 2003

	Non-college graduated parents		College graduated parents	
	Hours per day	% of total	Hours per day	% of total
Market work	4.4	27.5%	4.7	29.4%
Childcare	1.2	7.5%	1.6	10.0%
Leisure				
Personal leisure	6.4	40.0%	5.6	35.0%
Other	4.0	25.0%	4.1	25.6%
Total	16.0	100%	16.0	100%

Notes: This table provides information on time allocation by married parents in the United States. The data source is ATUS 2003. Childcare activity follows the BLS definition of "primary childcare activities." Market work contains all work-related and educational activities. Leisure includes all the other activities. For more details on activity grouping, see Appendix Table 1.A.1.

Table 1.D.2: Characteristics of Synthetic Neighborhoods

	Median household income (USD)	Fraction of people aged 25+ with college degree	Fraction below poverty level	Median rent over median household income	Fraction of households
	(1)	(2)	(3)	(4)	(5)
Neighborhood D1	20,638	0.113	0.358	0.286	0.071
Neighborhood D2	28,883	0.134	0.233	0.234	0.088
Neighborhood D3	34,259	0.158	0.167	0.211	0.093
Neighborhood D4	38,652	0.187	0.133	0.197	0.096
Neighborhood D5	42,957	0.212	0.105	0.187	0.100
Neighborhood D6	47,552	0.236	0.085	0.177	0.105
Neighborhood D7	52,547	0.268	0.069	0.170	0.107
Neighborhood D8	58,810	0.311	0.054	0.163	0.111
Neighborhood D9	67,780	0.386	0.042	0.156	0.114
Neighborhood D10	91,273	0.528	0.030	0.141	0.115

Notes: This table provides average sociodemographic characteristics in the ten synthetic neighborhoods. The data source is the NHGIS data from 2000 at the census tract level for the 100 biggest commuting zones. The associated distributions serve as targeted and untargeted moments for the model calibration.

1.D.3 Estimated Parameters: Neighborhood Choices and College Graduation

For two primary purposes, I use the AddHealth survey, detailed in Section 1.5.1. First, to compute intergenerational residential mobility. To do so, as in the NHGIS data analysis, I create ten synthetic neighborhoods in 1994-1995, 2008-2009, and 2016-2018 by ranking all census tracts by median household income and grouping them into ten groups of equal size. Thanks to the panel form of the data set, I can observe in which synthetic neighborhood an adolescent lived in 1994-1995 and in which artificial neighborhood she lived during adulthood, in 2008-2009 and 2016-2018. I restrict the sample to people no longer living at their parent's place. Even though, due to attrition, samples are smaller in Wave V than in Wave IV, I use estimates from Wave V.[58] In 2016-2018, interviewees were older, between 35 and 40, and more likely to be married than ten years before.[59] 30.2% of adolescents who lived in a first decile census tract in 1994 lived in the same decile census tract in 2016-2018. In the sixth decile, this percentage falls to 13.0%.

Second, I use AddHealth to estimate the individual probability of having a college degree conditional on parents and child variables. In the sample, anyone with at least a bachelor's degree is considered to have a college degree. Here again, because interviewees are older and more likely to have completed their education in 2016-2018 than in 2008-2009, I use data from Wave V to estimate the highest level of education. To proxy for children's realized human capital, I use an average of Wave II (1996) grades in English, mathematics, social science, and science. Information on parents' highest education level and income comes from the parent survey conducted in 1994-1995.

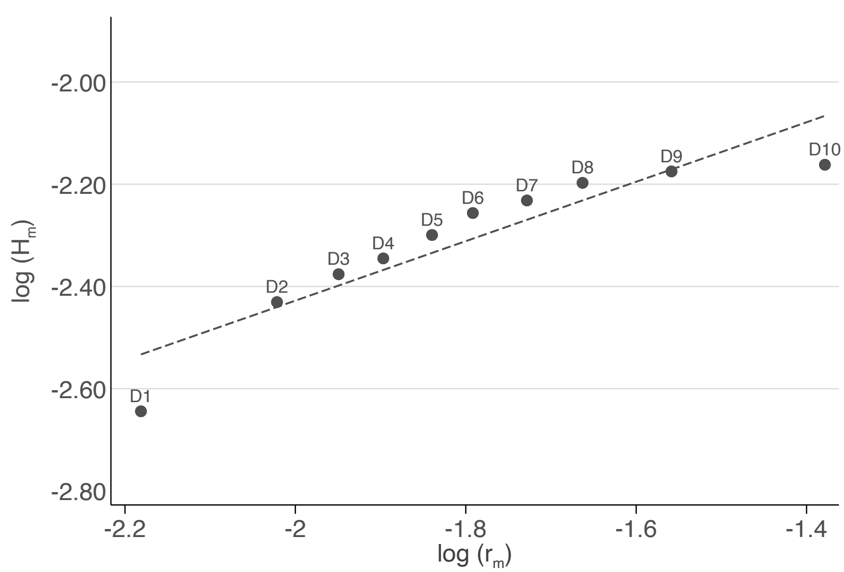
[58] All observations are weighted by the sampling weights of the corresponding wave provided by AddHealth. When variables from different waves are used simultaneously, the weights I use are from the latest wave.
[59] Intergenerational residential mobility patterns are similar whether I use Wave V or Wave IV.

Table 1.D.3: Estimated Parameters

	College graduation probability	
γ_1	-3.83	(0.35)
γ_2	0.35	(0.02)
γ_3	0.15	(0.02)
γ_4	1.11	(0.11)
Pseudo R^2	0.28	

Notes: The table shows the weighted logit regression results. The regression includes county fixed-effects. Robust standard errors are in parenthesis. These are all the estimated parameters.

Figure 1.D.1: Housing Market Estimation



Notes: This Figure displays the estimated housing supply equation, as a function of the relative rent price. Data points show actual rent prices and density for each of the ten synthetic neighborhoods.

1.D.4 Moments

Intergenerational residential mobility:

Data: I use the Add Health survey that contains census tract level information. The sample is restricted to interviewees who do not live in their parent's houses in adulthood. Census tracts are ordered by household median income and divided into ten synthetic neighborhoods. For each decile neighborhood, I compute the share of children who, in adulthood, choose to live in the same decile of the neighborhood quality distribution as their parents.

Model: For each neighborhood, I compute the share of children who, in adulthood, choose to live in the same decile of the neighborhood quality distribution as their parents.

Other moments:

Table 1.D.4: Moments Description

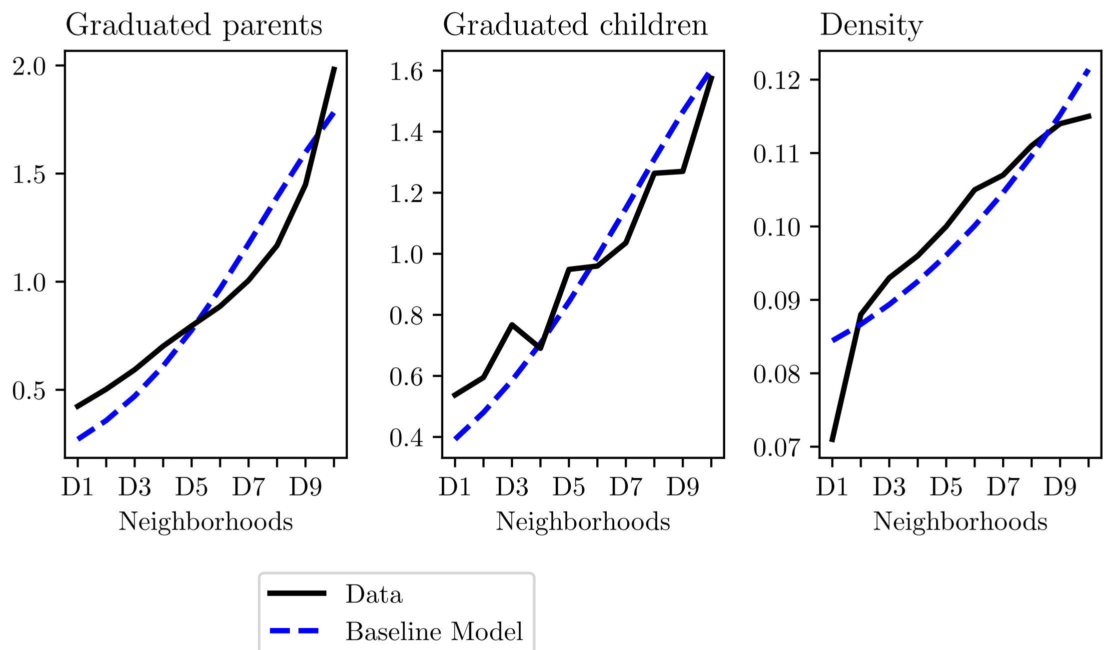
Moment	Description	Data restriction	Source
Earnings			
Share college parents in Q1 over Q4	Fraction of college parents in the first quartile of the income distribution over the fraction of college parents in the fourth quartile of the income distribution.	100 biggest commuting zones - families with a own child below 18	ACS 2000
Earnings ratio non-college - college	Household income of non-college parents over household income of college parents.	100 biggest commuting zones - families with a own child below 18	ACS 2000
Income Gini [inequality]	Families' income Gini, transformed in a 20 years Gini using the 2002 Shorrocks mobiliy index estimated by Kopczuk et al. (2010).	100 biggest commuting zones - families with a own child below 18	ACS 2000
Residential mobility			

Census tract Gini [segregation]	Gini coefficient across the ten synthetic neighborhoods household median income.	100 biggest commuting zones - families with a own child below 18	NHGIS 2000
Neighborhood effect (25th pct.)	For families with below-median income (p = 25). Simulate moves to every neighborhoods. Regress children's income on fixed effects for each neighborhood controlling for origin-by-destination fixed effects.	Tax records covering the U.S. population, spanning 1996-2012. 100 biggest commuting zones (CZ). Within CZ estimates. The authors use income ranks at age 26 because of outliers. In addition, they control for the fraction of childhood spent in a county.	Chetty and Hendren (2018a)
Neighborhood effect (75th pct.)	For families with above-median income (p = 75). Simulate moves to every neighborhoods. Regress children's income on fixed effects for each neighborhood controlling for origin-by-destination fixed effects.	Tax records covering the U.S. population, spanning 1996-2012. 100 biggest commuting zones (CZ). Within CZ estimates. The authors use income ranks at age 26 because of outliers. In addition, they control for the fraction of childhood spent in a county.	Chetty and Hendren (2018a)
Residential immobility (D1)	Fraction of children born in the first synthetic neighborhood who live in this same neighborhood when they are adults.	1994-2018. Interviewees who do not live in their parent's houses in adulthood (2018).	AddHealth Wave I and Wave V
Social mobility			
Rank-rank coefficient	Regression coefficient of child household income rank on parental househod income rank.	Tax records covering the U.S. population, spanning 1996-2012.	Chetty et al. (2014)

Parental time			
Parental time income coeff.	Income regression coefficient of parental time on income quartile and college graduation status of the parents.	Two-parent households with an own child below 18. Weighted. Additional controls on gender, child age, number of children, and date of interview.	ATUS 2003
Parental time education coeff.	Income regression coefficient of parental time on income quartile and college graduation status of the parents.	Two-parent households with an own child below 18. Weighted. Additional controls on gender, child age, number of children, and date of interview.	ATUS 2003
Parental time of non-college parents	Average parental time of non-college parents.	Two-parent households with an own child below 18. Weighted.	ATUS 2003

1.D.5 Model Fit for Additional Non-Targeted Moments

Figure 1.D.2: Non-Targeted Moments by Neighborhood



Notes: This Figure shows two non-targeted moments across neighborhoods: the share of graduated parents and the share of graduated children. The third panel presents the population density by neighborhood which is matched by estimating the housing supply function parameters. The solid line shows the data moments, and the dashed line shows the model-simulated analogs. Data Source: NHGIS.

1.D.6 Perfect Information Model

This section describes the calibration of a model without subjective beliefs. Parents' preferences take the following forms:

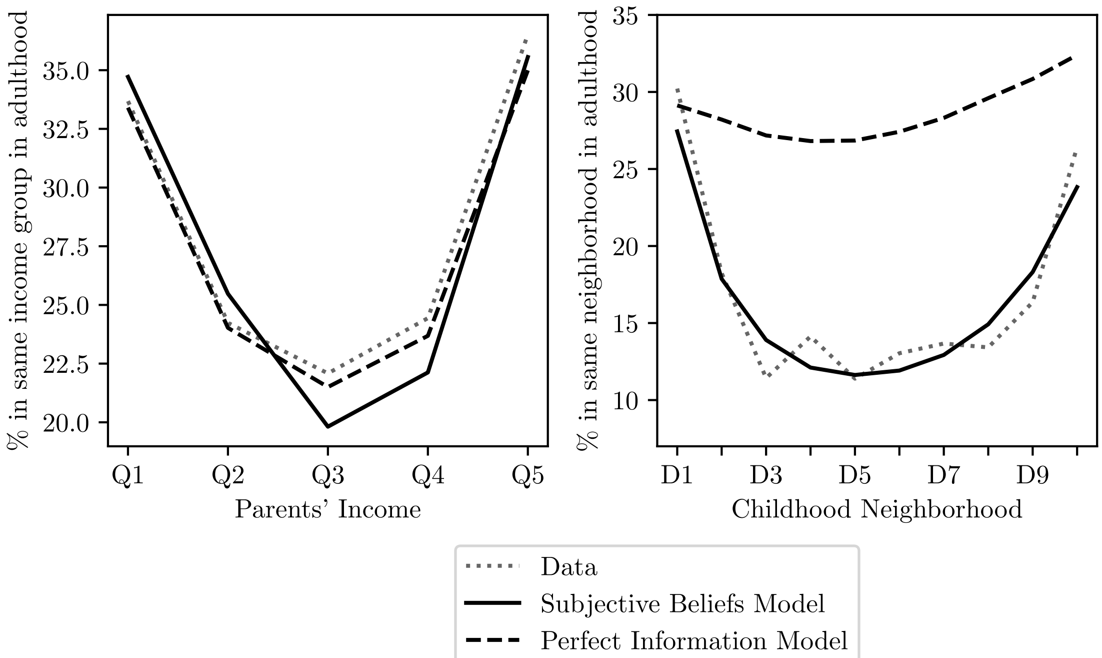
$$\log(c) + \frac{\left(1 - \overline{\ell_s} - \kappa\,\tau\right)^{1-\frac{1}{\psi}}}{1-\frac{1}{\psi}} + \iota\,\mathbb{1}_{m_0=m} + \nu\varepsilon_m + b\mathrm{E}\left[\mathcal{V}\left(h', s', m, \alpha\right)\right]$$

Table 1.D.5: Internally Calibrated Parameters - Perfect Information Model

Parameter	Description	Moment	Data	Model
Preferences and Labor Market				
$b = 0.65$	Altruism	Ratio share college parents Q1 to Q4	0.102	0.117
$\kappa = 0.7$	Parental time disutility	Parental time non-college parents	0.075	0.078
$\iota = 0.024$	Place of birth preference	Residential mobility (D1)	0.295	0.302
$\omega = 0.01$	College wage premium	Earnings ratio non-college - college	0.554	0.557
Neighborhoods				
$\sigma_m = 0.16$	Neighborhood quality	Neighborhood effect (25th pct.)	0.062	0.061
$\nu = 0.02$	Taste shock variance	Census tract Gini	0.231	0.157
Skill Formation: $h' = \left(\left(\gamma\ (\frac{\tau}{\overline{\tau}})^{\varphi} + (1-\gamma)\, m^{\varphi}\right)^{\frac{1}{\varphi}} + \underline{i}\right)^{\alpha} h^{1-\alpha}\ \exp(a)$ with $a \sim \mathcal{N}(0, \sigma_a)$				
$\alpha = 0.7$	Elasticity of investments	Rank-rank coefficient	0.341	0.348
$\gamma = 0.45$	Parental time share	Neighborhood effect (75th pct.)	0.046	0.057
$\varphi = -5$	Substitutability	Income gradient in parental time	0.140	0.091
$\sigma_a = 0.55$	Ability shock variance	Income Gini	1.247	1.218

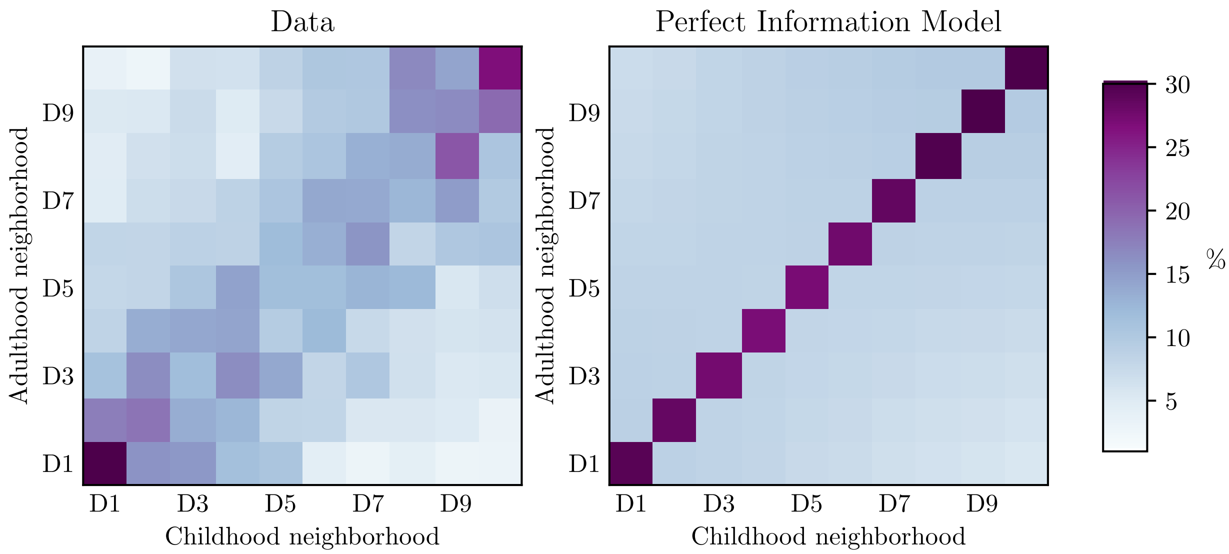
Notes: This table reports the internally calibrated parameters and the observed and simulated moments associated with the parameter estimates. Perfect Information model.

Figure 1.D.3: Residential Mobility in a Perfect Information Model



Notes: Left panel shows the share of children who, when they are adults, are in the same household income quintile as their parents. Right panel shows the share of children who still live in their childhood neighborhood quality when they are adults. Data moments are in dotted-line, subjective beliefs model simulated analogs in solid-line, and perfect information model in dashed-line.

Figure 1.D.4: Detailed Residential Mobility in a Perfect Information Model



Notes: This Figure shows residential mobility between childhood and adulthood in the data (left panel), and their perfect information model simulates analogs (right panel). The x-axis contains the ten synthetic childhood neighborhoods, and the y-axis is the ten synthetic adulthood neighborhoods. The frequency of moves is represented by cell colors, with larger values in darker squares.

1.D.7 Perfect Information Model with Heterogeneity

This section describes the calibration of a model without subjective beliefs but with heterogeneous preferences. Parents' preferences feature heterogeneous preferences by place of birth and I assume a quadratic moving cost function. Parents' preferences take the following forms:

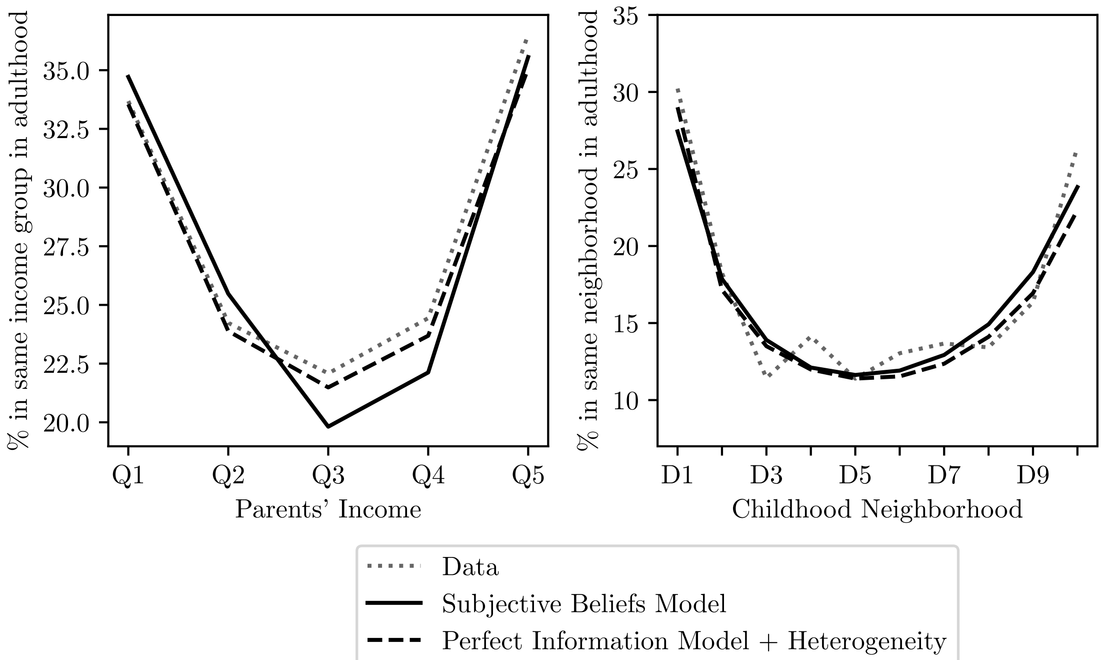
$$\log(c) + \frac{\left(1 - \overline{\ell_s} - \kappa\,\tau\right)^{1-\frac{1}{\psi}}}{1 - \frac{1}{\psi}} + \iota_{m_0}\,\mathbb{1}_{m_0=m} - \xi(m_0 - m)^2 + \nu\varepsilon_m + b\mathrm{E}\left[\mathcal{V}\left(h', s', m, \alpha\right)\right]$$

Table 1.D.6: Internally Calibrated Parameters - Perfect Information Model with Preference Heterogeneity

Parameter	Description	Moment	Data	Model
Preferences and Labor Market				
b = 0.65	Altruism	Ratio share college parents Q1 to Q4	0.102	0.118
κ = 0.7	Parental time disutility	Parental time non-college parents	0.075	0.077
ι_1 = 0.0135	Place of birth preference	Residential mobility (D1)	0.295	0.302
ι_2 = 0.0015	Place of birth preference	Residential mobility (D2)	0.182	0.172
ι_{3-9} = 0.0005	Place of birth preference	Residential mobility (D5)	0.112	0.114
ι_{10} = 0.0015	Place of birth preference (D10)	Residential mobility (D10)	0.235	0.223
xi = 0.0005	Quadratic moving costs	Correlation neighborhood quality	0.417	0.400
ω = 0.0001	College wage premium	Earnings ratio non-college - college	0.554	0.563
Neighborhoods				
σ_m = 0.16	Neighborhood quality	Neighborhood effect (25th pct.)	0.062	0.062
ν = 0.02	Taste shock variance	Census tract Gini	0.231	0.154
Skill Formation: $h' = \left(\left(\gamma\ \left(\frac{\tau}{\overline{\tau}}\right)^{\varphi} + (1-\gamma)\, m^{\varphi}\right)^{\frac{1}{\varphi}}\ +\ \underline{i}\right)^{\alpha} h^{1-\alpha}\ \exp(a)$ with $a \sim \mathcal{N}(0, \sigma_a)$				
α = 0.7	Elasticity of investments	Rank-rank coefficient	0.341	0.347
γ = 0.45	Parental time share	Neighborhood effect (75th pct.)	0.046	0.057
φ = -5	Substitutability	Income gradient in parental time	0.140	0.090
σ_a = 0.55	Ability shock variance	Income Gini	0.336	0.328

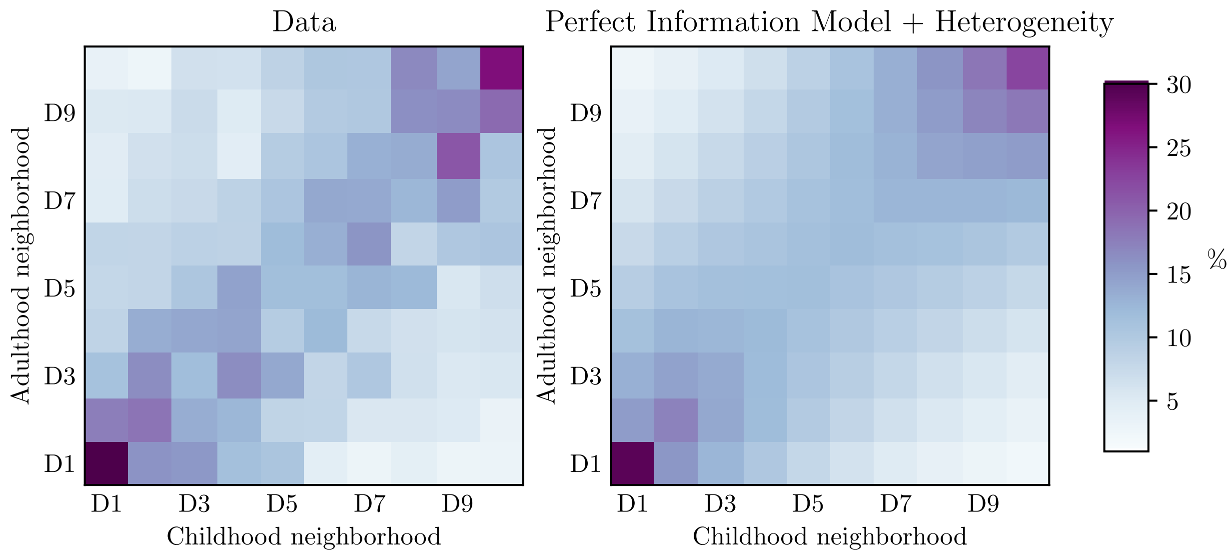
Notes: This table reports the internally calibrated parameters and the observed and simulated moments associated with the parameter estimates. Perfect Information model with preference heterogeneity.

Figure 1.D.5: Residential Mobility in a Perfect Information Model with Heterogeneity



Notes: Left panel shows the share of children who, when they are adults, are in the same household income quintile as their parents. Right panel shows the share of children who still live in their childhood neighborhood quality when they are adults. Data moments are in dotted-line, subjective beliefs model simulated analogs in solid-line, and perfect information model with heterogeneity in dashed-line.

Figure 1.D.6: Detailed Residential Mobility in a Perfect Information Model with Heterogeneity



Notes: This Figure shows residential mobility between childhood and adulthood in the data (left panel), and their perfect information model with heterogeneity simulates analogs (right panel). The x-axis contains the ten synthetic childhood neighborhoods, and the y-axis is the ten synthetic adulthood neighborhoods. The frequency of moves is represented by cell colors, with larger values in darker squares.

1.E Robustness checks

Table 1.E.1: Parental Time and Neighborhood Quality, Add-Health 1994-1995 and 2016-2018

	College Graduation 2016-2018		Parental Time 1994-1995		Neighborhood 2016-2018	
	(1)	(2)	(3)	(4)	(5)	(6)
Neighborhood 1994-1995	0.448***		0.758***	0.451***	0.25***	0.211***
	(0.1)		(0.195)	(0.185)	(0.037)	(0.032)
Parental Time 1994-1995		0.029***				
		(0.007)				
Race						-0.003
						(0.008)
Controls:						
Childhood SES	yes	yes	no	yes	yes	yes
Adulthood SES	no	no	no	no	no	yes
Same decile census tract	no	no	no	no	yes	yes
School ID FE	yes	yes	yes	yes	no	no
Commuting zone FE	no	no	no	no	yes	yes
No. of obs.	8,525	8,543	17,073	12,608	7,952	6,039
Clusters	101	101	120	105	97	90

Notes: This table provides between parental investment choices and college graduation (Columns (1) and (2)) correlations between parental investment choices (Columns (2) and (3)) and between childhood and adulthood neighborhood quality (Columns (5) and (6)). Socioeconomic status variables are family income, highest education level, and marital status. See Appendix Section 1.A.1 for details on variable construction. The dependent variables are indicated in the column title. All regressions control for age fixed effects. All observations are weighted by Wave specific sample weights provided by AddHealth. Standard errors displayed in brackets are clustered at the county level. *p<0.1; **p<0.05; ***p<0.01.

Chapter 2

Efficiency and Mobility of Education Tracking: A Quantitative Analysis[1]

Joint work with Lukas Mahler[2]

Abstract: We study the long-run aggregate and distributional effects of school tracking —the allocation of students to different types of schools— by incorporating school track decisions into a general-equilibrium heterogeneous-agent overlapping-generations model. The key ingredient in the model is the child skill production technology, where a child's skill development depends on her classroom peers and the instruction pace in her school track. We show analytically that this technology can rationalize reduced-form evidence on the effects of school tracking on the distribution of child skills. We calibrate the model to data from Germany, a country with a very early and strict school tracking policy. Our model suggests that eliminating the parental influence on the school track choice that arises purely from own-track preferences improves social mobility while keeping aggregate output constant. An education reform that postpones the tracking age from ten to fourteen generates even larger improvements in intergenerational mobility. However, these come at the cost of efficiency losses in aggregate economic output. The size of these losses depends on the design of the instruction levels in each school track.

[1] We are indebted to Antonio Ciccone, Michèle Tertilt, and Minchul Yum for their continued and invaluable support and guidance in this project. In addition, we are very grateful to the seminar and conference participants at the University of Mannheim for helpful discussions and suggestions. Suzanne Bellue gratefully acknowledges financial support from the German Academic Exchange Service (DAAD) and the German Research Foundation (through the CRC-TR-224 project A03). Lukas Mahler gratefully acknowledges the financial support from the German Science Foundation (through the CRC TR 224 Project A04) and the SFB 884 Political Economy of Reforms.

[2] Department of Economics, University of Mannheim

2.1 Introduction

School tracking, which involves the allocation of students into physically distinct types of schools that differ in the curriculum taught, intensity, and length, is a common feature of education policy in OECD countries.[3] While school tracking is designed to improve teaching efficiency, it may hamper equality of opportunity in access to education. The rationale behind tracking is that grouping children according to ability creates more homogeneous peer groups and allows for tailored instruction levels and curricula, leading to improved educational outcomes (Bonesrønning et al., 2022; Duflo et al., 2011). However, because the ability of young children varies over time and is subject to unpredictable shocks (Hanushek and Wössmann, 2006), introducing early tracking increases the risk of misallocation and the influence of parental background on the track decision. As a result, early tracking may impair mobility in educational and labor market outcomes across generations (Falk et al., 2021; Dustmann, 2004; Meghir and Palme, 2005; Pekkala Kerr et al., 2013). For that reason, the timing of school tracking is a recurrent issue in the public and academic debate about education reforms in countries with a strict and early tracking regime, such as Germany.[4]

We argue that the aggregate effects of changing the timing of school tracking depend on the efficiency-mobility trade-off that results from improving teaching efficiency through homogeneous peer groups and letting parents choose which track to assign their children under imperfect information about their children's skills. However, providing a quantitative assessment of the long-run aggregate, distributional and inter-generational effects of the timing of school tracking policies requires a macroeconomic model of mobility with an understanding of how parents will optimally adjust their track assignation choices in response to changes in the uncertainty about their child's skills and in the composition of the peers. Macroeconomic models of mobility provide a useful environment to consider such effects but have so far largely ignored how the development of child skills is during schooling years.

To empirically underpin our modelization of child skills formation during schooling years and parental track choices, we first document the evolution of children's test scores and track assignation using the German National Educational Panel Study (NEPS) (Blossfeld et al., 2011). We find that students' test score ranks vary across years, suggesting that child skills, measured by test scores, are imperfectly observed. Second, we see that the parental school track choices are largely consistent with teachers' school track recommendations. However, some parents deviate, and most often, it is in favor of their own education track. Students

[3] An overview about school tracking policies in OECD countries is given in Chapter 2 in OECD (2013).

[4] There is substantial variation in the timing of school tracking across OECD countries. While in some countries, such as Germany and Austria, tracking occurs already at the age of 10, in other countries, like the US and UK, do not track at all during secondary school.

who deviate toward the academic track are more likely than others to then fail. We interpret this data pattern as parents having a form of preference for their own education track.

As assessing the long-run effects of the timing of school tracking policies requires taking into account the effects of tracking on the educational outcomes of children, but also how these outcomes translate into labor market outcomes and outcomes across generations, we use a general equilibrium life-cycle Aiyagari framework of overlapping generations in which parents care about their offspring in the tradition of Becker and Tomes (1986).[5] Child skills during the schooling years evolve according to their parents' socioeconomic group, to the instruction pace in their track, and to their peers' average skills, and are subject to uninsurable skill shocks. The model is tailored to fit the German Education System, where children are tracked into two school tracks at the age of 10 based on a decision by the parents. As in the data, the track decision may be influenced by parental preferences for children to follow in their own education steps. While only one track directly facilitates access to college education, we allow for second-chance opportunities as children can decide to switch tracks after secondary school. Going to college incurs psychic costs, which are a function of child skills as well as time costs relative to non-college education. End-of-school child skills translate into adult human capital, which evolves stochastically over the working life and determines, together with the tertiary education decision, the labor earnings. The distribution of human capital across college and non-college workers affects prices, which in turn affects the school track choice. Finally, households can save into a non-state-contingent asset subject to life-cycle borrowing constraints. When children become independent, parents can also make a non-negative inter-vivos transfer.

The model builds around a parsimonious theory of how a child's skills are developed during school years. Going to a school that belongs to a particular school track affects children's skills directly through interactions with peers at their school and the pace of instruction that is taught in that school. Every child is assumed to have an ideal instruction pace at which she learns best. However, there can only be one instruction pace per school track, which is set endogenously by the policymaker. We show analytically that, under linear direct peer effects and complementarity assumptions between own skill and instruction pace, this gives rise to efficiency gains from tracking in terms of improving aggregate end-of-school skills. Indeed, absent any unforeseeable shocks to child skills, an optimal tracking policy should

[5] This is hard, if not impossible, to do in a purely reduced-form way, not only because of its demands on data, but also because a change in the allocation of children across tracks and, consequently, a change in the allocation of workers across skill levels, may entail general equilibrium effects. For example, suppose the share of children who are allocated to an academic track school increases substantially in the long run. In that case, the price of academically skilled labor in the economy should decrease. This, in turn, makes an academic track school less attractive, which affects the share again.

perfectly stratify children according to their skills as early as possible. In the presence of skill shocks, however, due to the risk of misallocation, it can be optimal to postpone tracking, even from an efficiency point of view. Finally, the theory implies that not all children gain from tracking and that the losses are often concentrated in the track with the lower average skill level. Thus, our child skill formation technology rationalizes some of the most robust empirical findings regarding school tracking in the literature.[6]

The model is solved numerically, and the parameters are calibrated in two steps. First, we estimate the child skill formation technology parameters directly from German data on school children using a latent variable framework as in Cunha et al. (2010). In particular, we use the information on achievement test scores to measure child skills at different stages of their school careers. We use an instrumental variable strategy similar to Agostinelli et al. (2019) to account for measurement error. We then calibrate the remaining parameters to match a set of critical moments from representative German survey data. The model matches the data well, both in terms of aggregate moments and with respect to the distribution of child skills across school tracks and parental backgrounds, as well as the transitions through the education system. To test the model's validity, we investigate the effects of the initial school track on later-in-life economic outcomes for a set of children who are, in equilibrium, just at the margin between the two school tracks. Dustmann et al. (2017) argue that for such marginal children in Germany, the initial track choice is inconsequential for labor earnings later in life. Simulated data from our model suggest that children who go to different school tracks solely based on small differences in skills at the time of the track decision experience very similar lifetime economic outcomes, where children in an academic track school track earn around 2% higher lifetime labor income compared to similar children that did not go an academic-track school.

Notwithstanding this, a variance decomposition exercise shows that skill formation during the school tracking years, and hence the school tracking policy, plays an important role in lifetime inequality across the population. In particular, variation in the initial school track alone can account for 12% of the variation in lifetime earnings and 13% of the variation in lifetime wealth. As in the data, parental education is, after child skills, the second most crucial determinant of initial school track choice. We use our model to show that most of this effect comes from direct parental preferences for children to follow their education track rather than college tastes or knowledge about the deterministic influence of parental education on child skill development. The parental bias in school track choice gives rise to inefficiencies in the allocation of children across tracks. For example, a college-educated parent may push

[6] For the case of Germany, see for instance Piopiunik (2014) who shows that low-achievers may be negatively affected by school tracking.

her child into an academic-track school even though her child's skills optimally suggest a vocational-track school. This harms her child's learning outcomes and affects average learning in that track as the instruction pace endogenously adjusts to the composition of skills in that track. We perform counterfactual experiments using our model that eliminate the parental bias in school track choice, for example, by introducing a strict skill threshold that governs school track allocation. Such a policy improves social mobility across generations with minor effects on cross-sectional inequality. The intergenerational income elasticity decreases by around 0.3%-2.3%. However, while this policy improves the average learning outcomes of children, we highlight that aggregate GDP remains essentially unchanged. This is because the improvements in child skills are quantitatively minor (0.1%-0.3%) in the first place and fade out over the remaining schooling career.

Finally, we use our model to study the long-run effects of an education reform that universally postpones the school tracking age by four years. Such a reform is often suggested in countries with traditionally early tracking systems, such as Germany, to improve equality of opportunity in access to academic education (Woessmann, 2013). We show that postponing the tracking age improves social mobility as it leads to a 2% decrease in the intergenerational elasticity of income. However, this comes at the cost of around 3% of the GDP. These mobility gains arise primarily because college education after secondary school becomes significantly less determined by the previous school track, so more children use the "second-chance" opportunity to go to college even after graduating from a vocational track school. This is aided by the fact that differences in end-of-school skills across tracks decrease when children stay in comprehensive schools for longer. As a result, college education, but also the school track choice, become less dependent on the parental background, improving social mobility. On the other hand, our results indicate that postponing tracking comes at the cost of a 0.2% drop in GDP. The reason for this is that prolonged learning in a comprehensive school track foregoes efficiency gains from tailored instruction levels in an early tracking system. Quantitatively, these learning losses cannot be recuperated by efficiency gains coming from the fact that the late tracking decision is based on more complete information about children's skills. Lower levels of child skills then translate into lower levels of human capital in the economy causing aggregate output to decline. In this context, we highlight the importance of considering general equilibrium effects on the labor market that influence school track and college decisions. In particular, the aggregate output losses in partial equilibrium would be significantly higher at around 0.8-1% of GDP. In sum, our results, therefore, suggest the presence of an efficiency-mobility trade-off of reforms that postpone the tracking age. Abolishing tracking in favor of comprehensive schooling altogether further exacerbates this trade-off.

Related Literature

This paper links several strands of the literature: the quantitative family-macroeconomics literature, the children's skill formation literature, and the school tracking literature.

First, we contribute to the quantitative family macroeconomics literature that studies determinants of the intergenerational transmission of economic status (Abbott et al., 2019; Caucutt and Lochner, 2020; Daruich, 2022; Jang and Yum, 2022; Fuchs-Schündeln et al., 2022; Fujimoto et al., 2023; Lee and Seshadri, 2019; Yum, 2022). Some of these studies incorporate a part of the educational system into their analysis, such as Abbott et al. (2019); Caucutt and Lochner (2020); Fuchs-Schündeln et al. (2022) who model high-school graduation choice. However, all of these studies except Fujimoto et al. (2023) focus on the United States, often concentrating on access to higher education and neglecting the importance of designing the (secondary) school system for macroeconomic outcomes. We explicitly focus on the secondary schooling system. Our paper is perhaps most closely related to to Fujimoto et al. (2023) who study the importance of free secondary schooling for misallocation driven by borrowing constraints in Ghana. Our contribution is to analyze a widespread education policy at the secondary school stage in developed countries: school tracking. In particular, we investigate the consequences of the school track choice and the age at which school tracking occurs for inequality and efficiency in a dynamic macroeconomic model. We thereby complement related research that focuses on the early, pre-school phases in a child's skill development (Daruich, 2022; Yum, 2022) and research that focuses on higher, post-secondary education (Abbott et al., 2019; Capelle, 2022).

Second, this paper builds on the literature on children's skill formation, which has described how children's skills evolve as a function of endowments, parental and environmental inputs, and recently also schooling inputs (see, for instance, Cunha and Heckman (2007); Cunha et al. (2010); Agostinelli et al. (2023, 2019)). Our main innovation relative to this literature is considering two forms of peer effects, which allows for rationalizing the empirical findings regarding school tracking. First, similar to Agostinelli (2018), we incorporate direct peer effects, which capture the idea that children benefit from high-quality peer groups. Second, following Duflo et al. (2011) 's evidence in Kenyan primary schools, we consider how the instruction levels adjust endogenously to the peer composition in schools of a particular track. More specifically, we assume that a child's optimal pace of instruction is unique and increases with her current skill level. Then, learning decreases monotonically with the distance between a child's optimal instruction pace and the one she is currently taught at. This parsimonious micro-funded model captures the main arguments about school tracking and allows us to evaluate the effects of delaying the tracking decision.

Third, this paper contributes to and builds on the literature that estimates the impact of early school tracking on efficiency and equity measures. An extensive empirical literature investigates the effects of age at school tracking on students' test scores and later outcomes. It either exploits temporal within-country variation in tracking practices (Meghir and Palme (2005), for Sweden; Aakvik et al. (2010), for Norway; Malamud and Pop-Eleches (2011), for Romania; Pekkala Kerr et al. (2013), for Finland; and Matthewes (2021); Piopiunik (2014) for Germany) or between-country variation with a difference-in-differences strategy (Hanushek and Wössmann, 2006; Ruhose and Schwerdt, 2016). Most studies suggest that earlier tracking raises inequality in educational outcomes and increases the effect of parental education on student achievement. Guyon et al. (2012) investigate an educational reform in Northern Ireland that led to a large increase in the share of students admitted to the elite track at age eleven. They find a strong positive overall effect of this de-tracking reform on the number of students passing national examinations at later stages and a negative effect on student performance in non-elite schools that lost their most able students. A notable exception is Dustmann et al. (2017), who use an individual-level instrumental variables strategy (the date of birth) and find no effect of track choice on educational attainment or earnings for German students at the margin between two tracks. While their result suggests that the misallocation of hard-to-assign students has little impact on their future outcomes, it does not rule out a potential adverse effect of early school tracking on the outcomes of non-marginal sub-groups of students, such as those from low-socioeconomic backgrounds.

The remainder of the paper is organized as follows. Section 2.2 provides empirical facts to guide our school-tracking modeling choices. Section 2.3 presents our model that uses a life-cycle Aiyagari GE framework of overlapping generations, and Section 2.4 helps build intuitions about the model mechanisms underlying school tracking using a parsimonious theory for the child skill formation. Section 2.5 explains how we estimate and parameterize the model. It also offers some validation exercises. In Section 2.6, we use the calibrated model to perform a series of counterfactual experiments to quantify the effects of different school tracking policy regimes. Finally, Section 2.7 concludes.

2.2 Data

In this section, we document the evolution of child skills along the parental socioeconomic background and school track dimensions in Germany. The results motivate our focus on school tracking and will serve as inputs for the calibration of the quantitative model. We use the German National Educational Panel Study (NEPS), which comprises detailed longitudinal data on the educational process, acquired competencies, as well as the learning environment and context persons of six cohorts of children in nationally representative samples in Germany (Blossfeld et al., 2011). A key component of the information collected is regular standardized assessment tests of the children's competencies in areas such as mathematics, reading, sciences, vocabulary, or grammar, combined with specific wave weights. In addition, there is information about school track recommendations and the final parental school track choices. Indeed, primary school teachers give track recommendations before transitioning to secondary school. They are based on reflecting on the child's achievement during primary school and the teacher's assessments.[7]

First, Table 2.1 shows that parents frequently deviate from teacher recommendations toward their own education. Research on school tracking has found that parents with higher socioeconomic status are more likely to send their child to an academic track school than parents with a lower socioeconomic status, even conditional on school performance or achievement test scores before the track decision (Falk et al., 2021). Consistently, we find that 74% of children from college-graduated parents receive a teacher recommendation for the academic track versus 43% of children from non-college-graduated parents.[8] In addition, Table 2.1 shows that while around 23% of parents who themselves have a college education overrule a vocational recommendation, only 6% of them overrule an academic recommendation. At the same time, while 16% of non-college graduated parents overrule an academic recommendation, only 12% of them overrule a vocational recommendation. There may be multiple reasons behind these deviations. For example, parents may have more information about their child's skills than teachers. However, the deviations are not symmetric across tracks, and parents are more likely to deviate from teachers' recommendations for their own education.

Parents may have several reasons for frequently overruling teachers' recommendations when they differ from their own education. For instance, they may be better equipped to support their child in a track with which they are more familiar. However, the last columns of Table 2.1 show that children of college-educated parents who deviate from the recommended vocational track do relatively poorly compared to those who received the academic recommendation. In fact, only 20% of children of college-educated parents who deviated from the vocational track recommendation belong to the top quartile of skills four years later in Grade 9. In contrast, the same number reaches 34% among those who received an academic track recommendation. This suggests that the support provided by college-educated parents does not fully compensate for relatively low skill levels. Conversely, children from non-college-educated parents who deviate toward the vocational track do remarkably well in Grade 9, with over half of them belonging to the top quartile of skills. As a comparison, only a quar-

[7] See also Appendix Section 2.D for more details on the tests.

[8] We define children from college parents if they have at least one of the parents with a college education.

ter of those with a vocational recommendation reached the top quartile in Grade 9. Those numbers indicate that these students might have succeeded in the academic track as well. Thus, we argue that the relatively high number of deviations towards parents' education is partly driven by a parental bias towards their own education, which is not only motivated by parents' ability to support the child or their intrinsic knowledge of their skills.

Table 2.1: School Track Choice

				% in the top 25% in G9	
	Recommendation (G5)	Shares	% deviate	if followed	if deviated
College Parents					
	Academic	74% (973)	6% (58)	34% (235)	51% (17)
	Vocational	26% (350)	23% (61)	31% (39)	20% (7)
Non-College Parents					
	Academic	43% (753)	16% (121)	24% (113)	56% (35)
	Vocational	57% (999)	12% (89)	25% (106)	23% (11)

Notes: This table provides information on school track choice by parental education and teacher recommendation. All observations are weighted. Source: NEPS, Cohort 3.

Second, we see that children's skills are subject to variation across academic years.[9] Table 2.2 displays the correlations of a child's math test score percentile rank across Grades.[10] Those correlations are below one, suggesting that a child's skills are subject to variation over time. Additionally, we see that the rank correlations between earlier Grades are lower than those between later ones. Specifically, the rank correlation between Grades 1 and 4 is 0.59 and rises to 0.72 between Grades 5 and 9. We interpret these data patterns as children's skills being subject to shocks, which may be larger in the early stages of a child's life.

Finally, in line with the literature, Table 2.3 shows that higher parental socioeconomic background is correlated with higher child skills very early on. As of Kindergarten, average child skills by parental education differ by 0.47 SD. This result is consistent with the literature that documents early skill level differences, which may be a combination of parental investment and genetic components.[11] Moreover, looking at the statistics by cohort, we see

[9] Here, we use mathematics test scores as the main measurement of a child's raw skills.

[10] The sample comprises only children who took the tests in both Grades. For this reason, we prefer unweighted estimates.

[11] See for instance Larson et al. (2015) who studies cognitive abilities of children at the entry of kindergarten in the United States by parental socioeconomic status.

Table 2.2: Rank-Rank Correlations

		Rank-Rank Correlation	Obs.
Cohort 2	G1-G4	0.59	3,116
	G4-G7	0.63	2,321
Cohort 3	G5-G9	0.72	5,311

Notes: This table provides the rank-rank correlations in math grades. Source: NEPS

that skill differences by parental background do not decrease over time, which is consistent with previous empirical studies.[12]

Table 2.3: Differences in Average Skills in Standard Deviation

		Difference by	
	Grade	Parent's Education	School Track
Cohort 1	K1	0.47	
Cohort 2	G1	0.46	0.77
	G2	0.45	0.80
	G4	0.58	0.86
	G7	0.49	0.97
Cohort 3	G5	0.58	1.00
	G7	0.67	1.07
	G9	0.70	1.16
Cohort 4	G9	0.62	1.06

Notes: This table provides information on average differences in math grades in one standard deviation unit by parental background and school track over time. All observations are weighted. Source: NEPS

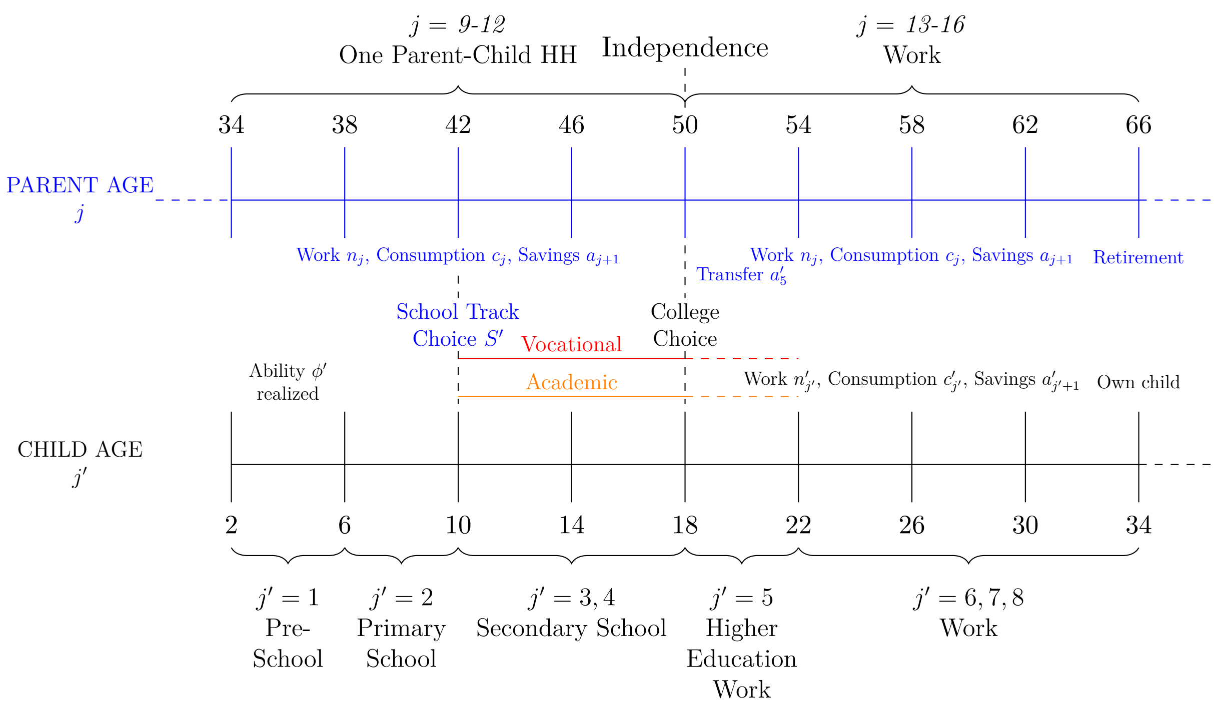
2.3 The Model

The model is designed to fit the German Education System and is guided by the data patterns documented above.[13] First, empirical variation in children's skills over time and parents'

[12] See for instance Passaretta et al. (2022); Nennstiel (2022); Schneider and Linberg (2022) who investigate the NEPS data and find stable or growing socioeconomic status gaps in children's skills.

[13] An overview of the German Education System is given in Appendix 2.C.

Figure 2.1: Timeline of Life-cycle Events



deviation from teacher recommendations toward their own education motivate us to include age-specific skill shocks to the child skill formation and a parental preference for their own education track. Second, investigating the effects of parents' decisions about their child's school track on their children's outcomes requires an overlapping-generations life-cycle model. We assume that time is discrete and infinite and that one model period, j, corresponds to the 4 years in between ages $[4j-2, 4j+2]$ in real life.[14] This frequency allows us to investigate meaningful variations in school tracking ages. The dynastic structure implies that there are 20 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, represented by j, as illustrated in Figure 2.1. For the remainder of the text, we will denote all child variables with primes. Since an adult becomes a parent at age 32, the child of a parent who is in period j is in period $j' = j - 8$.

Each individual goes through 20 stages in life and starts with childhood, $j = 1, 2, 3, 4$, during which she makes no decisions, is educated by her parent, and goes to the school chosen by her parent. School tracking between academic and vocational schools happens in period

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

$j = 3$. At age 18, in period $j = 5$, the child finishes school and becomes an independent adult. She decides how much to work and whether to pursue a college education or obtain a vocational/professional degree. Over the following periods ($j = 6, 7, 8$), the individual works while not yet having children. From periods $j = 9$ to $j = 16$, the adult goes through the parenting years while also making consumption, savings, and labor supply decisions. During these periods, her human capital grows stochastically. Moreover, when her child turns 10, the parent decides on her school track, and when she turns 18, the parent decides on an inter-vivos transfer. In the remaining model periods $j = 17, ..., 20$, the individual is retired and makes consumption and savings decisions while earning retirement benefits. Everyone dies with certainty after model period $j = 20$, that is, at age 82.

2.3.1 Evolution of Child Skills

In period $j = 1$, a 2-year-old child enters into a one-parent household, equipped with an initial learning ability ϕ', which is imperfectly transmitted from her parent as in Yum (2022).[15]

$$\log \phi' = \rho_\phi \, \log \phi \, + \epsilon'_\phi, \qquad \epsilon'_\phi \sim \mathcal{N}(0, \sigma^2_\phi), \tag{2.1}$$

where ϵ_ϕ is an intergenerational shock. The learning ability translates into an initial child skill level that can be summarized in a univariate level:[16]

$$\theta'_1 = \log \phi'. \tag{2.2}$$

Child Skill Formation

During the period of childhood, which consists of periods $j = 1,\ 2,\ 3,\ 4$, a child's skills are determined by past skills, household characteristics, and school inputs, as represented by the following equation:[17]

[15]The learning ability captures genetic components and investments made by parents into their child's skill development during early childhood, infancy, and even in-utero. The importance of these early life stages as well as policy interventions targeted at investments during these years, has been the focus of the child skill development literature (see e.g. Heckman and Mosso (2014) for a review).

[16]As in Cunha and Heckman (2007), we do not differentiate between abilities and skills, as both are partly endogenously produced and partly exogenously determined pre-birth. Moreover, we do not allow for potentially different production technologies of cognitive and non-cognitive skills as in Cunha et al. (2010) or Daruich (2022). Instead, in the tradition of Becker and Tomes (1986), we focus on one composite skill, which can be translated into adult human capital rewarded on the labor market after school.

[17]Note that recognizing the multistage nature of child development is crucial not only to capture the self-productive nature of skills and dynamic complementarity of investments into skills but to differentiate the effects of school tracking at different stages of the schooling years.

$$\theta'_{j+1} = \theta'_j + \alpha_j \bar{\theta'}_{-i,j} + \beta_j p_j - \frac{\gamma_j}{2} {p_j}^2 + \gamma_j \theta'_j p_j + \zeta_j E + \eta_{j+1}, \tag{2.3}$$

where η_{j+1} describes unobserved shocks to the formation of skills that are assumed to be independent and normally distributed around mean zero with a variance $\sigma^2_{\eta_{j+1}}$.[18] We assume the parameters $\gamma_j > 0$ and $\alpha_j,\ \beta_j,\ \zeta_j \geq 0$. The evolution of the child's skills is directly affected by past skills. We allow for the explicit dependence of child skill development on parental education (E) measured by ζ_j, which enables us to capture the effects of the household environment, as well as parental inputs, such as monetary investments in the child's skill development. The school input consists of two elements: linear and direct peer effects (α_j) and a non-linear effect of the pace of instruction p_j. The linear and direct peer effects capture the idea that children benefit from being surrounded by more able peers. The pace of instruction represents the tailored instruction levels and curricula.[19] Learning monotonically increases with the pace of instruction at rate β_j and decreases with the distance to the pace of instruction at rate γ_j. The fact that distance to the pace of instruction matters reflects the intuition that the closer a child is to the pace of instruction, the faster she learns.[20] It implies that for each skill level θ'_j, there exists a unique optimal pace of instruction p^*_j that maximizes the child's future skill level, keeping everything else fixed. The optimal pace of instruction p^*_j for a child with skills θ_j is given by $p^*_j = \frac{\beta_j}{\gamma_j} + \theta'_j$.[21] It is strictly increasing in current skill, such that higher-skilled children also prefer a higher pace of instruction.

Optimal Pace of Instruction

Consider a policymaker choosing the pace of instruction, which we think as reflecting the curriculum as well as the teaching intensity, for a group of children in a given school track S. Let us assume the policymaker seeks to maximize the expected aggregate next-period skills

[18] Our assumption of shocks as the source of child skill formation uncertainty is slightly different from the idea that the "true" academic potential of a child cannot be perfectly observed and must be learned over time from signals, such as school grades. We discuss the differences that would imply a model with imperfectly observed child skills in Appendix Section 2.F.

[19] Note that we assume the heterogeneity in instruction paces across tracks does not entail systematic differences in teacher quality or resources devoted to teaching across tracks that could also affect child skill development. The literature on international differences in student achievement tends to find limited effects of resources spent per student on learning outcomes (Woessmann, 2016). In Appendix 2.C, we summarize information on expenditure per student as well as teacher quality across different school tracks in Germany. While we do not necessarily abstract from these factors in affecting child skill development, we conclude that they are not correlated with school track.

[20] This introduces non-linear teacher effects. Duflo et al. (2011) empirically provides evidence of those.

[21] This follows from taking the first order condition of the child skill formation in (2.24) with respect to p^*_j.

$(\mathbb{E}(\overline{\theta_{j+1}}))$ and knows the child skill formation in (2.24).[22]

Lemma 1. The optimal pace of instruction, given a distribution of child skills in a given track S, is given by

$$P^*_{j,S} = \frac{\beta_j}{\gamma_j} + \bar{\theta}'_{j,S}. \tag{2.4}$$

Proof. Follows from taking the first order condition of the conditional expected value $\mathbb{E}(\theta_{j+1|S})$ in (2.24) with respect to p_j under the i.i.d. assumption of η_{j+1}. □

Lemma 1 establishes that the optimal pace of instruction a policymaker would pick for the group of children in track S will be a function of the average skill level of this group, $\bar{\theta}_{j,S}$.[23]

Replacing the pace of instruction in (2.24) with the optimal one results in:

$$\begin{aligned} \theta'_{j+1} &= g_j(\theta'_j,\ E,\ \bar{\theta}'_{j,S'}) \\ &= \kappa_{0,j} + \kappa_{1,j}\theta'_j + \kappa_{2,j}{\theta'_j}^2 + \kappa_{3,j}\bar{\theta}'_{j,S'} + \kappa_{4,j}(\theta'_j - \bar{\theta}'_{j,S'})^2 + \kappa_{5,j}E + \eta'_{j+1}, \end{aligned} \tag{2.5}$$

where $\kappa_{0,j} = \frac{\beta_j^2}{2\gamma_j}$, $\kappa_{1,j} = (1 + \ \beta_j)$, $\kappa_{2,j} = -\kappa_{4,j}$, and $\kappa_{3,j} = \alpha_j$, $\kappa_{4,j} = \frac{\gamma_j}{2}$ and $\kappa_{5,j} = \zeta_j$. We set $\kappa_{2,j}$, $\kappa_{3,j}$, and $\kappa_{4,j}$ to zero in the pre-school period ($j = 1$), as they correspond to the school inputs. In primary school, all children are in comprehensive schools ($j = 2$), and the average skills of peers $\bar{\theta}'_2$ correspond to the average skills in the cohort, assuming a representative classroom. At the beginning of secondary school ($j = 3$), children are assigned to either the academic or vocational track by their parents ($S' = A, V$). In periods $j = 3, 4$, the school component will depend on the composition of students in each of the tracks.

End-of-school skills determine the first adult human capital level, h_5:

$$h_5 = \exp(\theta_5). \tag{2.6}$$

[22] For example, in Germany, the curricula in the different tracks are set by each federal state under some general federal education goals. They consist of learning and competence goals as well as methods and specific topics that should be taught separately for each school track, subject, and school grade. The curricula are subject to frequent review and renewal. For example, as of 2023, 14 out of 16 federal states in Germany updated the curriculum in the last four years and 7 out of 16 in the last two years.

[23] It is thus only (first-best) optimal for a child with exactly the average skill level in that track. Every child with a skill level above or below the average loses in terms of future skills compared to a world in which she would be taught at her individually optimal level. Clearly, in a first-best world, a policymaker would like to provide every child with her preferred pace of instruction, which would trivially maximize end-of-school skills.

2.3.2 Preferences, Labor Income and Social Transfers

Preferences

We assume that the preferences of adult individuals over consumption and leisure take the following form:

$$u(c_j, n_j) = \frac{(c_j/q_j)^{1-\sigma}}{1-\sigma} - b\frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}, \tag{2.7}$$

where q_j is an adult consumption-equivalent scale that depends on the household composition so that c_j remains the consumption of the household (Yum, 2022). Risk aversion is captured by σ.

In addition, we assume that parents are biased toward their own educational background. When deciding on their child's school track, they incur a utility cost $\chi(E, S')$ that depends on their own education attainment E and their child's school track S'.[24] In particular, we assume that $\chi(E, S') = \chi_1 \mathbb{I}\{E = 1 \wedge S' = A\} + \chi_0 \mathbb{I}\{E = 0 \wedge S' = V\}$, so that the bias of college-educated parents is governed by χ_1 and that of vocational parents by χ_0.

Moreover, when individuals reach adulthood ($j = 5$), those who opt for and graduate from college face the following utility cost $\psi_1(S, \theta_5, \nu(E^p))$:[25]

$$\begin{aligned}\psi_1(S, \theta_5, \nu(E^p)) &= \exp(\psi + \psi_{S=V} + \psi_\theta \theta_5 + \nu(E^p)) \\ \nu(E^p) &\sim \mathcal{N}(\mu_{\nu,E^p}, \sigma_\nu^2).\end{aligned}$$

This cost depends on their secondary school track S, their end-of-school skill level θ_5, and an idiosyncratic "college taste" shock, $\nu(E^p) \sim \mathcal{N}(\mu_{\nu,E^p}, \sigma_\nu^2)$, which is influenced by their parent's education level E^p. This formulation represents two salient features of the transition between secondary and college education in the data that we ask our model to replicate. Firstly, the share of children with an academic track secondary school degree who end up getting a college degree is significantly higher than those with a vocational secondary school degree, which is why the shock depends explicitly on the school track S through $\psi_{S=V}$.[26] Secondly, independently of the school track, the likelihood of college education in

[24] We can view this as a result of unmodeled informational friction.

[25] The purpose of this cost is to account for the fact that it is, in principle, possible to obtain a college education even after not graduating from an academic track secondary school. However, college education through such "second-chance" opportunities are much less frequent. In Germany, every graduate from an academic track secondary school gets an official qualification 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.

[26] In Germany, most of this is coming from the fact that an academic track secondary school diploma auto-

the data is increasing in the end-of-school skills, so we allow the college cost to depend on the end-of-school skill level as well through ψ_θ.[27] Finally, to account for additional heterogeneity in the college decision, we allow for normally distributed taste shocks $\nu(E^p)$ that depend on parental education. Its purpose is to reflect heterogeneity in the higher education decision coming from parental background or from channels that are outside of this model.[28]

Finally, we assume that parents are altruistic. Parents take into account the continuation value of their child to a factor of δ when making inter-vivos transfers, thus exhibiting altruism. When entering adulthood ($j = 5$), individuals thus differ in their human capital h_5, their school track S, their learning ability ϕ, and their initial savings a_5.

Adult Human Capital, Labor Income and Borrowing

During their working life ($j = 5$ to $j = 16$), human capital grows stochastically,

$$h_{j+1} = \gamma_{j,E}\ h_j\ \varepsilon_{j+1}, \quad \log \epsilon_j \sim \mathcal{N}(0, \sigma_\epsilon^2), \tag{2.8}$$

and the earnings are given by:

$$y_j = w_E\ h_j\ n_j, \tag{2.9}$$

with n_j the number of hours worked and h_j the accumulated human capital. We assume that all work during the higher education stage ($j = 5$) is of a vocational nature, and therefore, earnings are $y_5 = w_0\ h_5\ n_5$. Note that all prices, including the wage rate w_E, implicitly depend on the distribution of agents in the economy, which we suppress for notational convenience. The human capital is subject to idiosyncratic market luck shocks ε_{j+1}, which we assume follows an i.i.d. normal distribution in logs, with zero mean and constant variance σ_ε^2, as in Huggett et al. (2011).

In the remainder of her life, the agent receives retirement benefits $\pi_j(h_{17})$, which depend on the last human capital level before retirement.[29] Finally, the value of death is normalized

matically qualifies for university entrance.

[27] Net of the above-explained effect coming through the secondary school track graduation, this may partly be due to the fact that for many university degrees, admission is competitive and often even requires a specific end-of-school grade average ("numerus clausus"). Of course, it could also simply reflect the selection of higher-skilled school graduates into an academic career, where these (mostly cognitive) skills are more useful.

[28] Indeed, we recognize that there are many pathways into academic higher education in Germany that fall outside the scope of our model. For example, in some vocational professions, it is possible to become eligible for university entrance after having worked a number of years in that profession. Moreover, the college decision seems to be influenced by parental education, even net of effects going through secondary school track decisions and end-of-school skills. These types of "psychic costs of higher education" are common in the literature (see e.g. Abbott et al. (2019); Daruich (2022)).

[29] As is common in the literature, we let benefits depend on human capital in this way to proxy for lifetime

to zero.

Throughout their life, adult model agents can save into a risk-free asset a, which pays a period interest rate r. As in Lee and Seshadri (2019), we assume that each agent's borrowing is constrained by the amount that can be 100% repaid in the next period using a government transfer g. Moreover, agents cannot borrow against the future income of their children. The per-period borrowing constraint can thus be written as

$$a_{j+1} \geq \frac{-g}{1+r}. \tag{2.10}$$

In the following, we provide the recursive formulation of the agent's decisions. We begin by discussing the decision problem of the working adult with a dependent child, as this is the time at which the agent chooses the school track for her child, which is at the heart of our model. Following this, we will discuss the decision problems of the working adult without dependent children and of the retiree.

2.3.3 The Optimization Problems

The timing of the model is as follows. At the beginning of each adulthood period, individuals learn about their productivity shock and, in case they have a child, about the child skill shock. Based on this information, an individual decides her consumption (c_j), her savings for the next period (a_{j+1}), and if she is not retired, her hours worked (n_j).

In addition, an adult must decide on her college education in period $j = 5$, on the inter-vivos transfer in period $j = 13$, and, importantly, on the school track of her child in period $j = 11$.

Work Life and Parenthood, $j = 9, 10, 11, 12, 13$

For each period of the work life, the adult's choices are subject to the human capital production technology in (2.8), the earnings in (2.9), a budget and time constraints:

$$c_j + a_{j+1} = y_j + (1+r)a_j - T(y_j, a_j), \quad c_j > 0, \quad n_j \in [0,1], \tag{2.11}$$

and the borrowing constraint (2.10). The interest rate is r, and taxes net of transfers is $T(y_j, a_j)$. Thus, in each period, the individual will optimally allocate her unit of time between hours worked n_j and leisure. She will also decide how to optimally spend her earnings y_j, capital gains, and taxes net of transfers $T(y_j, a_j)$ between consumption c_j and savings a_{j+1},

earnings, which form the basis of pension benefits in reality.

subject to the life-cycle borrowing constraint $\underline{a}$. For simplicity, we suppress those constraints in the following formulations.

Before The School Track Decision ($j = 9,\ 10$) The child's learning ability ϕ' is realized, and, given the initial skill level is given by (2.2), the child's skill $\theta'_{j'}$ evolves according to equation (2.5). A parent with a dependent child in the household solves the following life-cycle savings problem:

$$V_j(E, h_j, a_j, \phi', \theta'_{j'}) = \max_{c_j, a_{j+1}, n_j} \left\{ u(\frac{c_j}{q_j}, n_j) + \beta \int V_{j+1}(E, h_{j+1}, a_{j+1}, \phi', \theta'_{j'+1}) dF(\varepsilon_{j+1}) dF(\eta'_{j'+1}) \right\}$$
$$\text{s.t. } \theta'_{j'+1} = g_{j'}(\theta'_{j'}, E, \bar{\theta}'_{j'}) \ , \quad \eta'_{j'+1} \sim \mathcal{N}(0, \sigma^2_{\eta, j'+1}).$$

The state space includes the parent's and the child's variables. The first state, E, denotes whether the individual is vocational or college educated, h_j her level of human capital determined in the previous period, and a_j, and her savings. In addition, her state space comprises the child's learning ability ϕ' and current skills θ'_j. In the preschool period $j = 1$, the schooling component of the technology of skill formation is zero, and the evaluation of the child's skills does not depend on other children's skills. In the primary school period $j = 2$, the child goes to a comprehensive school, and the evolution of her skills depends on the average skill level of all children in that cohort $\bar{\theta}'_{j'}$.

The School Track Decision ($j = 11$) The school track choice happens at the beginning of secondary school. After the parent observes the realization of her child's skills, θ'_3, she makes the decision whether to send her child to the vocational or academic track school, $S' \in \{V, A\}$. This decision is affected by the value of placing her child in each track, V_{11}, but also by a fixed preference shifter, $\chi(E, S')$, that depends on the child's but also on the parent's educational attainment. Then the value at the beginning of period 11 is given by

$$V_{11}(E, h_{11}, a_{11}, \phi', \theta'_3) = \max_{S' \in \{V, A\}} \{W_{11}(E, h_{11}, a_{11}, \phi', \theta'_3, S') - \chi(E, S')\}, \tag{2.12}$$

where the values of sending a child to a school that belongs to school track S' are given by

$$W_{11}(E, h_{11}, a_{11}, \phi', \theta'_3, S') = \max_{c_{11}, a_{12}, n_{11}} \left\{ u(\frac{c_{11}}{q_{11}}, n_{11}) + \beta \int V_{12}(E, h_{12}, a_{12}, \phi', \theta'_4, S') dF(\varepsilon_{12}) dF(\eta'_4) \right\}$$
$$\text{s.t. } \theta'_4 = g_3(\theta'_3, E, \bar{\theta}'_{3,S'}) \ , \quad \eta'_4 \sim \mathcal{N}(0, \sigma^2_{\eta, 4}). \tag{2.13}$$

Future child skills are now affected by $\bar{\theta}_{3,S'}$, which are the average skill levels among children in school track S'. Parents need to form expectations over average skill levels in each track, which in equilibrium, must coincide with the actual distributions. As with prices, including the interest rate on savings r, and the wage rate w_E, we keep the dependence of average skill levels on the aggregate distribution implicit.

Remaining Parenthood ($j = 12, 13$) In period $j = 12$, the parent solves the following problem:

$$
\begin{aligned}
W_{12}(E, h_{12}, a_{12}, \phi', \theta'_4, S') = \max_{c_{12}, a_{13}, n_{12}} & \left\{ u(\frac{c_{12}}{q_{12}}, n_{12}) + \beta \int V_{13}(E, h_{13}, a_{13}, \phi', \theta'_5, S') dF(\varepsilon_{13}) dF(\eta'_5) \right\} \\
\text{s.t. } \theta'_5 = g_4(\theta'_4, E, \bar{\theta}'_{4,S'}) \ , & \quad \eta'_5 \sim \mathcal{N}(0, \sigma^2_{\eta,5}),
\end{aligned}
\tag{2.14}
$$

where the child's school track S' that has been decided in the previous period is now included in the parent's state space.[30]

Just before her child reaches the age of 18 and becomes independent, the parent decides on a financial inter-vivos transfer that her child receives, a_5, while taking into account the child's future value V'_5. As in Daruich (2022), we model this as an interim decision problem and assume that the parent already knows the realization of her own market luck shock and her child's final skill shock but does not know the realization of the college taste shock $\nu'(E)$. As is common, the transfer cannot be negative, so parents cannot borrow against the future income of their child. The value at the beginning of period 13 is then

$$
\begin{aligned}
V_{13}(E, h_{13}, a_{13}, \phi', \theta'_5, S') = \max_{a'_5 \geq 0} & \left\{ \widetilde{V}_{13}(E, h_{13}, a_{13} - a'_5) + \delta \, \mathbb{E}_\nu \, V'_5(\theta'_5, a'_5, \phi', S', E) \right\} \\
\nu'(E) \sim & \ \mathcal{N}(\mu_{\nu,E}, \sigma^2_\nu),
\end{aligned}
\tag{2.15}
$$

where $\widetilde{V}_{13}$ is the value for a parent with savings a_{13} after the inter-vivos transfer has been made

$$
\begin{aligned}
\widetilde{V}_{13}(E, h_{13}, a_{13}) = \max_{c_{13}, s_{14}, n_{13}} & \left\{ u(c_{13}, n_{13}) + \beta \int V_{14}(E, h_{14}, a_{14}) dF(\varepsilon_{14}) \right\} \\
\text{s.t. } c_{13} + a_{14} + a'_5 = & \ y_{13} + (1 + r) a_{13} - T(y_{13}, a_{13}).
\end{aligned}
\tag{2.16}
$$

where the transfer a'_5 enters the budget constraint.

[30] That is, we assume that there is no track-switching possibility during secondary school. For example, in 2010/11, only around 2.5% of children in the first stage of secondary school in Germany switched school tracks (Bellenberg and Forell, 2012). Moreover, this number includes switches among different tracks that we group into the vocational track, so is likely an upper bound of the track switches between the vocational and academic tracks. The great majority of track switches are from an academic track school to a vocational track school rather than the other way.

Work Life Without a Dependent Child, $j = 5, 6, 7, 8$ and $j = 14, 15, 16$

As before, for each period of the work life, the choices of the working adult's without a dependent child are subject to the human capital production technology in (2.8), the earnings in (2.9), the budget and the time constraints in (2.11) and the borrowing constraint (2.10). For simplicity, we also suppress them in the following formulations.

Independence ($j = 5$) At the beginning of adulthood ($j = 5$), the human capital h_5 depends on past child skills following (2.6). The newly independent adult solves the following problem:

$$\begin{aligned} V_5(\theta_5, a_5, \phi, S, E^p) = \max\{&W_5(E = 0, h_5, a_5, \phi), \\ &W_5(E = 1, h_5, a_5, \phi) - \psi_1(S, \theta_5, \nu(E^p))\} \\ \nu \sim \mathcal{N}&(\mu_{\nu,E^p}, \sigma_\nu^2), \end{aligned} \tag{2.17}$$

where W_5 denotes the values of the college and non-college education, given by

$$W_5(E, h_5, a_5, \phi) = \max_{\substack{c_5, a_6 \\ n_5 \in [0, \bar{n}(E)]}} \left\{ u(c_5, n_5) + \beta \int V_6(E, h_6, a_6, \phi) dF(\varepsilon_6) \right\}. \tag{2.18}$$

In addition to choosing her consumption level, savings, and hours worked the newly independent adult chooses whether to graduate from college or not. Her state space includes her past child skills θ_5, her savings that she received from her parent a_5, and her learning ability that she will eventually (imperfectly) transfer to her child. The last two state variables are her school track S and the higher education state of her parent $E^p \in \{0, 1\}$, which influence her utility cost of going to college $\psi(S, \theta_5, \nu(E^p))$. Note that during her college education, an individual may also work, but only at the vocational wage rate w_0. However, obtaining a college education reduces the time available for work, as individuals spend part of their total time endowment studying, thus $\bar{n}(E = 1) < 1$.[31]

Remaining Work Life ($6, 7, 8$ and $j = 14, 15, 16$) For the two following periods $j = 6, 7$, an adult without a child solves the following life-cycle savings problem:

$$V_j(E, h_j, a_j, \phi) = \max_{c_j, a_{j+1}, n_j} \left\{ u(c_j, n_j) + \beta \int V_{j+1}(E, h_{j+1}, a_{j+1}, \phi) dF(\varepsilon_{j+1}) \right\}. \tag{2.19}$$

The state space only includes the parent's related variables, including her learning ability that will be transferred (imperfectly) to her child.

[31] 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.

In period $j = 8$, the individuals know that they will become parents next period. For that reason, they take expectations over the learning ability of their future child, ϕ', on top of the expectations over the market luck shocks. We assume that ability is imperfectly transmitted from parents to children, according to $\phi^c \sim G(\phi'|\phi)$ and (2.1). Thus, the value in period 8 becomes

$$\begin{aligned} V_8(E, h_8, a_8, \phi) &= \max_{c_8, a_9, n_9} \left\{ u(c_8, n_8) + \beta \int V_9(E, h_9, a_9, \phi') dF(\varepsilon_9) dG(\phi'|\phi) \right\} \\ &\text{s.t. } \log \phi' = \rho_\phi \log \phi + \epsilon_\phi, \qquad \epsilon_\phi \sim \mathcal{N}(0, \sigma_\phi^2). \end{aligned} \tag{2.20}$$

For the rest of the periods $j = 14, 15, 16$, an adult whose child became independent solves the following life-cycle savings problem:

$$V_j(E, h_j, a_j) = \max_{c_j, a_{j+1}, n_j} \left\{ u(c_j, n_j) + \beta \int V_{j+1}(E, h_{j+1}, a_{j+1}) dF(\varepsilon_{j+1}) \right\}, \tag{2.21}$$

where the learning ability ϕ has already been transmitted to the child and does not enter the state space anymore.

Retirement, $j = 17, 18, 19, 20$

Everybody retires at the beginning of model period 17, which corresponds to age 66 in real life, and receives retirement benefits $\pi_j(h_{17})$. After period 20, that is, at age 82, agents die with certainty and exit the model. The values in these periods are

$$\begin{aligned} V_j(E, h_{17}, a_j) &= \max_{c_j > 0, a_{j+1} \geq \underline{a}} \{ u(c_j, 0) + \beta V(E, h_{17}, a_{j+1}) \} \\ \text{s.t. } c_j + a_{j+1} &= \pi_j(h_{17}) + (1 + r) a_j - T(0, a_j). \end{aligned} \tag{2.22}$$

2.3.4 Aggregate Production, and Government

We assume that a representative firm produces output according to the Cobb-Douglas production function $Y = K^\alpha H^{1-\alpha}$, where K is the aggregate physical capital stock and H is a CES aggregate of total labor supply, which is defined by:

$$H = [\omega H_0^\epsilon + (1 - \omega) H_1^\epsilon]^{\frac{1}{\epsilon}}. \tag{2.23}$$

Here, H_0 is the aggregate labor supply in efficiency units of workers with vocational higher education, and H_1 is that of workers with a college education. The physical capital stock depreciates at rate δ_f.

The government taxes labor income progressively, such that labor income net of taxes amounts to $y_{net} = \lambda y^{1-\tau_n}$ (Heathcote et al., 2017). It also taxes capital income linearly according to $\tau_a r a_j$ (Yum, 2022). All tax revenue is used to finance retirement benefits π_j as well as fixed lump-sum social welfare benefits g that are paid to every household. These may include child allowances, unemployment benefits, or contributions to health insurance.

2.3.5 Equilibrium

We solve for the model's stationary equilibrium and its associated distribution using the numerical strategy in Lee and Seshadri (2019). Stationarity implies that the cross-sectional distribution over all states in every period j is constant across cohorts. Our model economy consists of 20 overlapping generations or cohorts at each time. The equilibrium requires that households and firms make optimal choices according to their value functions and firm first-order conditions, respectively. Moreover, the aggregate prices for physical capital and both types of human capital r, w_0, and w_1 are competitively determined and move to clear all markets. Note that we do not require the government budget to clear as well. Instead, we assume that all government revenues that exceed the financing of all social welfare programs result in linearly independent spending.

A special feature of our model is that learning during the school years depends on the distribution of children across school tracks. Importantly, an equilibrium requires that parents form expectations over average skill levels in each track, which in equilibrium, coincide with the actual distributions. A detailed definition of the equilibrium is given in Appendix 2.B.

2.4 Developing Intuition: Child Skill Formation and School Tracking

In order to provide an intuition for the key mechanisms at work for the timing of school tracking, we simplify our full model and focus exclusively on schooling years (periods 2 to 4 in the full model) and the school tracking decision. We now ignore general equilibrium responses in prices and simplify parents' preferences. We assume parents only care about their expected child's end-of-school skills and are unbiased regarding the school track choice. As a result, the only decision concerns the school track choice.[32]

[32]In addition, we consider a policymaker who sets the instruction pace to maximize the expected aggregate next-period skills. This will help provide some reasoning for the functional form of the child skill technology (2.5).

2.4.1 Intuition for the Child Skill Formation

Child Skill Formation

Suppose each child i arrives just before entering school with a set of skills that can be summarized in a univariate level, $\theta_{i,j=1}$.[33] In the remaining, we simplify the notation by omitting the subscript i. We specify the following form for the child skill technology, where θ_j refers to the child i's skill level in period j:

$$\begin{aligned} \theta_{j+1} &= \theta_j + \alpha\bar{\theta}_{-i,j} + \beta p_j - \frac{\gamma}{2}(\theta_j - p_j)^2 + \frac{\gamma}{2}\theta_j^2 + \zeta E + \eta_{j+1} \\ &= \theta_j + \alpha\bar{\theta}_{-i,j} + \beta p_j - \frac{\gamma}{2}{p_j}^2 + \gamma\theta_j p_j + \zeta E + \eta_{j+1}, \end{aligned} \tag{2.24}$$

where η_{j+1} describes unobserved shocks to the formation of skills that are assumed to be independent and normally distributed around mean zero with a variance $\sigma^2_{\eta_{j+1}}$. We assume the parameters $\gamma > 0$ and $\alpha,\ \beta,\ \zeta \geq 0$. Parental background affects the child's skill evolution linearly through ζE. In addition, the evolution of the child's skills are directly affected by past skills and through interactions with peers $\bar{\theta}_{-i,j}$, similar to Duflo et al. (2011).[34] Finally, the pace of instruction is denoted by p_j. Learning monotonically increases with the pace p_j at rate β and decreases with the distance between a child's current skill and the pace of instruction. It implies that for each skill level θ_j, there exists a unique optimal pace of instruction p_j^* that maximizes the child's future skill level, keeping everything else fixed. The optimal pace of instruction p_j^* for a child with skills θ_j is given by $p_j^* = \frac{\beta}{\gamma} + \theta_j$.[35] It is strictly increasing in current skill, such that higher-skilled children also prefer a higher pace of instruction. With β reasonably small, these assumptions imply that for a child with a very low current skill level attending a class with a very ambitious, high instruction pace can be harmful to the point when she actually loses skills. Similarly, a high-skilled child might be so sub-challenged in a class with a very low pace that she actually loses skills.

---

[33] As in Cunha and Heckman (2007), we do not differentiate between abilities and skills, as both are partly endogenously produced and partly exogenously determined pre-birth. Moreover, we do not allow for potentially different production technologies of cognitive and non-cognitive skills as in Cunha et al. (2010) or Daruich (2022). Instead, in the tradition of Becker and Tomes (1986), we focus on one composite skill, which can be translated into adult human capital rewarded on the labor market after school.

[34] We concentrate on the case with a linear-only direct peer externality governed by α. As summarized in Epple and Romano (2011), many studies find that such linear-in-means peer effects are sizable and robust across settings. Evidence on non-linear peer effects in the classroom is more ambiguous. For that reason, we do not incorporate non-linearities in peer effects directly. Instead, we consider the endogenous setting of instruction levels across school tracks as a channel through which non-linear peer effects arise.

[35] This follows from taking the first order condition of the child skill formation in (2.24) with respect to p_j^*.

Optimal Pace of Instruction

Consider a policymaker choosing the pace of instruction, which we think as reflecting the curriculum as well as the teaching intensity, for a group of children in a given school track S. Let us assume the policymaker seeks to maximize the expected aggregate next-period skills ($\mathbb{E}(\overline{\theta_{j+1}})$) and knows the child skill formation in (2.24).

Lemma 2. The optimal pace of instruction, given a distribution of child skills in a given track S, is given by

$$P^*_{j,S} = \frac{\beta}{\gamma} + \bar{\theta}_{j,S} \tag{2.25}$$

Proof. Follows from taking the first order condition of the conditional expected value $\mathbb{E}(\theta_{j+1|S})$ in (2.24) with respect to p_j under the i.i.d. assumption of η_{j+1}. □

Lemma 1 establishes that the optimal pace of instruction a policymaker would pick for the group of children in track S will be a function of the average skill level of this group, $\bar{\theta}_{j,S}$. It is thus only (first-best) optimal for a child with exactly the average skill level in that track. Every child with a skill level above or below the average loses in terms of future skills compared to a world in which she would be taught at her individually optimal level.[36] Note that replacing the pace of instruction in (2.24) by the optimal one results in our full model child skill formation (2.5) where $\kappa_{0,j} = \frac{\beta^2}{2\gamma}$, $\kappa_{1,j} = (1+\ \beta)$, $\kappa_{2,j} = -\kappa_{4,j}$, and $\kappa_{3,j} = \alpha$, $\kappa_{4,j} = \frac{\gamma}{2}$ and $\kappa_{5,j} = \zeta$.

2.4.2 Underlying Mechanisms of School Tracking

Let the schooling system in a given stage be characterized by the number of distinct tracks S. If there is only one track to which all schools belong, we speak of a *comprehensive system*. If there are two distinct school tracks, we speak of a *tracking system*.[37] Consistently with our full model, we assume that once a comprehensive system is switched to a tracking system between school stages, it cannot switch back to a comprehensive system. Finally, we do not allow children to switch between school tracks.[38] The pace of instruction in each track

[36] Clearly, in a first-best world, a policymaker would like to provide every child with her preferred pace of instruction, which would trivially maximize end-of-school skills.

[37] Thus, while in principle, a large number of school tracks is conceivable, we restrict a tracking system to two school tracks as this corresponds to a typical number across OECD countries.

[38] To the best of our knowledge, there are no cases in OECD countries where a comprehensive system follows a tracking system. In virtually all countries, the schooling years start with comprehensive primary school. Tracking into distinct school types then occurs, if at all, at some point during secondary school. Among OECD countries, the first age of school tracking varies from age 10 in Austria and Germany to age 16 in Australia, Canada, Chile, Denmark, Finland, Iceland, New Zealand, Norway, Poland, Spain, Sweden, the United Kingdom, and the United States (OECD, 2013). While, in principle, switches between tracks

is chosen by a policymaker seeking to maximize aggregate end-of-school skills as in (2.25). The resulting child skill technology (2.5) is common knowledge.[39] Moreover, we assume the distribution of skill levels among all children and the skill shock distribution are common knowledge. In equilibrium, expectations about the paces of instruction are consistent with the actual paces of instruction.

In the following, we discuss the sorting mechanism of children across tracks and compare the distribution of the end-of-school skills between both systems, assuming away direct parental inputs ($\zeta = 0$).

One-period Schooling System

We start by considering the case with one period of schooling only, so that θ_2 are the skills at the end of school. This means that either a tracking system or a comprehensive system can be implemented for the entirety of the school year.

Sorting Mechanisms We consider two alternative allocation mechanisms. In the first one, a policymaker allocates children across tracks by maximizing the expected aggregate end-of-school skills ($\mathbb{E}(\overline{\theta_2})$). The second alternative consists of parents making the track decision for their child i, with the goal of maximizing their expected child's end-of-school skill level ($\mathbb{E}(\theta_{i,2})$).[40]

Proposition 1 shows that, in both alternatives, the track decision is governed by a sharp cut-off skill level. A policymaker would optimally split the distribution exactly at its median.[41] Intuitively, this generates the highest aggregate end-of-school skills as it minimizes the variance of skills in each track, or in other words, it creates peer groups that are as homogeneous as possible. In doing so, the policymaker internalizes that any effects coming from the direct peer externality exactly offset each other across tracks. Thus, all gains achieved

during the secondary school years are possible, they are very rare. Dustmann et al. (2017) document a share around 2% of track switches during secondary school in Germany.

[39] For example, in Germany, the curricula in the different tracks are set by each federal state under some general federal education goals. They consist of learning and competence goals as well as methods and specific topics that should be taught separately for each school track, subject, and school grade. The curricula are subject to frequent review and renewal. For example, as of 2023, 14 out of 16 federal states in Germany updated the curriculum in the last four years and 7 out of 16 in the last two years.

[40] This has become common practice in Germany, where in the majority of federal states, parents are completely free in making the secondary school track choice for their children. Only in three states, Bavaria, Thuringia, and Brandenburg, academic school track access is conditional on a recommendation by the primary school teachers. These recommendations are often tied to achieving a certain grade point average in math and German in primary school. However, even in these states, children without a recommendation can take advantage of a trial period in an academic track school, after which the child will be assessed again.

[41] Equivalent to the mean in this context. A similar argument has been made repeatedly in the theoretical literature. See for instance, Epple and Romano (2011).

from making average peer skills in one track higher are lost as the average level in the other track becomes smaller.

In contrast, if parents are the decision-makers, they ignore the aggregate outcomes. Whenever the direct peer effects are positive ($\alpha > 0$), the cut-off skill level is smaller than the optimal threshold a policymaker would pick. More children would go to the highest peer group track. Intuitively, parents do not internalize the effect their child's skills have on aggregate end-of-school skills. The threshold is characterized by the skill level $\tilde{\theta}_1^*$ at which the child's next period skill is in expectation, the same in both tracks.

Proposition 1. In Equilibrium, the allocation of children across tracks is characterized by a skill threshold $\tilde{\theta}_1$, such that all children with initial skills below $\tilde{\theta}_1$ go to one track and all children with initials skills above $\tilde{\theta}_1$ go the other track.

- If the track allocation is done by the policymaker, the optimal skill threshold corresponds to the average initial skill level $\tilde{\theta}_1^* = \mathbb{E}[\theta_1] = 0$.
- If the track allocation is done by parents, the endogenous skill threshold that emerges from this game depends on the direct peer externality α. With $\alpha > 0$, the threshold is smaller than $\tilde{\theta}_1^*$.[42]

Proof. In Appendix 2.A. □

The End-of-school Distribution Proposition 2 describes the end-of-school distribution of skills in both schooling systems in the one-period model. Independently of the sorting mechanism, the average end-of-school skills in a full optimal tracking system are always larger than in a comprehensive system. Intuitively, this advantage comes from more homogeneous peer groups in each track in terms of their initial skills. Since learning generally decreases in the variance of skills among children in a track, more homogeneity on average increases end-of-school skills.

A full tracking system necessarily leaves a non-negative mass of children worse off compared to a comprehensive system. These children have initial skills around the tracking threshold and would be closer to their optimal instruction pace in a comprehensive system. In an optimal tracking system with $\tilde{\theta}_1 = 0$, these children thus occupy the center of the distribution and would, given a choice, prefer a comprehensive system.[43] If there are no

[42] We rule out an (uninteresting) equilibrium of the track choice game in which parents randomly allocate their child into one of the two tracks, leading to the same distribution of skills in both tracks and, hence, the same pace of instruction.

[43] This is interesting in a political economy context as the median voter in this model would prefer a comprehensive system. This could partially explain why we see different tracking systems across different countries.

direct peer effects, an equal share of children in both tracks lose relative to the comprehensive counterpart. However, with positive peer effects the losses are concentrated among the track with the lower average peer level. This reflects a robust finding of the empirical school tracking literature that especially the children at the bottom of the skill distribution suffer from a tracking system (see, e.g., Matthewes (2021)).

Proposition 2.

- Aggregate end-of-school skills in a full tracking system are larger than in a full comprehensive system. This holds regardless of who makes the track decision, i.e. regardless of the tracking skill threshold. $\tilde{\theta}_1$
- Children with initial skills inside a non-empty interval lose from a full tracking system in terms of their end-of-school skills relative to a fully comprehensive system. With $\alpha = 0$, the losses are symmetric in both tracks. With $\alpha > 0$, the losses are concentrated in the track with the lower average skill level.

Proof. In Appendix 2.A. □

Note that these results are not affected by the presence of skill shocks in the one-period model. This is because these shocks are assumed to be mean zero and realized at the end of the period.[44] They matter in a two-period schooling system.

Two-period Schooling System

Let us now consider a two-period schooling system so that θ_3 are the skills at the end of school. We are interested in a comparison between the end-of-school skill distribution in an early tracking system, ET, and a late tracking system, LT, in which the allocation is done by the policymaker who maximizes the expected aggregate end-of-school skills ($\mathbb{E}(\overline{\theta_3})$). The early tracking system is characterized by an initial track allocation into two tracks, V and A, at the beginning of the school year, $j = 1$.[45] In an early tracking system, a child, therefore, remains in her school track for the two school periods.[46] The late tracking system is characterized by all children going to a comprehensive school in the first period, followed

[44] Therefore, the realization of shocks affects the end-of-school distribution only in that it raises the variance uniformly in both tracks.

[45] As argued before, we do not allow for track switches during the schooling years, including at the beginning of $j = 2$.

[46] Absent any costs of re-tracking, a school system that features such a re-tracking possibility would improve aggregate end-of-school skills in our model relative to the early tracking system we describe here. However, we focus on the early tracking system, as in reality, re-tracking (or second chance) opportunities *during* school years are used only relatively rarely.

by the allocation into schools that belong to one of two tracks, V and A, at the beginning of the second period. Importantly, while in the LT case, this allocation occurs after the skill shock η_2 is realized, in the ET case, the allocation occurs before.[47]

Proposition 3 shows that aggregate end-of-school skills in an optimal LT system can be larger than in an optimal ET system if the variance of the skill shocks is large enough. Intuitively, in the presence of skill uncertainty, early tracking generates misallocation that leads to learning losses not only for the misallocated individual child but also poses an externality for all other children as the instruction pace is endogenous to the peer composition. Indeed, in the second period, the unexpected skill shocks render the peer group more heterogeneous. Depending on the size of the skill uncertainty, the efficiency gains of the ET system in the first period can be outweighed by the losses due to misallocation in the second period.

Proposition 3. Average end-of-school skills in the two-period model are larger in an optimal late tracking system than in an early tracking system iff

$$\frac{\sigma^2_{\eta_2}}{\sigma^2_{\theta_1}} > 1 + \alpha + \alpha^2 + \beta + \frac{\beta^2}{2} + 2\alpha(1+\beta) + \frac{\gamma^2}{2\pi}\sigma^2_{\theta_1}. \tag{2.26}$$

Proof. In Appendix 2.A. □

This parsimonious modeling provides a reasoning for the child skill formation (2.5) and helps see the main mechanisms at work for school tracking: efficiency gains of homogenous peer groups and misallocation losses due to the child skill uncertainty. The efficiency of a tracking system's timing will depend not only on estimates of the (age-specific) child skill formation technology parameters but also on the size of the skill uncertainty.

However, this simplified version of the model cannot help us quantify the aggregate effect of a policy that would delay the tracking choice. Indeed, to obtain analytical results and form intuition, we ignored important features of the full model that also influence the efficiency of school tracking policies. In particular, we have defined simple objective functions, namely the maximization of (aggregate) end-of-school skills. Still, in a richer model, parents and policymakers can have different objectives. For example, the policymaker could maximize the aggregate output in (2.23), and parents would consider future labor market prospects that depend on the tracking choice and college graduation. Moreover, they may be biased toward their educational path, which is another source of misallocation to be considered when evaluating the effects of the timing of school tracking. Finally, the wage rate is a function of

[47] We do not consider a fully comprehensive system in which children remain in comprehensive track schools for the whole duration of their school career. Proposition 2 implies that such a system cannot achieve higher aggregate end-of-school skills compared to a late tracking system.

labor demand by firms that likely employ labor from both vocational and academic degrees. Thus, general equilibrium effects and later college choices are also important to consider to provide insights into the effects of the timing of school tracking policies on lifetime and intergenerational outcomes.[48] To consider all those factors in evaluating the effects of school tracking policies, we now calibrate a quantitative version of the full model.

2.5 Model Calibration

As is common in the literature, we parameterize the model following a two-step approach. In the first step, we estimate the parameters of the child skill formation technology during the school years, as well as other selected model parameters directly from the data. In the second step, the remaining parameters are estimated using the simulated method of moments by matching the moments from the stationary equilibrium distribution of the model to their empirical counterparts. A summary of the externally calibrated parameters is given in Table 2.6 and of the internally estimated ones in Table 2.7.

2.5.1 Data and Sample Selection

All externally estimated parameters in the first step and moments used in the second step are based on two data sources. The first source is the German National Educational Panel Study (NEPS) described in Section 2.2. We further restrict the sample to individual observations containing information on the school and class in that school a child attended in a given year.

The second data source is the German Socio-Economic Panel (SOEP), an annual representative survey from which we use the 2010-2018 waves. The data contains rich information on labor supply, income, and education on the individual level. We use this data source primarily to construct empirical moments for the working stage of the life cycle, as will be detailed below. For this reason, the only sample selection that we do is dropping those with hourly wages below the first and above the 99th percentile. Lastly, we convert all income data to 2015 Euros using a CPI index for inflation adjustment.

We begin by detailing how we measure, identify, and estimate the parameters of the child skill formation technology, as these constitute the most important ingredient of our model. Then, we describe the functional forms and estimation strategies for all remaining parameters.

[48] See Dustmann et al. (2017) for an excellent discussion about the pathways through the German Education System.

2.5.2 Estimation of the the Child Skill Formation Technology

We specify the production technology of (the logarithm) of child i's skills (2.5) that we take to the data as follows:[49]

$$\theta_{i,j+1} = \kappa_{0,j} + \kappa_{1,j}\theta_{i,j} + \kappa_{2,j}\bar{\theta}_{-i,j,S} + \kappa_{3,j}{\theta_{i,j}}^2 + \kappa_{4,j}{(\theta_{i,j} - \bar{\theta}_{j,S})}^2 + \kappa_{5,j}E_i + \eta_{i,j+1}, \qquad (2.27)$$

Note that (2.27) is just a rearranged version of the child skill technology (2.24) after substituting in the optimal pace of instruction. Moreover, we, in principle, allow all child skill technology estimates to be specific to the period j. We denote by $\bar{\theta}_{-i,j,S}$ the average skill level of child i's *classroom* peers, as opposed to $\bar{\theta}_{j,S}$, which refers to the average skill level of all children in a school that belongs to track *theta* and arises from optimal track-specific instruction paces. Note that in the model, $\bar{\theta}_{-i,j,S} = \bar{\theta}_{j,S}$, since we assume a representative school and class per track (or alternatively, identical classes conditional on school tracks). In the data, however, there is clearly heterogeneity across classes, even within a school track. Since we are interested in capturing skill development effects that arise from direct interactions with peers, which are likely occurring in a specific classroom, we, therefore, exploit that heterogeneity in the estimation. During primary school, we observe only one comprehensive track in the data. In that case, even with classroom-specific direct peer effects, we cannot fully identify the parameters in (2.27), which is why we drop θ^2.[50]

In the estimation, the parental educational attainment E is a time-constant dummy that equals 1 if child i comes from a household in which at least one parent is college educated. We use test scores to measure the evolution of this skill measure.[51] As is common in the child skill formation literature (Cunha et al., 2010; Agostinelli and Wiswall, 2016), we think of log child skills θ as latent variables that are only imperfectly measured in the data. For that reason, we employ a linear measurement system for the logarithm of latent skills in each period and identify the loadings on each measure in each period by ratios of covariances of

[49] Following the work in Cunha et al. (2010), much of the empirical and quantitative literature using child skill formation technologies employed parametric specifications of (2.5) of the constant elasticity of substitution (CES) form. As noted in Agostinelli and Wiswall (2016), this requires, under standard parameter restrictions, that all input factors are static complements. An alternative is to use a nested CES structure as in Fuchs-Schündeln et al. (2023); Daruich (2022). To retain tractability, we follow Agostinelli and Wiswall (2016) and opt for the trans-log approach.

[50] This is also the reason why we prefer (2.27) over a model that includes $\bar{\theta}^2_{j,S}$ and the interaction $\theta\bar{\theta}_{j,S}$ as separate regressors, such as (2.24), even when using class-specific peer effects. While in version (2.27), we just have to drop the squared term on skills, which is typically statistically insignificant even when two tracks are available, in version (2.24), we cannot identify either the coefficient in front of $\bar{\theta}^2_{j,S}$ or that of $\theta\bar{\theta}_{j,S}$.

[51] As argued in Borghans et al. (2008), achievement test scores measure both cognitive and non-cognitive skills.

the measures by subject (as in Agostinelli et al. (2019)).[52]

We present the estimates of the child skill production technology parameters in Table 2.4. Note that the estimates are based on the data from NEPS Starting Cohort 3, which follows children through secondary school. Since prior to grade 5, children are in a unique school track, we cannot estimate the age-specific coefficients for period 2. In addition, in grade 12, the tests are track specific, which makes the estimates unreliable for period 4. For those reasons, we assume that the estimates of the child skill technology parameters κ_2, κ_3, and κ_4 between school grades 5 and 9 are representative of the entire schooling career. That is, we drop the j index on those technology parameters. Recall that $\theta_{i,j}$ is defined as the logarithm of child skills. Hence, we can interpret the coefficients as elasticities. Thus, $\hat{\kappa}_{1,2} = 0.65$ means that a 1% increase in latent skills at the beginning of primary school is associated with an 0.65% increase with end-of-primary school skills. Generally speaking, the own-skill elasticity is close to one for the first two stages and decreases in the second half of secondary school, suggesting a relatively high own-skill productivity, as is commonly found in the literature (see estimates in Cunha et al. (2010); Agostinelli et al. (2019)). During secondary school, the estimated coefficient $\hat{\kappa}_2$ is positive. More importantly, we cannot reject the hypothesis that $\hat{\kappa}_3 = -\hat{\kappa}_4$ which is in line with Section 2.4.

The estimated coefficient $\hat{\kappa}_4$ is negative and statistically significant at the 10% level. It indicates that a 1% increase in the squared distance to the average skill level in a track is associated with an up to 0.05% decrease in the next period's skills. This lends empirical support to the idea that the instruction pace in every track is tailored to the average skill level, and deviations, in both directions, from this level can hurt individual skill development. The estimated $\hat{\kappa}_2$ is rather small and statistically insignificant.

The final estimates we use to parameterize the child skill formation technology in our model are then κ_n for $n = 2, 4$ as reported in Table 2.4. The intercept is calibrated internally, such that average log skills are always zero in the baseline model, which is one of our identifying assumptions. The parameter $\kappa_{1,3}$ comes from Table 2.4, while $\kappa_{1,2}$ and $\kappa_{1,4}$ are estimated internally.

2.5.3 Remaining Parameters

Preferences

We set the inverse elasticity of intertemporal substitution, $\sigma = 2$, a value that is common in the literature. The Frisch elasticity of labor supply is set to 0.5. The disutility shifter b

[52] Appendix Section provides details on skills measurement. Appendix Tables 2.E.1 and 2.E.2 describe the evolution of child skills over time using the identified latent variables.

Table 2.4: Child Skill Technology Parameters Estimates

Dependent Variable: $\theta_{i,j+1}$ Grade 9 on Grade 5		
Coefficient	Variable	
$\hat{\kappa}_1$	$\theta_{i,j}$	0.65 (0.026)
$\hat{\kappa}_2$	$\bar{\theta}_{-i,j,S}$	0.12 (0.082)
$\hat{\kappa}_3$	$\theta^2_{i,j}$	0.02 (0.02)
$\hat{\kappa}_4$	$(\theta_{i,j} - \bar{\theta}_{j,S})^2$	-0.05 (0.025)
$\hat{\kappa}_5$	$E = 1$	0.10 (0.04)
N Children		1,675

Notes: This table presents the coefficients of regressions of skills in grade 9 on skills in grade 5, skills squared, the average skill level of peers, distance to the average skill in the track squared, and parent's education dummy. All observations are weighted. Standard errors are clustered at the school level. Models control for age, gender, and school-fixed effects. Source: NEPS.

is estimated internally in order to match the average time worked in the SOEP data, given that the total time available after sleep and self-care is normalized to 1.

We set the time discount factor β, such that the equilibrium interest rate amounts to 4% annually. The altruism parameter δ is calibrated such that the ratio of average inter-vivos transfers to average labor income in the model corresponds to that of average higher education costs of children to average 4-year labor income in the data. According to a survey by the German Student Association in 2016, the monthly costs of living during the higher education stages range from 596 to 1250 Euros (Middendorf et al. (2019)). We expect the parents to bear the bulk of these costs and assume that they support their child for an average of 5 years (the length of time it takes to complete studies that are equivalent to a masters level). Then, the ratio of total costs to average 4-year labor income ranges from 0.32 to 0.67. In our baseline calibration, we take as a target a ratio of 0.6.

Finally, we estimate the bias parameters χ_1 and χ_0 to match the share of deviations from secondary school track recommendations by parental education in the data.

Initial Child Skills and Child Skill Shocks

Initial child skills just before entering primary school are a function of the learning ability of a child, which is imperfectly transmitted from the parent following an AR(1) process with inter-generational correlation coefficient ρ_ϕ, and variance σ^2_ϕ. Since the learning ability is correlated with the eventual higher education outcome of a parent, we pick as the target moment for ρ_ϕ the difference in average preschool child skills by parental education in one standard deviation unit. The variance σ^2_ϕ is then estimated to match the variance of initial math test scores in the data.

An integral part of the child skill development is the presence of unforeseeable, permanent shocks to child skills. As discussed in Section 2.4, the size of such shocks has important implications for the effects of school tracking policies as they can give rise to efficiency losses if "late-bloomer" effects are large. To quantify the importance of child skill shocks in our model, we internally estimate the shock variance $\sigma^2_{\eta,j+1}$, for $j = 2, 3, 4$. As target moments, we choose the correlation of a child's math test score percentile rank across periods.[53] In this way, we capture all changes in a child's relative position in the skill distribution in a given period that cannot be accounted for by the skill formation technology or track choices.[54]

[53] Appendix Table 2.E.2 describes the correlations of child skills between periods using the identified latent variables.

[54] In reality, such changes may also arise from factors that are outside the scope of this model but can put children on a different skill formation path. These could be, for example, a change of schools within a school track, a change of teachers within a class, or even tutoring sessions that are uncorrelated with parental education.

College Costs

We estimate the two parameters ψ and $\psi_{S=V}$ of the college costs to match the share of graduates from an academic secondary school who follow up with a college education and the share of vocational secondary school graduates that end up in college. We discipline the coefficient ψ_θ that multiplies end-of-school skills by matching the regression coefficient on log math test scores from a regression of a college graduation dummy on end-of-school test scores.

The normally distributed college taste shock ν, with parental education specific means μ_{ν,E^p} and variance σ_ν^2, account for additional heterogeneity in the college decision. We calibrate the two parameters $\mu_{\nu,E^p=0}$ and $\mu_{\nu,E^p=1}$ to match the share of children from each parental education background that receive a college degree in the data. Finally, we calibrate the variance, σ_ν^2 to match the variance of the residuals from the above regression of college education on end-of-school skills, as in Daruich (2022).

The final component of college costs is not a part of the "psychic" costs ψ_1 but reflects the time cost of obtaining a college education. We assume that studying for a college degree takes away around 60% of the total time available for work for four years or one model period.[55] Thus, we set the maximum remaining time during the higher education stage to $\bar{n}(E=1) = 0.40$.

Human Capital Growth

We set the child-skill-to-human-capital anchor, ξ, such that in equilibrium average labor income before taxes is equal to 1 (Lee and Seshadri, 2019). We estimate the deterministic human capital growth profiles for both types of education, $\{\gamma_{j,E}\}$, $j = 5, ..., 16$ using wage regressions in the SOEP data, following the approach in Lagakos et al. (2018).[56] The resulting

[55] A common estimate is that full-time studying takes around 40 hours per week, which amounts to around 60% of the maximum weekly work hours, which we set to 65. Moreover, the average study length in Germany is 8 semesters or 4 years.

[56] Concretely we create, separately for each education group, 4-year work experience bins. We then estimate Mincer regressions of wages on years of schooling and potential work experience, controlling for time and cohort effects of the form:

$$\log w_{ict} = \alpha + \beta s_{ict} + \delta x_{ict} + \gamma_t + \zeta_c + \epsilon_{ict},$$

where w_{ict} is the wage of individual i, who belongs to birth cohort c and is observed at time t. Wages are defined as total annual labor earnings divided by hours worked. We denote by s_{ict} the years of schooling and by x_{ict} work experience, which is defined as

$$x_{ict} = age_{ict} - 18 \text{ if } s_{ict} < 12$$
$$x_{ict} = age_{ict} - s_{ict} - 6 \text{ else.}$$

experience-wage profiles for 4-year experience bins are shown in Table 2.5, expressed in growth relative to the previous bin. We set the $\{\gamma_{j,E}\}_{j=5}^{16}$ parameters to these values.

Finally, we calibrate the variance of the market luck shocks, σ_ε^2 such that our model replicates the standard deviation of labor income across all workers in the data.

Table 2.5: Human Capital Growth Profiles

Experience (Years)	Wage Growth Non-College	College
0	1.00	1.00
4	0.96	1.15
8	1.09	1.19
12	1.10	1.11
16	1.04	1.06
20	1.02	1.01
24	1.00	0.99
28	1.01	0.97
32	0.99	0.98
36	1.01	0.99
40	0.99	1.01

Notes: This table provides wage growth estimates by year of experience and educational attainment. Source: SOEP

Firms and Government

Following large parts of the literature, we set the capital share in the aggregate production function to $\alpha = 1/3$. Moreover, we set $\sigma_f = 1/3$ such that the elasticity of substitution between vocational and academic human capital in the firm production is equal to 1.5 (Ciccone and Peri, 2005). The weight on vocational human capital in the CES aggregator, ω is estimated internally. Following the arguments in Lee and Seshadri (2019), we calibrate it to match the share of college-educated workers in the SOEP data.

Regarding the tax-related parameters, we set the labor income tax scale to $\lambda = 0.679$ and the labor tax progressivity parameter to $\tau_l = 0.128$ following estimates in Kindermann et al. (2020). The linear capital tax is set to $\tau_a = 0.35$. The size of the lump sum government transfers is set to $g = 0.06$, which in equilibrium amounts to 6% of average labor earnings in the economy. Finally, we set pension benefits to $\pi_j = \Omega h_j w_E$ during retirement and calibrate the scale parameter Ω internally, such that the average replacement rate corresponds to 40%.

To disentangle time from cohort effects, we assume that there is no experience effect on wage growth in the last 8 years of work, following the HLT approach in Lagakos et al. (2018).

Table 2.6: Parameters calibrated externally

Parameter	Value	Description	Source
Household			
σ	2.0	Inverse EIS	Lee and Seshadri (2019)
γ	0.5	Frisch Elasticity	Fuchs-Schündeln et al. (2022)
q	1.56	Equiv. Scale	Jang and Yum (2022)
$\bar{n}(E=1)$	0.40	Time Cost of College	
Firm			
σ_f	1/3	E.o.S Vocational and Academic Human Capital	Ciccone and Peri (2005)
δ_f	6%	Annual Depreciation	Kindermann et al. (2020)
Government			
τ_n	0.128	Labor Tax Progressivity	Kindermann et al. (2020)
λ	0.679	Labor Tax Scale	Kindermann et al. (2020)
τ_a	0.35	Capital Tax Rate	
g	0.06	Lump-sum Transfers	

Notes: This table presents the externally calibrated parameters and their corresponding sources.

2.5.4 Method of Simulated Moments Estimation Results

In total, we estimate 20 parameters internally using the method of simulated moments to match 20 target data moments. The parameters, their estimated values, model-implied moments, and target data moments are presented in Table 2.7.

The model generally fits the data well, both in terms of aggregate moments and concerning the distribution of child skills, school tracks and higher education. For example, the share of of college graduates in the simulated economy is 35.6%, which is in line with the German data in the 2010s. Given that the model also matches the transition rates from academic and vocational secondary school into college higher education (at 70% and 11%), this implies that the share of children in an academic track school in the model, 42% is in accordance with the data.

Parental preferences towards their own track affect the school track decision significantly, both in the model and in the data. In particular, around 20% of parents from each education background overrule a different track recommendation by teachers in the NEPS data. In the model simulated data, roughly the same shares of parents would send their child to a different track if it was not for the own-track bias.

The model is further successful in capturing the transitions between secondary and tertiary education. Around 70% of graduates from an academic track secondary school achieve a college education, while that share is 11% from a vocational track secondary school. Thus, despite the fact that in principle only an academic school degree qualifies for university en-

Table 2.7: Internally Calibrated Parameters

Parameter	Value	Description	Target	Data	Model
Preferences					
β	0.935	Discount Factor	Annl. Interest Rate	0.04	0.04
b	6.8	Labor Disutility	Avrg. Labor Supply	0.53	0.53
Δ	0475	Parental Altruism	Transfer/Income	0.60	0.59
χ_0	0.017	Own V-Track Bias	Share of Deviations	0.16	0.18
χ_1	0.021	Own A-Track Bias	Share of Deviations	0.23	0.21
College Costs					
ψ	0.88	Intercept	Share A $\rightarrow$ College	0.71	0.71
ψ_V	0.25	Add. Costs for V-Track	Share V $\rightarrow$ College	0.11	0.08
ψ_θ	0.74	Coefficient on θ_5	Regression Coefficient	0.80	0.94
$\mu_{E^p=0}$	0.1	Mean Taste Shock if $E^p = 0$	Share in CL from Non-CL HH	0.20	0.18
$\mu_{\nu,E^p=1}$	-0.1	Mean Taste Shock if $E^p = 1$	Share in CL from CL HH	0.64	0.66
σ_ν	0.001	Std. Taste Shock	Variance of Residual	0.218	0.122
Idiosyncratic Shocks					
σ_ε	0.008	Std. Luck Shock	Std(Log Labor Income)	0.73	0.82
σ_ϕ	0.07	Std. Ability Shock	Var(Test Scores Grade 1)	0.12	0.12
ρ_ϕ	0.65	Persistence of Ability	Test Score Diff. (Grade 1) by E	0.47	0.52
σ_{η_3}	0.07	Std. Learning Shock $j = 3$	$\text{Rank}_{j=2}$-$\text{Rank}_{j=3}$	0.59	0.60
σ_{η_4}	0.06	Std. Learning Shock $j = 4$	$\text{Rank}_{j=3}$-$\text{Rank}_{j=4}$	0.63	0.68
σ_{η_5}	0.05	Std. Learning Shock $j = 5$	$\text{Rank}_{j=4}$-$\text{Rank}_{j=5}$	0.72	0.74
Miscellaneous					
Ω	0.14	Pension Anchor	Replacement Rate	0.40	0.39
A	2.5	TFP	Avrg. Labor Earnings	1.0	1.0
ω	0.54	Weight V. Human Capital	College Share	0.35	0.35

Notes: This table presents the internally calibrated parameters, targeted moments, and their model-generated counterfactuals.

trance, the model can account for the various "second chance" opportunities in the German Education System. In doing so, the college taste shocks play an important role as they ensure that the correlation between parental and child higher education in the model matches the data.

In order to match the correlation between child skill ranks across school periods, the model requires rather large child skill shocks. This is in part, because the estimated own-skill productivity, κ_1, in the child skill formation technology is also quite large. Despite this, the model slightly overstates the differences in initial child skills by parental education prior to entering school. In particular, while children from college-educated parents have an average initial skill level that is around 0.46 standard deviations larger than the average level of non-college-educated parents in the data, this difference is 0.64 standard deviations in the model.

2.5.5 Validation Exercises

We assess the model's validity using two approaches. First, as is standard in the literature, we compare non-targeted moments from our model simulated data to their counterparts in the NEPS data or using estimates from other research papers. Second, we investigate the effects of school track choice on later-in-life economic outcomes for a set of *marginal* students and compare the results to the null-effects reported in Dustmann et al. (2017) for Germany.

Non-targeted Moments

We summarize selected non-targeted moments and their data or external counterparts in Table 2.8. The first set of moments pertains to child skills. While we target the difference (in terms of standard deviations) in average initial child skills prior to entering primary school in the calibration, we do not track how this difference evolves over the school career. In the data, the differences in parental education and school track increase slightly during secondary school.[57]

Similarly, the differences in average child skills across school tracks (in terms of standard deviations) increase throughout secondary school, both in the model and in the data. These differences are around twice as large compared to the differences across parental education.

The second set of moments concerns the relationship between track choice and parental education. In the data, the ratio of college-educated parents who choose an academic track school for their child relative to the average is 1.46. For non-college-educated parents, this ratio is only 0.81. The model implies a slight overestimation of the first and an underestima-

[57] Appendix Table 2.E.1 describes the evolution of child skills over time using the identified latent variables.

tion of the second. Moreover, we regress a dummy variable that equals one if a child attends an academic track school on the percentile rank of the child's skills prior to secondary school, in order to assess the skill gradient in academic track choice. The estimated coefficient is 0.75 in the data and 1.42 in the model. Taken together, these moments suggest that our model somewhat overestimates the importance of child skills for the track choice.

The third set of moments relates to intergenerational mobility. To assess the model's validity here, we compare its implications vis-à-vis the estimates on social mobility in Germany reported in Dodin et al. (2021). Using a different data set than we, they regress a dummy of academic-track school graduation of a child on the percentile income rank of her parents, finding that a 10 percentile increase in the parental rank is associated with a 5.2 percentage point increase in the probability of graduating from an academic track school. In our model, a comparable estimate yields a 4.4 percentage point increase. Moreover, Dodin et al. (2021) report absolute graduation rates for children from the first quintile of the income rank distribution (Q1) of 34%, and a ratio of the fifth income rank quintile over the first quintile of 2.13, which our model matches well. We also compare our model-implied estimate of the intergenerational elasticity of income (IGE) to estimates on German data by Kyzyma and Groh-Samberg (2018). Compared to their findings, the model produces IGEs that are at the lower bound of their data counterparts.

Finally, the model understates the degree of inequality in labor incomes as measured by the Gini coefficient. However, the average college wage premium is consistent with the data.

Long-term effects of Track Choice for Marginal Students

Dustmann et al. (2017) analyse the long-term labor market effects of early school track choice in Germany using a quasi-experimental setting. Their identification strategy makes use of the existence of a (fuzzy) cut-off age for school entry in the German system. Children that are born just before the cut-off age are less likely to go to an academic track secondary school, simply because they are younger at the time of the track decision relative to their class peers. This induces a quasi-randomness in secondary school track choice based on the date of birth. The authors then investigate the effect of that date of birth on later-in-life wages, employment and occupation. They find no evidence that the track attended in secondary school affects these outcomes for the marginal children around the school entry cut-off.[58]

We use our model-simulated data to perform a similar exercise. In particular, we are interested in comparing the later-in-life outcomes of children that are very similar in terms

[58] Note that Dustmann et al. (2017) control for the effect that being born after the cut-off age directly harms a child's later wages since it means that her labor market entry is later, so that at any given age, she will have accumulated less work experience.

Table 2.8: Non-targeted moments

Moment	**Data**	**Model**
Child Skill Moments		
Mean Differences by Parental Background (in Standard Deviations)		
Beginning Secondary School	0.58	0.65
Middle Secondary School	0.70	0.71
Mean Differences by School Track (in Standard Deviations)		
Beginning Secondary School	0.87	0.92
Middle Secondary School	1.01	0.77
School Track Choice		
Relative share A-track children from CL. HH	0.74	0.72
Relative share A-track children from Non-CL HH	0.24	0.25
Coefficient A-track on Skill Rank	0.87	1.02
Intergenerational Mobility		
Parental Income Gradient (Dodin et al., 2021)	0.52	0.50
Q5/Q1 A-track on income (Dodin et al., 2021)	2.13	2.50
Q1 A-track on income (Dodin et al., 2021)	0.34	0.26
IGE (Kyzyma and Groh-Samberg, 2018)	0.27-0.37	0.30-0.33
Inequality - Returns to College		
Gini Coefficient of Labor Income	0.29	0.26
CL/Non-CL Earnings	1.69	1.76

Notes: This table presents the non-targeted moments and their model-generated counterfactuals.

of their state variables at the point of school track choice but end up going to different school tracks. Naturally, in our model, we cannot distinguish the date of birth for children of the same cohort. For that reason, we distinguish children by their skills prior to the secondary school track choice ($\theta_{j=3}$). As detailed in Section 2.4, our child skill development technology implies that, conditional on parental background, the school track choice is characterized by a skill threshold, such that all children with skills above that threshold go to the academic school track and all below go the vocational track school. Conditional on all other states at the time of the track choice – parental human capital, assets, education, and learning ability – differences in child skills and hence differences in school track choice in our model arise from randomly drawn skill shocks. Analogously to Dustmann et al. (2017), we could alternatively argue that these shocks are (at least partly) the result of within-cohort age differences of children, which affect their skill development but are not explicitly modeled. Thus, comparing the later-in-life outcomes of otherwise very similar children with skills around the tracking threshold can be interpreted as estimating the effect of school track choice induced by random (age or skill) shocks.

Concretely, we compare children with skills in a 5% interval around the tracking threshold who go to different school tracks, conditional on all other states.[59] We evaluate these marginal children in terms of their labor income at age 30, the present value of their lifetime labor income, and the present value of their lifetime wealth.[60] We find that going to the academic track instead of the vocational track is associated with a 6.7% higher labor income at age 30, a 2.2% higher present value of lifetime labor income, and a 4.1% higher present value of lifetime wealth.

While not zero, these differences seem rather small in relation to overall inequality in these outcomes. For example, the 2.2% higher present value of lifetime labor income is around 1/20th of a standard deviation of lifetime labor income. Moreover, in our model, the track choice is only between one vocational and one academic track, whereas Dustmann et al. (2017) consider three tracks, of which two can be classified as vocational. We would generally expect that children at the margin of these two vocational tracks show fewer differences in lifetime outcomes. In sum, we conclude that the implications our model entails with respect to the effect of tracking on *marginal* children are not at odds with the reduced-form evidence

[59] This interval amounts to around 1/5 of a standard deviation of child skills prior to the school track choice. We form quintiles of the continuous states of parental human capital and parental assets and allocate children into discrete groups pertaining to these quintiles. Moreover, we partition the distribution of the learning ability ϕ' into three ability states. For these reasons, the skill threshold can become fuzzy in the sense that even conditional on these groups a child with slightly higher skills goes to the vocational track whereas a child with slightly lower skills goes to the academic track.

[60] Lifetime labor income is computed as the discounted sum of all labor income during the adult periods, and lifetime wealth is that sum plus the initial monetary transfer from the parent to their independent child.

presented in Dustmann et al. (2017).

2.6 Quantitative Results

The benefit of our model is that we can use it to understand the effects of school tracking not only for marginal children but for the whole distribution of children, their educational and labor market outcomes, as well as their economic mobility relative to their parents. To that end, we first use our model to quantify the sources of lifetime and inter-generational inequality in the spirit of Huggett et al. (2011) and Lee and Seshadri (2019). Then, we investigate the determinants and consequences of secondary school track choice, as this constitutes the main novelty of our model. In this context, we perform counterfactual analysis of economies in which the school track decision is not affected by an own-track bias of the parents or in which a policymaker enforces a strict tracking skill threshold. Finally, we study the effects of a counterfactual policy reform that postpones the school tracking age to 14.

2.6.1 Sources of Inequality

Using our model, we can decompose how much of the variation in lifetime economic outcomes of our model agents can be explained by various factors at various ages. Following the literature, we focus on lifetime labor income and lifetime wealth as our economic outcomes of interest. We begin by computing the contribution of each state variable of a freshly independent child at age 18 to the variation in lifetime labor income and wealth.[61] These states are the school track in secondary school, S, initial adult human capital h_5, initial transfers received from the parent a_5, parental education E^p, and innate learning ability ϕ.

Row 1 of Table 2.9 summarizes that 70% of the variation in lifetime labor income can be accounted for by all states at the age of 18. In terms of lifetime wealth, this number is around 65%.[62] Thus, our model suggests that the majority of lifetime outcomes is already predetermined when agents become independent and can enter the labor market. Note that at this stage, all uncertainty regarding initial human capital as well as the college decision has been made. The remaining unresolved uncertainty over human capital (market luck) shocks during the working years has, therefore, only limited effects on lifetime inequality.

As Row 2 of Table 2.9 shows, the explained share of variation in lifetime outcomes remains

[61] Concretely, we follow the approach in Lee and Seshadri (2019) and calculate conditional variances of lifetime labor income and wealth, after conditioning on the state variables. As before, we partition the continuous states into three equally sized groups.

[62] These numbers are comparable with estimates for the U.S. (Lee and Seshadri, 2019; Huggett et al., 2011; Keane and Wolpin, 1997)

Table 2.9: Contributions to Lifetime Inequality

Life Stage	States	Share of Explained Variance Lifetime Earnings	Lifetime Wealth
Independence (age 18)	$(S, \phi, h_5, a_5, E, E^p)$	70%	65%
	(S, ϕ, h_5)	63%	60%
	(S, ϕ, a_5, E, E^p)	54%	45%
School Track Choice (age 10)	$(S, \phi', \theta_3, h_{11}, a_{11}, E)$	23%	30%
	(S, θ_3, ϕ')	20%	21%
	(S)	12%	13%
Pre-Birth (parent age 30)	(E, ϕ, h_8, a_8)	10%	20%

Notes: This table shows how much of the variation in lifetime economic outcomes is explained by different factors at different ages.

relatively high if we only condition on the states before the college decision has been made and the inter-vivos transfers have been realized: the child's secondary school track S, her learning ability ϕ and her end-of-school skills that are transformed into initial adult human capital, h_5. This suggests that the size of the parental transfer a_5 and the college choice E, even when affected by parental education E^p are not major sources of lifetime inequality. Instead, if we only exclude initial adult human capital h_5 (Row 3), the share of explained variance in lifetime earnings drops by almost 16 percentage points, and the share of explained variance in lifetime wealth by 20 percentage points. This highlights the importance of variation in initial human capital, and therefore of end-of-school skills, as a driver of lifetime inequality.[63] Interestingly, the correlation between initial adult human capital and transfers received from parents is negative in the model. This suggests that parents partially offset the disadvantage their children experience in the labor market from having lower skills by giving them higher transfers.[64]

Using the same methodology, we can also evaluate how much lifetime inequality is already determined at the time of the school track choice. Conditioning on all states at that age, around 23% of lifetime earnings and 30% of lifetime wealth variation is explained (Row 4). Again, the majority of this variation seems attributable to differences in child states at that age. Yet the explained share is clearly smaller than after school, suggesting that the learning outcomes during secondary school play an important role in shaping later-in-life inequality. Conditioning on the initial school track choice alone can account for 12% of lifetime earnings variation and 13% of lifetime wealth variation. However, this should not be interpreted as

[63] We cannot, however, attribute these drops exclusively to initial human capital differences, given the possible correlation between states.

[64] This channel is also present, albeit to a smaller degree, in Lee and Seshadri (2019).

the marginal effect of school track choice on lifetime outcomes, as the initial school track choice is, for example, highly correlated with child skills at that age. In fact, as we argued in Section 2.5.5, for children with similar skills, the track choice has only small independent effects on lifetime outcomes. We investigate the determinants and consequences of the school track choice in more detail below.

The last row of Table 2.9 shows the contribution of parental states prior to the birth of their children to their children's lifetime outcomes. At this stage, none of the uncertainty regarding child skill and human capital shocks nor regarding the child's learning ability has been realized. Still, around 10% of the variance in lifetime earnings of the yet-to-be-born child is predetermined by parental education, ability, human capital, and wealth. For lifetime wealth, this share is even higher at 20%, pointing to the important role of wealth transfers. For example, using the same decomposition of the unconditional variance of transfers into parental states pre-birth, we find that almost 31% of variation in transfers is predetermined prior to the birth of the child. In contrast, only 16% of the variation in human capital at age 18 is predetermined prior to birth, which highlights the role of shocks to child skills during their childhood and school years.

2.6.2 School Tracking Age Counterfactual Experiments

An important feature of school tracking policies is the age at which children are allocated across the tracks. Generally, OECD countries differ remarkably in the school tracking age (see Figure IV.2.2 in OECD (2013) for an overview). In countries with an early tracking system in place, such as Germany, it is often argued that postponing the tracking age will improve equality of opportunity in terms of access to academic education without incurring efficiency losses in terms of learning outcomes (Woessmann, 2013). While some reduced-form estimates, exploiting cross-country, federal-state level, or time differences in tracking policies exist, little is known about the aggregate, distributional, and inter-generational consequences or welfare effects of a large-scale reform that postpones the tracking age.

To evaluate such a reform in the context of Germany, we conduct a series of counterfactual experiments using our calibrated model, in which we postpone the tracking age from age 10 to age 14, or abolish tracking during secondary school altogether. In each experiment, we assume that in the periods preceding tracking ($j = 3$), all children attend a school that belongs to a comprehensive school track, just like during primary school in $j = 2$. In each counterfactual experiment, all parameters, in particular those governing school track preferences and college costs, remain the same as in the baseline economy.

We present the relative changes of selected aggregate and social mobility outcomes of the counterfactual experiments relative to the baseline economy in Table 2.10. In addition, we calculate the effects on the policies on average welfare. As our welfare measure, we use the percent change in consumption that a newborn in the baseline economy would require to be equally well off as in the policy counterfactual. As is common in the literature, we calculate this consumption equivalence welfare measure under the veil of ignorance, meaning that all policy functions remain unchanged.[65]

The experiments differ in the way we assume that prices and instruction paces are allowed to adjust. In Column (2), all prices (wages per efficiency unit for college and non-college human capital w_0, w_1 and the interest rate r) are assumed to remain at the same values as in the baseline case. That is, we compare the partial equilibrium outcomes of the policy counterfactual. Moreover, we assume that the instruction pace during the *second stage* of secondary school does not adjust. That is, the policymaker sets the same pace as in the baseline case in both academic and vocational track schools during $j = 4$. As a result, parents do not need to form expectations over the average skill levels in each track when they make the postponed track choice.

Postponing School Tracking by Four Years

In this economy, aggregate output Y is around 0.8% lower than in the baseline case. The share of college-educated agents decreases by almost 7%, and the share of children in the academic track in $j = 4$ similarly decreases by 7.4%. Average human capital is significantly less than in the baseline economy, which is ultimately a result of less efficient learning during secondary school. In particular, average end-of-school skills in period $j = 5$ are around 17% lower in the late tracking case than in the baseline economy.

As we derived in Section 2.4, it is theoretically not clear whether later tracking results in such learning efficiency losses. In particular, later tracking could even increase average learning outcomes if the variance of the child skill shocks is sufficiently large. The reason for that is that with large skills shocks, the gain from more homogeneous peer groups in each track during the last stage of secondary school can outweigh the losses incurred due to one more period of learning in a comprehensive track during the first stage of secondary school. However, despite sizable estimates of the child skill shocks variances, our model predicts that the learning losses from postponing four years of tracking in Germany cannot be recuperated by more-efficient learning during the remainder of secondary school.

At the same time, academic track attendance becomes less dependent on the income and education of the parent after the late tracking policy reform. For example, while in

[65] Appendix Section 2.G provides welfare definition.

the early tracking case, 72% of children of college-educated parents go to an academic track school, this number drops to 68% in the late tracking counterfactuals. The share of children from non-college households, who attend an academic track school, however, does not drop relative to the early tracking case, resulting in effectively more children from non-academic households in academic track schools. In a similar vein, the regression coefficient of academic track attendance on parental income decreases by around 13%.

On top of that, the college decision becomes significantly less dependent on the secondary school track in the late tracking counterfactual. Concretely, while the share of academic-track graduates that go to college drops slightly, the share of vocational-track graduates going to college triples. This signals that, in the late tracking counterfactual, the benefits from academic track attendance arising from better chances to go to college are smaller than in the early tracking case. One reason for this result is that late tracking results in a less polarized distribution of end-of-school skills compared to early tracking. For example, the overall variance of end-of-school skills decreases by around 2.5%. Moreover, the difference in average skills between academic and vocational track children decreases by almost 10% in the late tracking counterfactual. Since college utility costs are decreasing in end-of-school skills, this makes the attractiveness of college education become more equal across vocational and academic school track graduates.

As a consequence, both cross-sectional inequality, as measured by the Gini coefficient of labor income, and the intergenerational elasticity of earnings decrease in the late tracking counterfactual. Thus, our quantitative exercise suggests that postponing tracking results in efficiency losses in terms of learning and aggregate output but comes with the benefit of reduced inequality and improved social mobility. This result can be viewed in a similar spirit to the efficiency-mobility trade-off in **?**, who has shown that desegregation policies may entail penalties in terms of growth.

If we allow the instruction pace in each track to adjust endogenously while still keeping prices at their baseline values (in Column (3)), these two opposing effects become slightly more pronounced. For example, the share of children ending up in the academic track school drops by over 9%, and the share of college workers drops by around 8% relative to the baseline, early tracking economy. Aggregate learning also decreases more, resulting in an output loss of almost 1% in this economy. We can interpret this result again through the lens of the theoretical illustrations derived in Section 2.4. In particular, we have argued before that the equilibrium allocation of children across school tracks that results from a game played among parents need not be equal to the optimal one that a policymaker seeking to maximize learning would implement if there are positive direct peer effects. Against this backdrop, the results in Columns 2 and 3 of Table 2.10 then suggest that the unadjusted instruction pace carried

Table 2.10: Timing of Tracking Counterfactual Experiments - Results

	(1)	(2)	(3)	(4)	(5)
Tracking Age	10	14	14	14	Never
Instruction Pace	Baseline	Baseline	Adjusts	Adjusts	Adjusts
Wages	Baseline (GE)	PE	PE	GE	GE
Outcome	Panel A.				
Y	2.05	-0.8%	-0.95%	-0.2%	-0.5%
College Share	0.35	-6.9%	-8.1%	0.0%	-4.3%
A-Track Share	0.42	-7.4%	-9.3%	-5.5%	-
CL/Non-CL Earnings	1.773	0.0%	0.0%	-0.2%	5.0%
Gini Earnings	0.26	-0.4%	-0.8%	0.0%	0.0%
$Pr(S=1\|E^p=1)$	0.72	-5.6%	-6.9%	-6.9%	-
$Pr(S=1\|E^p=0)$	0.25	-0.4%	-2.8%	-0.8%	-
$Pr(E=1\|S=1)$	0.71	-2.8%	-4.2%	-1.4%	-
$Pr(E=1\|S=0)$	0.08	200%	187%	225%	-
$Pr(E=1\|E^p=1)$	0.66	-3.9%	-4.6%	-1.7%	-13.2%
$Pr(E=1\|E^p=0)$	0.18	-4.3%	-4.3%	1.6%	15.2%
$Pr(S=A)$ on Income	0.50	-13.4%	-15.4%	-14%	-
IGE	0.31	-1.9%	-2.0%	-1.9%	-2.1%
Welfare (Cons. Equiv.)	-	-0.5%	-0.6%	-0.1%	-0.5%
	Panel B. (Differences in Log Skills)				
$\bar{\theta}_5$	0.04	-17.1%	-20.0%	-5.7%	-0.7%
$Var(\theta_5)$	0.12	-0.3%	-0.3%	-0.1%	0.04%
$Std(\theta_{5\|S=V})$	0.39	1.0%	0.8%	-0.8%	-
$Std(\theta_{5\|S=A})$	0.40	-6.9%	-6.9%	-2.0%	-

Notes: This table presents changes in outcomes due to delaying the school tracking choice by four years (from the age of ten to the age of fourteen). Column (1) shows aggregate outcomes in the baseline economy, and Columns (2) to (4) display percentage changes due to the policy change in different scenarios. Column (2): if the pace of instruction and prices are unchanged. Column (3): if the pace of instruction adjusts but prices are unchanged. Column (4): if the pace of instruction and prices adjust. Column (5):

over from the early tracking economy is actually closer to the one a policymaker would optimally choose than the adjusted one resulting from parents making the track decision while correctly anticipating the distribution of child skills across tracks in equilibrium. Thus aggregate learning drops. At the same time, the effects on intergenerational mobility and cross-sectional earnings inequality remain approximately the same as before.

Naturally, some of the loss in efficiency when postponing tracking could be due to the fact that, in the partial equilibrium late tracking counterfactuals, the share of college-educated workers overall declines markedly, which impacts aggregate human capital during the working years as this is assumed to grow at a college-specific rate $\gamma_{j,E}$.

Once we allow for general equilibrium effects in Columns (4), the college share returns to be approximately the same as in the baseline economy, as wages adjust to clear the labor markets for college and non-college type labor. As college education becomes more attractive again, also the share of children in the academic track rises. Nonetheless, it is (at around 40%) still slightly lower than in the early tracking baseline economy. Moreover, postponing tracking still decreases the average end-of-school skills relative to early tracking (by 5.7%), yet markedly less so than in the partial equilibrium cases.

Thus, despite the fact that the variances of end-of-school skills in each school track in the late tracking counterfactual are smaller than in the baseline economy, this gain in homogeneity in peer groups cannot overcome the disadvantage in terms of average skills stemming from one more period of comprehensive track schooling. As a result, total output still decreases by 0.2% relative to the early tracking economy.

On the other hand, the resulting general equilibrium continues to feature more mobility between generations, as the intergenerational earnings elasticity drops by almost 2%. This is again a consequence of significantly more children going to college after a vocational track secondary school and a declining share of children from academic parents going to academic track schools. Perhaps surprisingly, cross-sectional inequality as measured by the Gini coefficient on labor income, does not change relative to the early tracking case.

The main takeaways of the policy reform that postpones school tracking to age 14 in our model can be summarized as follows. First, postponing school tracking incurs efficiency losses from worse learning outcomes in the additional period of comprehensive school. The losses cannot be compensated by gains in later years that arise from more homogeneous peer groups across tracks as the track decision is based on more complete information about children's skill evolution. Second, later tracking incentivizes fewer parents to send their child to an academic track secondary school as the likelihood of college education depends less on the secondary school track.

Third, this results in more equal access to academic secondary education by parental background, which leads to more equal access to higher education and more equal labor market outcomes. The quantitative size of this effect depends on whether the school tracking age reform is evaluated in the short run, when wages and possibly the instruction paces in schools have not reacted, or in the long run, when general equilibrium effects are taken into account. Finally, in all cases, later tracking reduces the persistence of economic status across generations, inducing an efficiency-mobility trade-off.

Abolishing School Tracking

Column (5) of Table 2.10 reports the results of a counterfactual economy, in which we abolish tracking altogether while allowing wages and the interest rate to adjust. All children go to comprehensive schools for the entirety of their schooling years, and instruction occurs at the same pace that is optimal for the overall average skill level. As a consequence, the correlation between parental background and end-of-school skills drops sharply (-26.2%). As a result, and despite college-specific preferences, mobility in higher education also increases; children from non-college parents are more likely to graduate from college than in the baseline economy, and children from college parents are less likely to do so. Overall, mobility as measured by the (negative of the) intergenerational income elasticity improves significantly (+2.1%).

On the other hand, abolishing tracking altogether makes learning even less efficient relative to the late tracking economy. In particular, we see large losses in aggregate human capital, leading to a decrease in output (-0.5%).

Thus, despite considerable mobility gains, a completely comprehensive school system worsens average welfare by 0.05%.

2.6.3 School Track Allocation Counterfactuals

According to the theoretical predictions laid out in Section 2.4, the initial school track should, to a large degree, be based on child skills. A regression of an academic school track dummy on all states at the time of the tracking decision using model-generated data confirms that this is true in our model. Column 1 of Table 2.11 reports the standardized coefficient estimates of this regression, indicating that child skills at the time of the track choice, θ_3 have the strongest impact on the track decision. In particular, increasing log child skills by one standard deviation increases the probability of going to the academic track by 53 percentage points.

Notwithstanding this, Column 1 in Table 2.11 also indicates that parental education is the second most important independent driver of the school track choice. In the model, parental education can influence the track choice, net of the effects coming through child skills, human capital, or wealth, in three ways. First, college-educated parents know that their children learn faster than their non-college-educated counterparts. This comes from the estimated direct parental education effect in the child skill production technology, κ_5. This knowledge may prompt college parents to send their child to the academic track even if their child's skills are lower than those of a child from a vocational parent. Second, parents know that their child will receive a college taste shock that depends on their parent's education, governed by μ_{ν,E^p}. In anticipation of this, college parents, for instance, may have a stronger incentive to send their child to an academic track school as this, everything else equal, increases the likelihood of college admission. However, (non-pecuniary) college costs also decrease in end-of-school skills. As derived in Section 2.4, for a set of children with low preschool skills, end-of-school skills are maximized if they attend the vocational school track. This force counteracts the incentive of college parents to send their child to the academic track described before. Third, even net of college tastes, we assume that parents bias the school track choice towards their own education level. We motivated this bias by the significant number of deviations from teacher recommendations in the school track choice. The bias then directly implies a stronger direct effect of parental education on the school track choice.

Table 2.11: School Track Choice Determinants

	Dependent Variable: $S = A$			
	Stand. Coefficient Estimates			
	(1)	(2)	(3)	(4)
	Baseline	$\omega_{5,j=3,4} = 0$	$\mu_{\nu,1} = \mu_{\nu,0} = 0$	$\chi_0 = \chi_1 = 0$
ϕ'	0.09	0.10	0.10	0.10
θ_3	0.53	0.54	0.58	0.56
$E = 1$	0.34	0.24	0.23	0.21
h_{11}	0.00	-0.01	0.00	0.00
a_{11}	0.02	0.01	0.02	0.02

Notes: This table reports the standardized coefficient estimates of regressions of an academic school track dummy on all states at the time of the tracking decision. Column (1) corresponds to the baseline economy. In Column (2), we shut down the channel of differential parental inputs in periods 3 and 4. Column (3) considers the case of identical college taste shock by parental education. In Column (4), we remove the parental preference bias for education.

To understand how important each of these channels for the school track choice is, we perform a series of three counterfactual experiments using the calibrated model, in which we isolate each effect, respectively.[66] In particular, we isolate the effects of the first channel

[66] In doing so, we again solve for the stationary general equilibrium allowing prices to clear the markets and average child skills across tracks to be consistent with the parents' track decision.

by solving the model with $\kappa_{5,j=3,4} = 0$ yet leaving $\kappa_{5,j=3,4} > 0$ in the simulation of the distribution. That is, we assume that parents do not take into account the direct effect of their own education on child skill development during secondary school when making the track decision. The skills, however, still evolve as in the baseline model. Column (2) in Table 2.11 reports the (standardized) results of the regression of academic track choice on all state variables in this counterfactual scenario. The coefficient on parental education drops as expected, while the coefficient on child skills prior to the track decision increases. This confirms that the knowledge of direct parental effects on future child skill development prompts parents to send their child to the same track as their own, net of effects of parental education through child skills that are already formed. The magnitude of this channel, however, seems relatively small. In particular, the results suggest that this channel accounts for around 8.5% of the direct effect of parental education on the probability of academic school track attendance of her child.

Column (3) reports the resulting coefficient estimates when isolating the second channel, working through college tastes. If we equalize the means in college taste shocks across parental education (to zero), once again, the coefficient on direct parental influence on school track choice decreases, and the one on child skill increases. Quantitatively, these effects are comparable to the first channel. Similarly, as reported in Column (4) of Table 2.11, the direct influence of parental education on the school track of a child drops by almost 40% if we shut down parental preferences in school track choice directly by setting $\chi_E = 0$ for both education levels. At the same time, a child's own skills become more important for the track decision.

As discussed in Section 2.4, any such forces that impact the school track allocation net of child skills can, in theory, be detrimental to the efficiency of skill development in secondary school if they dilute the homogeneity of peer groups in each track.[67] An important question is whether the consequences of such "misallocation" effects are visible not only in terms of child skill outcomes but also in the aggregate and distributional outcomes in the economy. Our model provides a suitable environment to investigate such effects. Table 2.12 provides an overview of selected outcomes in the baseline model (Column (1)) and compares the resulting percentage change of these outcomes in two counterfactual scenarios: In Column (2), we shut down direct parental preferences shifting the school track choice ($\chi_0 = \chi_1 = 0$) as before.[68]

[67] Suppose for example, college-educated parents send their children to an academic track school, despite the fact that their skill level would optimally suggest the vocational track. In that case, this will not only harm their child's development but also cause the instruction pace in that track to adjust. This, in turn, harms the average learning gains of everyone in that track. The same effect occurs in the vocational track school if parents from non-college backgrounds send their overqualified children there purely based on preferences.

[68] We focus on this experiment as we view this as being the easiest to address by policies. In particular, if

Moreover, in Column (3), we report the relative changes in the outcomes from another counterfactual experiment, in which we enforce that the school track allocation is governed exclusively by a sharp skill threshold, such that all children with skills below the threshold are allocated to the vocational track, while all children with skills above the threshold go to the academic track, regardless of the parental background. This threshold is chosen, such that the overall share of children in the academic track is constant relative to the baseline economy.[69]

In both counterfactual scenarios, aggregate output in the economy increases slightly relative to the baseline economy. Note that the share of college-educated agents in the economy remains constant relative to the baseline case, which is due to the general equilibrium effects on the labor market as wages adjust to keep demand for college and non-college labor approximately constant. In contrast, the share of children that attend an academic track school increases. In the case without preference-based school track choice, the share increases by 2.4%. By construction, this share increases even further in the case of the sharp track threshold, as this threshold implies that roughly 50% of the children go to either track. The reason for the positive effects on output becomes clear when we study the distribution of skills in counterfactual experiments.

In particular, the first row in Panel B. of Table 2.12 suggests that both counterfactual scenarios lead to an increase in average child skills at the end of secondary school. This increase arises from the fact that the variance in child skills within the school tracks changes relative to the baseline case, which impacts learning efficiency. In the first counterfactual experiment, skills in the academic track become more homogeneous, while the variance of skills in the vocational track increases only marginally. In the second counterfactual, the variance of skills in the vocational track decreases while the variance of skills in the academic track increases. The latter is likely due to the fact that the share of children in the academic track also increases. Overall, however, the effect on average end-of-school skills is positive, which then translates into higher output. This is consistent with the explanation of the efficiency-reducing misallocation effects that arise when parental background drives the school

preferences for school tracks are coming from information frictions, as argued before, mentoring programs have proven very effective and almost cost-free in alleviating some of these frictions as argued by Falk et al. (2021) and Resnjanskij et al. (2021).

[69] As derived in Section 2.4, the optimal tracking policy from the point of view of a policymaker who is only interested in maximizing aggregate end-of-school skills and cannot condition on the parental background, would be to track children at a threshold that is exactly equal to the average child skill level prior to the track decision. Given that the distribution of child skills is quite symmetric, this would result in a roughly equal split of children between tracks, which ensures that the variance of child skills in each track is minimized. However, to be comparable to the baseline economy, we select a threshold that will result in the top 42% of children in terms of their skills being allocated to the academic track and the rest to the vocational track.

track choice, independently from skills.

Row 4 of Panel A. reports that without direct parental preferences in school track choice and even more so with a sharp, purely skill-based allocation rule, the dependence of school track choice on parental income decreases. Unsurprisingly, skills themselves become more important in explaining the track choice and the college choice, as shown in Row 5. However, while the intergenerational elasticity between parent's and child's income drops in the first counterfactual experiment, it slightly increases when introducing a strict skill threshold. Again, this is likely due to the fact that the share of academic track children also increases in that case.

Table 2.12: Effects of School Track Choice Counterfactuals

Outcome	(1) Baseline Economy	(2) $\chi_0 = 0$ $\chi_1 = 0$	(3) Skill Threshold
	Panel A.		
Y	2.05	0.1%	0.3%
College Share	0.35	0%	-1.4%
A-Track Share	0.42	2.4%	0%
$Pr(S = A)$ on Income	0.50	-25.2%	-37.2%
$Pr(S = A)$ on Skills	1.02	1.4%	42.4%
$Pr(E = 1)$ on Skills	0.94	0.4%	2.2%
$Pr(E = 1\|S = V)$	0.08	13.4%	-3.7%
$Pr(E = 1\|E^p = 0)$	0.18	6.0% 10.9%	
IGE	0.31	-2.3%	-0.3%
Gini Earnings	0.26	0.4%	0.8%
Welfare	-	-0.2%	-0.1%
	Panel B.		
$\bar{\theta}_5$	0.04	0.1%	8.6%
$Std(\theta_{3\|S=V})$	0.25	-1.2%	-34.0%
$Std(\theta_{3\|S=A})$	0.38	-0.9%	-18.8%

Notes: Column (1) shows aggregate outcomes in the baseline economy, Column (2) displays percentage changes entailed by the absence of parental preference for education, and Column (3) displays percentage changes entailed by skill treshold-rule for school tracking. All results are coming from the new general equilibrium distribution.

Overall, Table 2.12 paints the following picture. Both counterfactual scenarios achieve an improvement in child learning during the secondary school years. This improvement yields a positive effect on aggregate output in the macroeconomy, which is larger when the track allocation is based on a pure skill threshold, though still relatively modest at 0.2%. At the same time, while decreasing parental track choice preferences improves social mobility relative to the baseline economy, this cannot be said about the case with an optimal, threshold-based track allocation.

2.7 Conclusion

How important is the design of education policies for the macroeconomic analysis of inequality and social mobility? This paper argues that school tracking, a standard policy across many advanced countries, influences not only equality of educational opportunities for children from different parental backgrounds but also shapes aggregate learning and, as a consequence, aggregate economic efficiency. We add a macroeconomic perspective to the predominantly reduced-form literature by building a macroeconomic GE model of overlapping generations that specifically zooms in on the children's schooling years. To that end, we formulate a simple theory of child skill formation, where child skills depend linearly on her classroom peers and non-linearly on the instruction pace specific to each school track.

We show that this child skill formation technology alone entails theoretical implications for the effect of school tracking policies on the distribution of child skills that align with the most robust findings of a vast empirical literature and the most popular arguments in the public debate about tracking. In particular, not every child gains from tracking; the losses are often concentrated among lower-skilled children. Additionally, tracking can lead to increased inequality in end-of-school skills. Finally, the effects of tracking on learning efficiency, while typically positive on average, depend on the age at which children are tracked and the size of uncertainty regarding the evolution of child skills, highlighting the importance of the timing of tracking.

We embed this theory into a standard Aiyagari-style life-cycle framework in which parents make a school track decision for their children. We tailor the model to fit the German Education System, where the track decision occurs at the age of 10 of the child, and calibrate it on German data. Our quantitative results suggest that variation from the initial school track alone can account for around 12% of the variation in eventual lifetime earnings. Conditional on prior child skills, the track choice is strongly influenced by parental preferences that cannot be explained by parental inputs into child skills or tastes for higher education. This gives rise to efficiency-reducing misallocation of children across tracks. Our results indicate that policies that reduce the parental influence on the school track choice, such as mentoring policies (Falk et al., 2021), can, therefore, not only improve social mobility but also lead to modest efficiency gains in terms of aggregate output in the macroeconomy.

Our paper also shows that the timing of the school tracking age entails a macroeconomic trade-off between efficiency and social mobility. Concretely, a policy reform that delays the school tracking decision by four years (to age 14) in Germany leads to aggregate output losses, in the long run, that amount to around 0.2% of GDP while decreasing the intergenerational income elasticity by around 2%, thereby improving social mobility. Key in the evaluation of this trade-off is the consideration of general equilibrium effects in the labor market that affect the incentives governing the school track choice. The output losses from this reform fundamentally stem from learning efficiency losses due to the prolonged time of comprehensive schooling. At the same time, the social mobility gains result from the track decision depending less on the parental background and the college decision depending less on the secondary school track.

Bibliography

Aakvik, A., Salvanes, K. G., and Vaage, K. (2010). Measuring heterogeneity in the returns to education using an education reform. *European Economic Review*, 54(4):483–500.

Abbott, B., Gallipoli, G., Meghir, C., and Violante, G. L. (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, 127(6):2569–2624.

Agostinelli, F. (2018). Investing in children's skills: An equilibrium analysis of social interactions and parental investments. *Unpublished Manuscript, University of Pennsylvania.*

Agostinelli, F., Doepke, M., Sorrenti, G., and Zilibotti, F. (2023). It takes a village: the economics of parenting with neighborhood and peer effects. Working Paper w27050, National Bureau of Economic Research.

Agostinelli, F., Saharkhiz, M., and Wiswall, M. (2019). Home and school in the development of children. Working Paper w26037, National Bureau of Economic Research.

Agostinelli, F. and Wiswall, M. (2016). Estimating the technology of children's skill formation. Working Paper w22442, National Bureau of Economic Research.

Ballou, D. (2009). Test scaling and value-added measurement. *Education finance and Policy*, 4(4):351–383.

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

Bellenberg, G. and Forell, M. (2012). Schulformwechsel in deutschland. *Durchlässigkeit und Selektion in den*, 16.

Blossfeld, H., Roßbach, H., and von Maurice, J. (2011). The german national educational panel study (neps). *Zeitschrift für Erziehungswissenschaft: Sonderheft*, 14.

Bonesrønning, H., Finseraas, H., Hardoy, I., Iversen, J. M. V., Nyhus, O. H., Opheim, V., Salvanes, K. V., Sandsør, A. M. J., and Schøne, P. (2022). Small-group instruction to improve student performance in mathematics in early grades: Results from a randomized field experiment. *Journal of Public Economics*, 216:104765.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of human Resources*, 43(4):972–1059.

Capelle, D. (2022). The great gatsby goes to college: Tuition, inequality and intergenerational mobility in the u.s. Working paper.

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

Ciccone, A. and Peri, G. (2005). Long-run substitutability between more and less educated workers: evidence from us states, 1950–1990. *Review of Economics and statistics*, 87(4):652–663.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American economic review*, 97(2):31–47.

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

Daruich, D. (2022). The macroeconomic consequences of early childhood development policies. Working Paper 2018-29, FRB St. Louis.

Dodin, M., Findeisen, S., Henkel, L., Sachs, D., and Schüle, P. (2021). Social mobility in germany. Discussion Paper DP16355, CEPR.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74.

Dustmann, C. (2004). Parental background, secondary school track choice, and wages. *Oxford Economic Papers*, 56(2):209–230.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2017). The long-term effects of early track choice. *The Economic Journal*, 127(603):1348–1380.

Epple, D. and Romano, R. (2011). Peer effects in education: A survey of the theory and evidence. In *Handbook of social economics*, volume 1, pages 1053–1163. Elsevier.

Falk, A., Kosse, F., and Pinger, P. (2021). Mentoring and schooling decisions: Causal evidence. Working Paper 8382, CESifo.

Fuchs-Schündeln, N., Krueger, D., Kurmann, A., Lale, E., Ludwig, A., and Popova, I. (2023). The fiscal and welfare effects of policy responses to the covid-19 school closures. *IMF Economic Review*, pages 1–64.

Fuchs-Schündeln, N., Krueger, D., Ludwig, A., and Popova, I. (2022). The long-term distributional and welfare effects of covid-19 school closures. *The Economic Journal*, 132(645):1647–1683.

Fujimoto, J., Lagakos, D., and Vanvuren, M. (2023). Aggregate and distributional effects of 'free' secondary schooling in the developing world. Working Paper w31029, National Bureau of Economic Research.

Guyon, N., Maurin, E., and McNally, S. (2012). The effect of tracking students by ability into different schools a natural experiment. *Journal of Human resources*, 47(3):684–721.

Hanushek, E. A. and Wössmann, L. (2006). Does educational tracking affect performance and inequality? differences-in-differences evidence across countries. *The Economic Journal*, 116(510):C63–C76.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics*, 132(4):1693–1754.

Heckman, J. and Mosso, S. (2014). The economics of human development and social mobility. *Annu. Rev. Econ.*, 6(1):689–733.

Henninges, M., Traini, C., and Kleinert, C. (2019). Tracking and Sorting in the German Educational System. Working Paper 83, Leibniz Institute for Educational Trajectories (LIfBi).

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

Jang, Y. and Yum, M. (2022). Aggregate and Intergenerational Implications of School Closures: A Quantitative Assessment. Working Paper 234v1, CRC TR 224.

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

Kindermann, F., Mayr, L., and Sachs, D. (2020). Inheritance taxation and wealth effects on the labor supply of heirs. *Journal of Public Economics*, 191:104127.

Kyzyma, I. and Groh-Samberg, O. (2018). "intergenerational economic mobility in germany: Levels und trends ". Working paper, DIW.

Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2018). Life cycle wage growth across countries. *Journal of Political Economy*, 126(2):797–849.

Larson, K., Russ, S. A., Nelson, B. B., Olson, L. M., and Halfon, N. (2015). Cognitive ability at kindergarten entry and socioeconomic status. *Pediatrics*, 135(2):e440–e448.

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

Malamud, O. and Pop-Eleches, C. (2011). School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics*, 95(11-12):1538–1549.

Matthewes, S. H. (2021). Better together? heterogeneous effects of tracking on student achievement. *The Economic Journal*, 131(635):1269–1307.

Meghir, C. and Palme, M. (2005). Educational reform, ability, and family background. *American Economic Review*, 95(1):414–424.

Nennstiel, R. (2022). No matthew effects and stable ses gaps in math and language achievement growth throughout schooling: Evidence from germany. *European sociological review*.

Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., and Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal for educational research online*, 5(2):80–109.

OECD (2013). Pisa 2012 results: What makes schools successful (volume iv): Resources, policies and practices, pisa. Report TD/TNC 114.1481, OECD.

Passaretta, G., Skopek, J., and van Huizen, T. (2022). Is social inequality in school-age achievement generated before or during schooling? a european perspective. *European Sociological Review*, 38(6):849–865.

Pekkala Kerr, S., Pekkarinen, T., and Uusitalo, R. (2013). School tracking and development of cognitive skills. *Journal of Labor Economics*, 31(3):577–602.

Piopiunik, M. (2014). The effects of early tracking on student performance: Evidence from a school reform in bavaria. *Economics of Education Review*, 42:12–33.

Pohl, S. and Carstensen, C. H. (2012). Neps technical report-scaling the data of the competence tests. Working Paper 14, NEPS.

Pohl, S. and Carstensen, C. H. (2013). Scaling of competence tests in the national educational panel study-many questions, some answers, and further challenges. *Journal for Educational Research Online*, 5(2):189–216.

Resnjanskij, S., Ruhose, J., Wiederhold, S., and Woessmann, L. (2021). Can mentoring alleviate family disadvantage in adolscence? a field experiment to improve labor-market prospects.

Ruhose, J. and Schwerdt, G. (2016). Does early educational tracking increase migrant-native achievement gaps? differences-in-differences evidence across countries. *Economics of Education Review*, 52:134–154.

Schneider, T. and Linberg, T. (2022). Development of socio-economic gaps in children's language skills in germany. *Longitudinal and Life Course Studies*, 13(1):87–120.

Woessmann, L. (2013). Die entscheidende säule. *Wirtschaftswoche Global*, 2(24.06.2013):110–111.

Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *Journal of Economic Perspectives*, 30(3):3–32.

Yum, M. (2022). Parental time investment and intergenerational mobility. *International Economic Review*.

Appendices

2.A Proof of Propositions

Proposition 1

First, we show that maximizing the aggregate end-of-school skills in a tracking system implies a threshold skill level $\tilde{\theta}_1$, such that all $\theta_1 < \tilde{\theta}_1$ go to one track, call it $S = V$ and all $\theta_1 > \tilde{\theta}_1$ go to the other track, $S = A$ (and those with $\theta_1 = \tilde{\theta}_1$ are indifferent). That is, the existence of a skill threshold is a necessary condition for optimal end-of-school skills. We restrict ourselves to the case with different instruction paces across school tracks.

To that end, it is useful to rewrite θ_2 in (2.24) of a child in a given school track S with instruction pace P_S^* using Lemma 1 as:

$$\theta_2 = \theta_1\ (1+\beta) + \alpha\bar{\theta}_S + \frac{\beta^2}{2\gamma} + \gamma\theta_1\bar{\theta}_S - \frac{\gamma}{2}\bar{\theta}_S^2 + \eta_2. \tag{2.28}$$

After adding and subtracting $\frac{\gamma}{2}\theta_1^2$, this can be expressed as

$$\begin{aligned}\theta_2 &= \theta_1\ (1+\beta) + \alpha\bar{\theta}_S + \frac{\beta^2}{2\gamma} + \frac{\gamma}{2}\theta_1^2 - \frac{\gamma}{2}\left(\theta_1^2 - 2\theta_1\bar{\theta}_S + \bar{\theta}_S^2\right) + \eta_2 \\ &= \theta_2(p_{\theta_1}^*) - \frac{\gamma}{2}(\theta_1 - \bar{\theta}_S)^2,\end{aligned} \tag{2.29}$$

where $\theta_2(p_{\theta_1}^*)$ denotes end-of-school skills if the child with skills θ_1 is taught at her individually optimal teaching pace $p_{\theta_1}^*$. Thus, in a given track, end-of-school skills are a strictly decreasing function of the *distance* to the average skill level $\bar{\theta}_S$ in that track. This is intuitive given Lemma 1, as it is solely the average skill level to which the instruction pace is optimally targeted.

Next, assume for contradiction that the expected value of end-of-school skills across tracks $\mathbb{E}[\theta_2]$ is maximized under a track allocation mechanism that does not feature a skill threshold. Suppose that $P_V^* < P_A^*$ without loss of generality. By Lemma 1, these are the optimal instruction paces for the average skill level in track V and A, respectively. Therefore, $\mathbb{E}(\theta_1|S = V) < \mathbb{E}(\theta_1|S = A)$. Then, because there is no strict threshold, this means that for any initial skill level θ_1, there must be at least two children with initial skill levels smaller or equal to θ_1 that go to different tracks or at least two children with initial skill levels larger or equal than θ_1 that go to different tracks. This implies that there exists a child with $\theta_1' \leq \mathbb{E}(\theta_1|S = V)$ that goes to track $S = A$, and/or a child with $\theta_1' \geq \mathbb{E}(\theta_1|S = A)$ that goes to track $S = V$, and/or two children with skills $\theta_1' < \theta_1''$, with $\theta_1', \theta_1'' \in [\mathbb{E}(\theta_1|S = V), \mathbb{E}(\theta_1|S = A)]$, where the child with the smaller skill level goes to track A and the child with the larger skill level to track V.

However, given the condition in (2.29), this child with θ_1' would always benefit from being in the other track as the distance between her skill level and the average skill level in that track is smaller than in the track she is in. Note that moving just one child to another track does not change the average skills in both tracks. Thus, the policymaker can improve aggregate end-of-school skills by moving this child.

The same line of argument holds in the implied game that parents play when they endogenously sort their children into two tracks. If no skill threshold level exists, there is always a child that would unilaterally gain if sent to a different track.

Thus, we have established that the existence of a skill threshold is necessary for optimal end-of-school skills both if a policymaker makes the track allocation directly and when parents play a sorting game. Next, we characterize the thresholds for both cases. Let $\tilde{\theta}_1$ be the skill threshold and let S again indicate to which track a child is allocated, now with $S = V$ for all $\theta_1 \leq \tilde{\theta}_1$ and $S = A$ for all $\theta_1 > \tilde{\theta}_1$.

A policymaker solves

$$\begin{aligned} & \max_{\tilde{\theta}_1} \quad \mathbb{E}(\theta_2) \\ \iff & \max_{\tilde{\theta}_1} \quad \mathbb{E}(\mathbb{E}(\theta_2|S)) \\ & \text{subject to} \\ & \qquad P_S \text{ chosen optimally given Lemma 1.} \end{aligned} \tag{2.30}$$

Using (2.28) and the law of iterated expectations, this maximization problem boils down to

$$\begin{aligned} & \max_{\tilde{\theta}_1} \quad \frac{\beta^2}{2\gamma} + \frac{\gamma}{2}\,\mathbb{E}\left(\mathbb{E}(\theta_1|S)^2\right) \\ \iff & \max_{\tilde{\theta}_1} \quad \frac{\beta^2}{2\gamma} + \frac{\gamma}{2}\left(F(\tilde{\theta}_1)\,\mathbb{E}(\theta_1|\theta_1 \leq \tilde{\theta}_1)^2 + (1 - F(\tilde{\theta}_1))\,\mathbb{E}(\theta_1|\theta_1 > \tilde{\theta}_1)^2\right), \end{aligned} \tag{2.31}$$

where $F(.)$ denotes the cumulative distribution function of the normal distribution. Note that the right term is just the expected value (across tracks) of the conditional expected values of initial skills squared, conditional on the school track. This corresponds to the variance of the conditional expected values, which depend on the skill threshold $\tilde{\theta}_1$. Using the law of total variance, the maximization problem can thus be rewritten as (dropping the constant term)

$$\begin{aligned} &\max_{\tilde{\theta}_1} \quad \mathbb{E}(\theta_2) \\ \iff &\max_{\tilde{\theta}_1} \quad \frac{\gamma}{2}\left(\sigma^2_{\theta_1} - \mathbb{E}(Var[\theta_1|S])\right). \end{aligned} \tag{2.32}$$

Thus, the policymaker chooses optimally a threshold such that the expected variance of skills in each track is minimized. The unique solution is then to set $\tilde{\theta}^*_1 = \mathbb{E}\,\theta_1 = 0$, that is, to split the distribution exactly in half. This makes the peer groups in each track as homogeneous as possible, which maximizes average and aggregate learning.

Next, we characterize the threshold that arises endogenously from the sorting game played by the parents. The equilibrium condition maintains that at this threshold, a parent is just indifferent between tracks as her child's skills would be equivalent in both tracks. A parent of a child with skill $\hat{\theta}_1$ is indifferent between tracks V and A iff

$$\begin{aligned} &\left(\alpha + \gamma\,\hat{\theta}_1\right)\mathbb{E}(\theta_1|\theta_1 \le \hat{\theta}_1) - \frac{\gamma}{2}\,\mathbb{E}(\theta_1|\theta_1 \le \hat{\theta}_1)^2 \\ &= \left(\alpha + \gamma\,\hat{\theta}_1\right)\mathbb{E}(\theta_1|\theta_1 > \hat{\theta}_1) - \frac{\gamma}{2}\,\mathbb{E}(\theta_1|\theta_1 > \hat{\theta}_1)^2 \\ \iff &\left(-\alpha - \gamma\,\hat{\theta}_1\right)\sigma_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)}{F(\hat{\theta}_1/\sigma)} - \frac{\gamma}{2}\sigma^2_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)^2}{F(\hat{\theta}_1/\sigma)^2} \\ &= \left(\alpha + \gamma\,\hat{\theta}_1\right)\sigma_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)}{1 - F(\hat{\theta}_1/\sigma)} - \frac{\gamma}{2}\sigma^2_{\theta_1}\frac{f(\hat{\theta}_1/\sigma)^2}{(1 - F(\hat{\theta}_1/\sigma))^2} \end{aligned} \tag{2.33}$$

where $F(\cdot)$ denotes the CDF of a standard normally distributed random variable, and $f(\cdot)$ is its density function. We solve for the root $\hat{\theta}_1$ that solves (2.33) numerically. In all cases with reasonable parameter values, (2.33) is a monotone function, such that the root is unique if it exists. In the special case without direct peer externality, i.e., $\alpha = 0$, the solution is $\hat{\theta}_1 = 0$, as can be directly seen from (2.33). When $\alpha > 0$, the root is smaller than 0, i.e. $\hat{\theta}_1 < 0$.

Proposition 2

The proof of this Proposition follows directly from (2.A). In a comprehensive system, the variance of initial skills across tracks is just equal to the overall variance since there is only one track. In a tracking system, the expected value of the conditional variances of skills across tracks is smaller than the overall variance, by the law of total variance and provided that the instruction paces are different across tracks. This holds for every skill threshold, not just for the optimal one. Thus average learning is higher.

Next, we show that a full tracking system leads to a "fatter" right tail of the end-of-school skill distribution compared to a comprehensive system. To see this, consider the child who, in expectation, has the highest end-of-school skill in a comprehensive system. Since θ_2 is monotonically increasing in θ_1 in a given track (see (2.28)), this is the child with the highest initial skill, say $\theta_{1,max}$. Moreover, from the properties of a truncated normal distribution, we know that, for any skill threshold $\tilde{\theta}_1$, average skills in the A track, $\bar{\theta}_{1,A}$ are larger than the unconditional average, $\bar{\theta}_{1,C} = 0$. Thus, the squared distance between $\theta_{1,max}$ and $\bar{\theta}_{1,A}$ in a tracking system is smaller. Taken together, (2.29) implies that the child with initial skill $\theta_{1,max}$ ends up with larger end-of-school skills compared to a comprehensive system, which skews the distribution positively.

Finally we derive the range of winners and loser from a tracking system relative to a comprehensive system. Given that θ_2 are monotonically increasing in θ_1 in every track, the range is characterized by the intersection of the linear function $\theta_{2,C}(\theta_1, \bar{\theta}_{1,C})$ with $\theta_{2,V}(\theta_1, \bar{\theta}_{1,V})$ and $\theta_{2,A}(\theta_1, \bar{\theta}_{1,A})$. For any skill threshold, the lower intersection $\theta_{1,L}$ hence solves

$$\begin{aligned}
& & \theta_{1,L} + \alpha\bar{\theta}_{1,C} + \frac{\beta^2}{2\gamma} + \beta\theta_{1,L} + \gamma\bar{\theta}_{1,C}\theta_{1,L} - \frac{\gamma}{2}\bar{\theta}^2_{1,C} + \eta_2 \\
& = & \theta_{1,L} + \alpha\bar{\theta}_{1,V} + \frac{\beta^2}{2\gamma} + \beta\theta_{1,V} + \gamma\bar{\theta}_{1,V}\theta_{1,L} - \gamma\bar{\theta}^2_{1,V} + \eta_2 \qquad (2.34) \\
\iff \quad \theta_{1,L} & = & \frac{1}{2}\bar{\theta}_{1,V} - \frac{\alpha}{\gamma}.
\end{aligned}$$

Similarly, the upper intersection is given at

$$\theta_{1,U} = \frac{1}{2}\bar{\theta}_{1,A} - \frac{\alpha}{\gamma}. \qquad (2.35)$$

For any skill threshold $\tilde{\theta}_1$, the interval $[\theta_{1,L}, \bar{\theta}_{1,U}]$ is non-empty. Hence, there are always children with initial skill levels inside this interval who lose in terms of end-of-school skills in a full tracking system relative to a comprehensive system. Every child outside of this interval gains relative to the comprehensive system.

With $\alpha = 0$, the tracking skill threshold is at $\tilde{\theta}_1 = 0$ even if parents endogenously sort their children. Hence, children with initial skills inside a symmetric interval around 0, $[\frac{1}{2}\bar{\theta}_{1,V}, \frac{1}{2}\bar{\theta}_{1,A}]$, lose relative to a comprehensive track since $\bar{\theta}_{1,V} = -\bar{\theta}_{1,A}$ if $\tilde{\theta}_1 = 0$. The average loss of a child in this interval is equal to $\frac{\gamma}{2}\bar{\theta}^2_{1,V} = \frac{\gamma}{2}\bar{\theta}^2_{1,A}$.

If $\alpha > 0$, and the policymaker enforces the tracking skill threshold $\tilde{\theta}_1 = 0$, the losses from tracking are concentrated among children in the V track. To see this, note that every child with initial skill in the interval $[\theta_{1,L}, 0]$ is allocated into the V track but loses relative to a comprehensive system. Similarly, every child with an initial skill inside $[0, \theta_{1,U}]$ is allocated to track A but loses relative to a comprehensive system. With $\alpha > 0$, $|\theta_{1,U}| < |\theta_{1,L}|$ and hence, the range of children in the A track that lose is smaller. The interval $[0, \theta_{1,U}]$ may even be empty in which case only children in the V track lose from tracking.

Proposition 3

First, we characterize the variance of θ_2. We start by collecting expressions for conditional and unconditional first and second moments.

The unconditional expected value of θ_2 in track V, if everyone went to V is

$$\begin{aligned} \mathbb{E}(\theta_{2,V}) &= \frac{\beta^2}{2\gamma} + \alpha\bar{\theta}_{1,V} - \frac{\gamma}{2}\bar{\theta}^2_{1,V} \\ &= \frac{\beta^2}{2\gamma} - \alpha\sigma_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})}{F(\tilde{\theta}_1/\sigma_{\theta_1})} - \frac{\gamma}{2}\sigma^2_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})^2}{F(\tilde{\theta}_1/\sigma_{\theta_1})^2}. \end{aligned} \tag{2.36}$$

The unconditional expected value of θ_2 in track A, if everyone went to A is

$$\begin{aligned} \mathbb{E}(\theta_{2,A}) &= \frac{\beta^2}{2\gamma} + \alpha\bar{\theta}_{1,A} - \frac{\gamma}{2}\bar{\theta}^2_{1,A} \\ &= \frac{\beta^2}{2\gamma} + \alpha\sigma_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})}{1 - F(\tilde{\theta}_1/\sigma_{\theta_1})} - \frac{\gamma}{2}\sigma^2_{\theta_1}\frac{f(\tilde{\theta}_1/\sigma_{\theta_1})^2}{(1 - F(\tilde{\theta}_1/\sigma_{\theta_1}))^2}. \end{aligned} \tag{2.37}$$

The variance of θ_2 in a comprehensive system is

$$\begin{aligned} Var(\theta_{2,C}) &= \mathbb{E}\Big(\big(\theta_2 - \mathbb{E}(\theta_2)\big)^2\Big) \\ &= (1+\beta)^2\sigma^2_{\theta_1} + \sigma^2_{\eta_2} \\ &\sigma^2_{\theta_2,C} + \sigma^2_{\eta_2}, \end{aligned} \tag{2.38}$$

where we define $\sigma^2_{\theta_2,C}$ to be the variance of θ_2 net of the additive skill shock variance.

Second, we can derive the expected value of end-of-school skills in the 2-period model in a late tracking system as

$$
\begin{aligned}
\mathbb{E}(\theta_{3,LT}) &= \mathbb{E}(\mathbb{E}(\theta_{3,LT}|S^2_{LT})) \\
&= \mathbb{E}(\theta_{2,LT}) + \frac{\beta^2}{2\gamma} + (\alpha+\beta)\,\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S_{LT})) + \frac{\gamma}{2}\,\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S_{LT})^2) \\
&= (2+\alpha+\beta)\frac{\beta^2}{2\gamma} + \frac{\gamma}{2}[\sigma^{\theta_2,LT} - \mathbb{E}(Var(\theta_{2,LT}|S_{LT}))],
\end{aligned}
\tag{2.39}
$$

where $\mathbb{E}(\theta_{2,LT})$ and $\sigma^2_{\theta_2,LT}$ are just equal to the mean and variance of the comprehensive system in the one-period model (see equation (2.38)). The variable S_{LT} indicates the track selection in period 2, which follows the cut-off rule $S_{LT} = V$ if $\theta_{2,LT} \leq \tilde{\theta}_{2,LT}$ and $S_{LT} = A$ otherwise. The cut-off that maximizes (2.39) is $\tilde{\theta}^*_{2,LT} = \mathbb{E}(\theta_{2,LT}) = \frac{\beta^2}{2\gamma}$. This follows as (2.39) mirrors that of average end-of-school skills in the full tracking system of the one-period model in that average and aggregate $\theta_{3,LT}$ decrease in the expected variance of skills in period 2 across tracks.

Similarly, we find the expected value of end-of-school skills in the 2-period model in an early tracking system as

$$
\begin{aligned}
\mathbb{E}(\theta_{3,ET}) &= \mathbb{E}(\mathbb{E}(\theta_{3,ET}|S^2_{ET})) \\
&= \frac{\beta^2}{2\gamma} + (1+\alpha+\beta)\,\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})) + \beta\frac{\gamma}{2}\,\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})^2) \\
&= \frac{\beta^2}{2\gamma} + (1+\alpha+\beta)\left(\frac{\beta^2}{2\gamma} + \beta\frac{\gamma}{2}[\sigma^2_{\theta_1} - \mathbb{E}(Var(\theta_{1,ET}|S_{ET}))]\right) + \beta\frac{\gamma}{2}\,\mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})^2 \\
&= \frac{\beta^2}{2\gamma} + (1+\alpha+\beta)\left(\frac{\beta^2}{2\gamma} + \beta\frac{\gamma}{2}[\sigma^2_{\theta_1} - \mathbb{E}(Var(\theta_{1,ET}|S_{ET}))]\right) \\
&+ \beta\frac{\gamma}{2}[\sigma^2_{\theta_2,ET} - \mathbb{E}(Var(\theta_{2,ET}|S_{ET}))].
\end{aligned}
\tag{2.40}
$$

Comparing (2.39) and (2.40), the condition that governs if average end-of-school skills in a late tracking system are larger than in an early tracking system reads

$$
\begin{aligned}
&\mathbb{E}(\theta_{3,LT}) - \mathbb{E}(\theta_{3,ET}) \\
&= \beta\frac{\gamma}{2}\left(\mathbb{E}(\mathbb{E}(\theta_{2,LT}|S_{LT})^2) - \mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})^2)\right) \\
&- (1+\alpha+\beta)\beta\frac{\gamma}{2}\,\mathbb{E}(\mathbb{E}(\theta_1|S_{ET})^2) > 0.
\end{aligned}
\tag{2.41}
$$

The last term of (2.41) represents the advantage of early tracking in the first stage of the schooling years. It stems from the smaller expected conditional variances of initial skills among children that are tracked relative to the overall variance. The conditional expected value of θ_2 in a late tracking system is given by

$$\mathbb{E}(\theta_{2,LT}|S_{LT}=V)=\frac{\beta^2}{2\gamma}-\sigma_{\theta_2,LT}\frac{f(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})}{F(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})} \tag{2.42}$$

and

$$\mathbb{E}(\theta_{2,LT}|S_{LT}=A)=\frac{\beta^2}{2\gamma}+\sigma_{\theta_2,LT}\frac{f(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})}{1-F(\tilde{\theta}_{2,LT}/\sigma_{\theta_2,LT})}, \tag{2.43}$$

where the unconditional variance of θ_2 in a late tracking system is given by $\sigma^2_{\theta_2,LT}=\sigma^2_{\theta_2,C}+\sigma^2_{\eta 2}$, i.e. by the one computed in equation (2.38). Since late tracking occurs *after* the realization of skill shocks in period 2, this variance additively *includes* the variance of these shocks.

Condition (2.41) is generally ambiguous and hard to interpret for arbitrary skill thresholds. We focus again on the optimal tracking case, that is, the case with skill threshold $\tilde{\theta}_1=\mathbb{E}(\theta_1)=0$ and $\tilde{\theta}_2=\mathbb{E}(\theta_{2,LT})=\frac{\beta^2}{2\gamma}$. In that case, we can write the expressions for the various expected square conditional expected values as follows:

$$\begin{aligned}\mathbb{E}(\mathbb{E}(\theta_1|S_{ET})^2)&=2\chi\sigma^2_{\theta_1}\\ \mathbb{E}(\mathbb{E}(\theta_{2,LT}|S_{LT})^2)&=\frac{\beta^4}{4\gamma^2}+2\chi(\sigma^2_{\theta_{2,LT}}+\sigma^2_{\eta 2})\\ \mathbb{E}(\mathbb{E}(\theta_{2,ET}|S_{ET})^2)&=\frac{\beta^4}{4\gamma^2}+2\chi\sigma^2_{\theta_1}\left(\alpha^2+\gamma^2 f(0)^2\sigma^2_{\theta_1}-\frac{\beta^2}{2}\right)\\ +2f(0)\sigma^2_{\theta_1}\left(\beta^2+2\alpha(1+\beta)-(2\gamma f(0)\sigma_{\theta_1})^2\right)&+2\chi(\sigma^2_{\theta_{2,LT}}+2\chi\gamma^2\sigma^2_{\theta_1}).\end{aligned}$$

Condition (2.41) then becomes

$$\begin{aligned}&\mathbb{E}(\theta_{3,LT})-\mathbb{E}(\theta_{3,ET})\\ &=\beta\frac{\gamma}{2}\Bigg(2\chi\sigma^2_{\eta 2}-2\chi\sigma^2_{\theta_1}\Bigg(\alpha^2+\gamma^2 f(0)^2\sigma^2_{\theta_1}-\frac{\beta^2}{2}\\ &+\beta^2+2\alpha(1+\beta)-4\gamma^2 f(0)^2\sigma^2_{\theta_1}+2\chi\gamma^2\sigma^2_{\theta_1}+1+\alpha+\beta\Bigg)\Bigg)\\ &=\frac{\gamma}{\pi}\left(\sigma^2_{\eta 2}-\sigma^2_{\theta_1}\left(1+\alpha+\alpha^2+\beta+\frac{\beta^2}{2}+2\alpha(1+\beta)+\frac{\gamma^2}{2\pi}\sigma^2_{\theta_1}\right)\right)>0.\end{aligned} \tag{2.44}$$

From this, Proposition 3 follows.

2.B Equilibrium Definition

We introduce some notation to define the equilibrium more easily. Let $x_j \in X_j$ be the age-specific state vector of an individual of age j, as defined by the recursive representation of the individual's problems in Section 2.3. Let its stationary distribution be $\Theta(X)$. Then, a stationary recursive competitive equilibrium for this economy is a collection of: (i) decision rules for college graduation $\{d^E(x_5)\}$, for school track $\{d^{S'}(x_{11})\}$, consumption, labor supply, and assets holdings $\{c_j(x_j), n_j(x_j), a_j(x_j)\}$, and parental transfers $\{a'_5\,(x_j)\}$; value functions $\{V_j\,(x_j)\}$; (iii) aggregate capital and labor inputs $\{K, H_0, H_1\}$; (iv) prices $\{r, w^0, w^1\}$; and (v) average skill levels among children in school track S' $\{\bar{\theta'}_{j',S'}\}$ such that:

1. Given prices and average skill levels among children in each school track, decision rules solve the respective household problems and $\{V_j\,(x_j)\}$ are the associated value functions.

2. Given prices, aggregate capital and labor inputs solve the representative firm's problem, i.e. it equates marginal products to prices.

3. Given average skill levels among children in each school track, allocation of children in school track solves the parent's problem, i.e. actual average skill levels are consistent with parents' prior.

4. Labor market for each education level clears.
   For high-school level:
$$H_0 = \sum_{j=5}^{16} \int_{X_j} n_j(x_j)\; h_j\,(x_j)\, d\Theta(X \mid E=0) + \sum_{j=5}^{5} \int_{X_j} n_j(x_j)\; h_j\,(x_j)\, d\Theta(X \mid E=1)$$
   where the first summation is the supply of high-school graduates while the second is the labor supply of college students.
   For college level:
$$H_1 = \sum_{j=6}^{16} \int_{X_j} n_j(x_j)\; h_j\,(x_j)\, d\Theta(X \mid E=1).$$

5. Asset market clears
$$K = \sum_{j=5}^{20} \int_{X_j} a_j(x_j) d\Theta(X),$$
   which implies that the goods market clears;

6. The distribution of X is stationary: $\Theta(X) = \int \Gamma(X) d\Theta(X)$.

2.C German Education System

In this section, we provide an overview of the most important features of the German Education and School System. A more extensive description can be found, for example, in Henninges et al. (2019). Figure 2.C.1 illustrates a simplified structure of the system, starting in Grade 4 and ending with tertiary education.

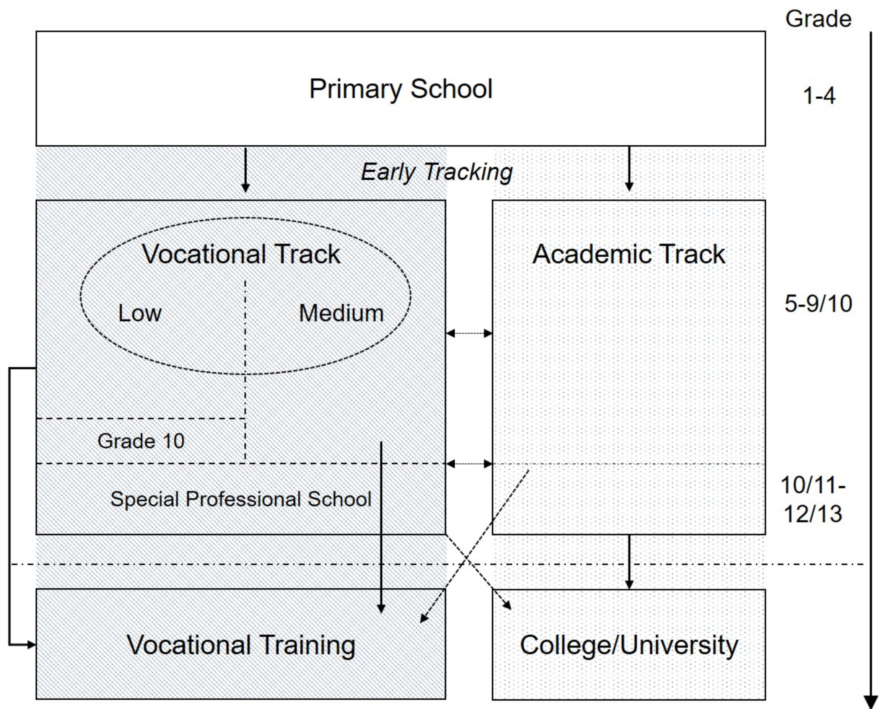
Generally, schooling is mandatory in Germany for every child starting at age six and lasting for nine or ten years. At age six, all children visit a comprehensive primary school that lasts the first four grades.[70] After that, children are allocated into traditionally three different secondary school tracks: A lower vocational track, a medium vocational track, and an academic track. However, triggered by the so-called PISA shock in the early 2000s, federal states in Germany have started reforming their secondary school system. In particular, the two vocational tracks have often been combined into one, resulting in a two-track system in the majority of federal states (Bellenberg and Forell, 2012). For that reason, and because even if still two vocational tracks exist, they are much more similar in comparison to the academic track schools, we opt to restrict our analysis in this paper to two school tracks.

Generally, the school tracks differ in the curricula taught, the length of study, and the end-of-school qualifications that come with graduation. In particular, only the academic track schools end with a university entrance qualification that directly allows children to go to college. This requires the completion of the second stage of secondary school, typically grades 10/11 to 12/13. Graduating from a vocational track occurs after Grades 9 and 10 and allows children to take up vocational training in blue-collar jobs or proceed to a professional school that prepares for entry into white-collar, business, or skilled trade occupations. At this stage, there is considerable scope for mobility between tracks. Firstly, professional degrees often allow access to university studies in selected fields. Secondly, children can directly switch to an academic track school if their school marks and achievements admit that. Finally, after having worked for a number of years in vocational jobs, access to some college degrees can be possible. At the same time, it is of course possible to switch from an academic track school to a vocational training or job after the mandatory education has been completed.

The public expenditure per student does not differ significantly across school tracks. Table 2.C.1 lists average per-student expenditures across the various school types in the years 2010 to 2020. Across these years, public expenditures by student were highest in pure

[70] In two federal states, Berlin and Brandenburg, comprehensive primary school lasts the first 6 grades.

Figure 2.C.1: Simplified Structure of the German Education System



lower vocational track schools. Expenditures in academic track schools were roughly equal compared to expenditures in joint vocational track schools. The bulk of these expenditures is attributable to teacher pay (around 80%) and the rest for investments into buildings, equipment etc. This suggests that resource differences across school tracks should not be a main driver behind achievement differences, on average.

A remaining driver behind achievement differences across school tracks could be the teaching quality. In particular, higher-quality teachers could select for academic track schools. However, regardless of the secondary school track, becoming a teacher requires university studies in the range of 7 to 10 semesters and a similar university degree. On top of that, the differences in wages across school tracks are no longer significant in many federal states. For example, both tenured teachers at vocational track schools and teachers at academic track schools are eligible for the same public pay grade in most northern and eastern federal states already.

2.D Measuring Child Skills in the NEPS

In this section, we provide an overview of our measures of child skills. One of the main goals of the NEPS project is to document the development of competencies of individuals over their

Table 2.C.1: Per-Student Public Expenditures across School Types and Years

Year	Primary	Lower Voc.	Upper Voc.	Joint Voc.	Acad.	Compr.
2010	5,200 €	7,100 €	5,300 €	8,000 €	6,600 €	6,600 €
2011	5,500 €	7,300 €	5,600 €	8,000 €	7,100 €	7,100 €
2012	5,400 €	7,900 €	5,700 €	7,700 €	7,200 €	7,200 €
2013	5,600 €	8,200 €	5,900 €	7,700 €	7,500 €	7,500 €
2014	5,900 €	8,700 €	6,200 €	8,000 €	7,800 €	7,800 €
2015	6,000 €	8,900 €	6,400 €	8,000 €	7,900 €	8,000 €
2016	6,200 €	9,300 €	6,700 €	8,100 €	8,100 €	8,200 €
2017	6,400 €	9,800 €	7,000 €	8,300 €	8,500 €	8,600 €
2018	6,700 €	10,400 €	7,400 €	8,700 €	8,800 €	9,100 €
2019	7,100 €	11,200 €	7,900 €	9,200 €	9,300 €	9,500 €
2020	7,400 €	12,200 €	8,200 €	9,500 €	9,600 €	10,000 €

Source: Statistisches Bundesamt (Bildungsfinanzbericht, Bildungsausgaben - Ausgaben je Schüler, Sonderauswertung)

lifespan (Neumann et al., 2013). To that end, the NEPS carefully designs and implements regular tests of the respondents' competencies along several domains. Given its central role not only in educational contexts but also as a predictor for later labor market success, we focus on mathematical competencies. Following the guidelines set by the Program for International Student Assessment (PISA), the mathematical competence domain is not just designed to assess the extent to which children have learned the content of school curricula but also to judge a child's ability to use mathematics to constructively engage with real-life problems (Neumann et al., 2013). The test, therefore, includes items related to "overarching" mathematical content areas that are consistent across all ages, such as quantity, change & relationships, space & shape, as well as several cognitive components, such as mathematical communication, argumentation, or modeling. The age-specific test items include primarily simple and complex multiple-choice questions, as well as short-constructed responses.[71]

In order to use these questions for the analysis of latent competencies, they need to be scaled. The NEPS (similar to the PISA) uses a scaling procedure that follows item response theory (IRT). IRT is a popular instrument in psychometrics to extract latent ability or other factors from test data. To quote the NEPS: "IRT was chosen as scaling framework for the newly developed tests because it allows for an estimation of item parameters independent of the sample of persons and for an estimation of ability independent of the sample of items. With IRT it is possible to scale the ability of persons in different waves on the same scale, even when different tests were used at each measurement occasion" (Pohl and Carstensen, 2013).

[71] A simple multiple choice question consists of one correct out of four answer categories, and complex multiple choice questions consist of a number of subtasks with one correct answer out of two options. Short-constructed responses typically ask for a number (Pohl and Carstensen, 2012). The mathematical competence test primarily consists of simple multiple-choice questions.

The most important scaling model used by the NEPS is the Rasch model. This model assumes that the right answers given to a set of questions by a number of respondents contain all information needed to measure a person's latent ability as well as the question's difficulty. It does so by positing that the probability that person v gives the right answer to question i is given by:

$$p(X_{vi} = 1) = 1 - p(X_{vi} = 0) = \frac{\exp(\theta_v - \sigma_i)}{1 + \exp(\theta_v - \sigma_i)}, \tag{2.45}$$

where θ_v denotes the latent ability of person v and σ_i is a measure of the question's difficulty. Thus, this model maps the total sum score of an individual into an ability parameter estimate. The scale is arbitrary. However, the ability estimate is cardinal.[72] This model is estimated via (weighted) conditional maximum likelihood under a normality assumption on latent ability.

Table 2.D.1 describes NEPS samples of mathematics assessments by cohort and Grade level.

Table 2.D.1: NEPS Mathematic Assessment Samples

			Information on Parents' Education		Information on School Track	
		Obs.	Obs.	% College Parents	Obs.	% Ac. Track
Cohort 1	K1	2,014	1,709	51%		
Cohort 2	G1	6,352	5,784	46%	2,731	63%
	G2	5,888	5,425	47%	2,651	62%
	G4	6,610	6,068	46%	3,229	63%
	G7	2,479	2,410	51%	2,208	58%
Cohort 3	G5	5,193	3,856	38%	4,369	52%
	G7	6,191	4,214	38%	5,525	49%
	G9	4,888	3,387	38%	4,356	47%
	G12*	3,785	2,830	41%	3,331	58%
Cohort 4	G9	14,523	8,474	35%	14,215	40%
	G12*	5,733	3,767	24%	5,530	23%

Notes: This table describes NEPS mathematics assessments by cohort. Note that in Grade 12, the assessments are different by school track, which makes the comparison of test scores by parental education or school track impossible. Source: NEPS

[72] It is interval-scaled asBallou (2009) puts it. That means an increase of 5 points from 15 to 20 represents the same gain in achievement as from 25 to 30.

2.E Details on Child Skill Technology Estimation

2.E.1 Skills Measurement

We employ a linear measurement system for the logarithm of latent skills in each period that is given by

$$M_{i,k,j} = \mu_{k,j} + \lambda_{k,j}\theta_{i,j} + \epsilon_{i,k,j}, \tag{2.46}$$

where $M_{i,k,j}$ denotes the kth measure for latent log skills of child i in period j. In each period, we have at least 3 different measures in our data, which typically constitute the achievement (item response theory) test scores of each child and are discussed in detail below. The parameters $\mu_{k,j}$, and $\lambda_{k,j}$ denote the location and factor loading of latent log skills, respectively. By $\epsilon_{i,k,j}$, we denote the measurement error. The parameters and measures are defined conditional on child's age and gender, which we keep implicit.

Following Cunha et al. (2010), we normalize $\mathbb{E}(\theta_j) = 0$ and $\lambda_{1,j} = 1$ for all j. That is, the first-factor loading is normalized to 1 in all periods.[73] We further normalize the measurement errors, such that $E(\epsilon_{k,j}) = 0$ for all j. Given that, the location parameters $\mu_{k,j}$ are identified from the means of the measures. In order to identify the factor loadings, we further assume that the measurement errors are independent of each other across measures and independent from latent skills. Under these assumptions and given that we have at least three measures of latent skills available in each period, we can identify the loadings on each measure in each period by ratios of covariances of the measures (as in Agostinelli et al. (2019)):

$$\lambda_{k,j} = \frac{Cov(M_{k,j}, M_{k',j})}{Cov(M_{1,j}, M_{k',j})} \tag{2.47}$$

for all $k, k' > 1$ and $k \neq k'$. Rescaling the measures by their identified location and scale parameters then gives us error-contaminated measures of latent skills for each period as

$$\theta_{i,j} = \frac{M_{i,k,j} - \mu_{k,j}}{\lambda_{k,j}} - \frac{\epsilon_{i,k,j}}{\lambda_{k,j}} = \widetilde{M}_{i,k,j} - \frac{\epsilon_{i,k,j}}{\lambda_{k,j}}. \tag{2.48}$$

Equipped with identified latent variables up to measurement error for all periods, we can plug these into the child skill technology (2.27), which yields

[73] We are aware of the potential bias that can arise from this assumption (see Agostinelli and Wiswall (2016)). However, contrary to their case, we measure three different stages of child development, where each stage comes with a new cohort of children (see below). Thus we cannot follow children over multiple periods. Moreover, even if we could, the data we use does not contain age-invariant measures according to their definition.

$$\begin{aligned}\widetilde{M}_{i,k,j+1} = \kappa_{0,j} + \kappa_{1,j}\widetilde{M}_{i,k,j} + \kappa_{2,j}\widetilde{M}^2_{i,k,j} + \kappa_{3,j}\overline{\widetilde{M}}_{-i,j,S} \\ + \kappa_{4,j}(\widetilde{M}_{i,k,j} - \overline{\widetilde{M}}_{j,S})^2 + \kappa_{5,j}E_i + \zeta_{i,k,j+1},\end{aligned} \tag{2.49}$$

where $\overline{\widetilde{M}}_{-i,j,S}$ refers to the expected value of the kth transformed measure across all children other than i in a classroom in track S and $\overline{\widetilde{M}}_{j,S}$ to that of the expected value of the measures across all children in a school that belongs to track S.

Importantly, the residual $\zeta_{i,k,j+1}$ now contains not only structural skill shocks, $\eta_{i,j+1}$, but also the measurement errors, $\epsilon_{i,k,j}$ as well as interactions of the measurement error with the rescaled measures and even the variance of the measurement errors. For that reason, even if a standard assumption of mean independence of the structural shocks η conditional on all independent variables holds, an OLS estimator of (2.49) will be biased. To account for that, we follow the literature and use excluded measures as instrumental variables, which we describe in Appendix 2.E.[74]

Tables 2.E.1 and 2.E.2 describe the evolution of child skills over time using the identified latent variables.

2.E.2 Instrumental Variables and Data

We estimate (2.49) using the (IRT) measure of mathematics tests. This is because we have this measure available at every stage j. We consider three stages of the schooling career, corresponding to the timing of our model. The first stage is the second period in a child's life and therefore indexed by $j = 2$ and corresponds to 4 years of primary school in real life, where children are typically aged 6 to 10. To estimate the parameters of this stage, we use the NEPS Starting Cohort 2. To account for measurement error, we instrument the math test scores in the first grade of primary school using test scores on science and vocabulary in the first grade of primary school as well as a math test in the second grade of primary school.

The second stage corresponds to the third period in a child's life, $j = 3$, when they are between 10 and 14 years old and typically go to secondary school. To estimate the parameters, we use data from the NEPS Starting Cohort 3. This data set, unfortunately, contains a relatively low number of observations for the first two years of secondary school.

[74]Under the assumption that measurement error is uncorrelated across measures, this strategy will take of measurement error and the interaction terms included in (2.49) but not of the variances of the measurement error. These will show up in the estimated intercept, thus biasing the constant. Since this constant does not have an economic meaning in our model, we disregard this bias for now. In the future, we can recover the variance of the measurement errors using ratios of covariances of the measures again, as in Cunha et al. (2010).

Table 2.E.1: Differences in Average Skills in Standard Deviation

	Grade	Difference by Parent's Education	Difference by School Track
Cohort 2	G1	0.46	0.77
	G4	0.55	0.83
	G7	0.45	0.93
Cohort 3	G5	0.48	0.87
	G7	0.55	0.92
	G9	0.57	1.01
Cohort 4	G9	0.54	0.97

Notes: This table provides information on average differences identified latent math grades up to measurement error in one standard deviation unit by parental background and school track over time. All observations are weighted. Source: NEPS.

Table 2.E.2: Rank-Rank Correlations

		Rank-Rank Correlation	Obs.
Cohort 2	G1-G4	0.57	3,116
	G4-G7	0.61	2,267
Cohort 3	G5-G9	0.67	4,927

Notes: This table provides the rank-rank correlations in identified latent math grades up to measurement error. Source: NEPS

For that reason, we estimate the child technology parameters in that stage once on math test scores between grades 5 and 9 and once on math test scores between grades 7 and 9, after an increase in the sample size. The instruments are reading test scores in grade 5 and science test scores in grade 6 in the former case, and reading and orthography test scores (both in grade 7) in the latter case.

The third stage corresponds to the fourth period in a child's life, $j = 4$ when they are between 14/15 and 18 years old and typically finish secondary school. We use the NEPS Starting Cohort 4 to estimate the technology parameters in this case, again relying on transformed math test scores in grade 9 and grade 12.[75] The instruments we employ are vocabulary, science, and reading test scores in grade 9.

For the primary school stage, we restrict the sample to children who are in classes with a size of at least 5 children such that we can compute a meaningful class average. In both secondary school stages, we restrict the class sizes to be at a minimum of 8 children. This is because it is not uncommon that some primary schools, especially in rural areas, feature quite small class sizes in Germany. In contrast, class sizes are typically in the range of 20-30 in secondary school.

2.F Discussion on Child Skill Shocks

As for the adult human capital, we assume child skills are subject to idiosyncratic shocks. These shocks represent unexpected heterogeneity in child development speeds (such as late-bloomers) and any shock that can arise during childhood and affect the child's learning, such as health issues, a move, parents' divorce, meeting an influential mentor, etc.

An alternative model would assume child skills are not subject to shocks but imperfectly observed by parents. In this section, we elaborate on an alternative model based on our baseline model that introduces this feature and compare it to our baseline model.

Specifically, in this alternative modeling, θ would be the true (log) skills that matter for the child skill evolution and future earnings and evolve according to the stage-specific function f, defined by:

[75] In Germany, the vocational track schools typically end after grade 9 or grade 10 and so-called upper secondary schooling only happens in academic track schools. However, the NEPS data keeps track of the students even if they are no longer enrolled in a school and tests them at the same age. A remaining issue is, of course, that even though we know the classroom compositions in grade 9, we do not know how long learning in that classroom continues in a vocational track school. For that reason, we make the assumption that children who went to a vocational track school that finished before they are 18 years old continue to learn in an environment that is the same as if the vocational school had continued. In reality, students who graduate from vocational schools often continue with an apprenticeship, where we think it reasonable to assume that the peer composition is similar to the one in school.

$$\theta_{j+1} = f(\theta_j, P_j^S, \bar{\theta}_j^S, E) \tag{2.50}$$
$$= \kappa\theta_j + \alpha\bar{\theta}_j^S + + \beta\, P_j^S + \gamma\,\theta_j\, P_j^S - \frac{\delta}{2}\, {P_j^S}^2 + \zeta E, \tag{2.51}$$

where, similarly to the baseline model, P^S is the instruction pace in track S, the average peer skills is denoted by $\bar{\theta}^S$ and E stands for parental background. However, in this alternative version, parents would not directly observe their child's skills θ_j. Instead, in every period, they would receive an unbiased signal $\hat{\theta}_j$ about their child skills, with:

$$\begin{aligned} \hat{\theta}_j &= \theta_j + \epsilon_{\theta,j} \\ \epsilon_{\theta,j} &\sim \mathcal{N}(0, \sigma^2_{\epsilon_\theta}). \end{aligned} \tag{2.52}$$

Given the parents' initial prior $\tilde{\theta}_{j-1}$, that is unbiased and follows a normal distribution $\mathcal{N}(\theta_{j-1}, \sigma_{j-1})$, parents update their perception of their current child's skills $\tilde{\theta}_j^P = f(\tilde{\theta}_{j-1}, P_j^S, \bar{\theta}_j^S, E)$ using Bayesian updating:[76]

$$\begin{aligned} \tilde{\theta}_j &= k\,\hat{\theta}_j + (1-k)\tilde{\theta}_j^P \\ \sigma_j^2 &= \sigma_{j-1}^2 - k\sigma_{j-1}^2 \\ k &= \frac{\sigma_{j-1}^2}{\sigma_{j-1}^2 + \sigma^2_{\epsilon_\theta}}, \end{aligned} \tag{2.53}$$

where k is the Kalman gain and is increasing in the precision of the signal $(\frac{1}{\sigma^2_{\epsilon_\theta}})$.

Since the perception of child skills is unbiased, the perception of the peer skills is equal to the truth in the limit. Consequently, $\bar{\theta}_j^S$ is assumed to be perfectly observed by the parents and stable in equilibrium. Similarly, in the limit, the policymaker perfectly observed the average child skills in every school track and set the pace of instruction P_j^S according to Lemma 1. Then, we can define the child skill production function as

$$\begin{aligned} \theta_{j+1} &= f(\theta_j, \bar{\theta}_j^S, E) \\ &= \frac{\beta^2}{2\delta} + \ (\kappa + \frac{\beta\gamma}{\delta})\theta_j + \ (\alpha)\bar{\theta}^S - \ \frac{\gamma^2}{2\delta}{\bar{\theta}^S}^2 + \ \frac{\gamma^2}{\delta}\theta_j\bar{\theta}^S + \zeta\, E \\ &= \omega_0 + \ \omega_1\,\theta_j + \ \omega_2\,\bar{\theta}^S + \ \omega_4\,{\bar{\theta}^S}^2 - \ 2\omega_4\,\theta_j\bar{\theta}^S + \ \omega_5\,E. \end{aligned}$$

[76]We could assume the first initial prior to be equal to the signal they receive in $j = 1$.

Notice that the child skill evolution is identical to one in the baseline model but for the idiosyncratic shock η that are here absent. As a result, the average skill threshold that determines the school track allocation would be identically determined in both model versions. Indeed, in the baseline model, the expected future child skills are independent of the shocks η that are assumed to be normally distributed and centered to zero. To see this, notice that in both models, the average skill threshold θ^* for a given parental background E and current (perceived) skills θ_3, is determined by the following equation:

$$
\begin{aligned}
E(\theta_5, E'|S = A, E) &= E(\theta_5, E'|S = V, E) \\
E(f(\theta_4, \bar{\theta}_4^A, E), E'|E) &= E(f(\theta_4, \bar{\theta}_4^V, E), E'|E) \\
E(\omega_1\, \theta_4 +\, \omega_2\, \bar{\theta}_4^A +\, \omega_4\, \bar{\theta}_4^{A^2} -\, 2\omega_4\, \theta_4\bar{\theta}_4^A, E'|E) &= E(\omega_1\, \theta_4 +\, \omega_2\, \bar{\theta}_4^V +\, \omega_4\, \bar{\theta}_4^{V^2} -\, 2\omega_4\, \theta_4\bar{\theta}_4^V, E'|E).
\end{aligned}
$$

Assuming $\bar{\theta}_j^A$ and $\bar{\theta}_j^V$ for $j = 3, 4$ are known and fixed, by the linearity of the function, we can replace θ_4 in the expectation by its expected value $E(\theta_4|S, E) = f(\theta_3, \bar{\theta}_3^S, E)$. So θ^* is independent of η in the baseline model and identically determined as in this alternative model.

Conceptually, misallocation sources, however, differ between the two models. In the alternative model, at the time of the school track choice $j = 3$, parents make their decision based on their perception of their child's skills $\tilde{\theta}_3 \sim \mathcal{N}(\theta_3, \sigma_3^2)$. Part of the misallocation will be driven by σ_3, which governs how imprecise the parental perception of the skills is. In the baseline model, parents perfectly observed their current child's skills, but skills are subject to shocks. Part of the misallocation is then governed by $\eta_4 \sim \mathcal{N}(0, \sigma_{\eta_4}^2)$, and more precisely by its variance σ_{η_4}. While allowing for re-tracking would solve the issue of misallocation driven by skill uncertainty in the baseline model, it would not completely solve the issue driven by imperfectly observed skills in the alternative model. Indeed, skills are still imprecisely observed in period 4—even though the precision is greater than in period three due to the learning process.

Finally and crucially, earnings variance would be entirely determined at the earliest age without child skills shocks. As a result, comparing early and late tracking in the two model versions leads to different results. While in the baseline model, late tracking versus early tracking makes the school track choice less dependent on early skill conditions, it is the reverse in the alternative model. Postponing tracking allows parents to make a more informed decision about the school track choice, strengthening the relationship between early (true) skill conditions and the school track. Still, the effect on mobility is ambiguous as late tracking shrinks the difference in skills across socioeconomic groups.

In reality, it is probably a mix of both modeling versions. However, the data does not allow us to differentiate between the two mechanisms. We use latent skills for calibration purposes and don't have information on parents' perceptions of their child's skills. Since we think skills are likely subject to shocks during childhood, as human capital is likely subject to shocks during adulthood, we favor the modelization with child skill shocks. A noise in the preference shifter could be regarded as a reduced form of capturing the imprecision in the parents' perception of their child's skills.

2.G Welfare Measure

Our analysis centers on evaluating aggregate welfare under scenarios that feature different policies. Welfare is defined by the consumption equivalence under the veil of ignorance in the baseline economy relative to the economy with the counterfactual policy in place. Formally, let $C \in \{0, 1, 2, ...\}$ denote the set of counterfactuals, with C = 0 being the baseline economy (early tracking) in a steady state. We refer to the consumption equivalence as the percentage change in consumption Δ in the baseline economy that makes individuals indifferent between being born in the baseline economy (C = 0) and the one in which the counterfactual policy $C \neq 0$ is in place. Denote $\mathcal{V}_5^C(\theta_5, a_5, \phi, S, E^p, \Delta)$ be the welfare of agents with the initial state of the economy j=5 if their consumption (and that of their descendants) were multiplied by $(1 + \Delta)$:

$$\mathcal{V}^C(\theta_5, a_5, \phi, S, E^p, \Delta) = \mathbb{E}^C \sum_{j=5}^{j=20} \beta^{j-5} v_j \left(c_j^{*C}(1+\Delta), n_j^{*C}, E^{*C}, \theta_5, S, E^p \right) + \beta^{13-5} \delta \mathcal{V}_{j^5}^C \left(\theta_5', a_5', \phi', S', E^{*C}, \Delta \right),$$

where E^p is the education of the parent, and for j = 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20

$$v_j(c_j, n_j, E, \theta_5, S, E^p) = \frac{(c_j/q)^{1-\sigma}}{1-\sigma} - b \frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}, \tag{2.54}$$

for j = 5

$$v_j(c_j, n_j, E, \theta_5, S, E^p) = \frac{(c_j/q)^{1-\sigma}}{1-\sigma} - b \frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} - 1\{E = 1\} \, \psi(S, \theta_5, \nu(E^p)), \tag{2.55}$$

for j = 11

$$v_j(c_j, n_j, E, \theta_5, S, E^p) = \frac{(c_j/q)^{1-\sigma}}{1-\sigma} - b\frac{n_j^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}} - 1\{S = A\}\ \chi(E). \tag{2.56}$$

Note that the policy functions are assumed to be unchanged when Δ is introduced. The average welfare is:

$$\bar{\mathcal{V}}^C(\Delta) = \sum_{S,E^p} \int_{\theta_5, a_5, \phi} \mathcal{V}^C(\theta_5, a_5, \phi, S, E^p, \Delta)\mu_C(\theta_5, a_5, \phi, S, E^p)$$

where μ_C is the distribution of initial states $\{\theta_5, a_5, \phi, S, E^p\}$ in the economy C.

We define Δ^C as the consumption equivalence that makes individuals indifferent between being born in the baseline economy $C = 0$ and one in which policy $C \neq 0$ is in place, such that:

$$\bar{\mathcal{V}}^0(\Delta^C) = \bar{\mathcal{V}}^C(0)$$

This dissertation is the result of my own work, and no other sources or means, except the ones listed, have been employed.

Suzanne Bellue

Curriculum Vitae

2018 - 2023	University of Mannheim
	Ph.D. in Economics
2013 - 2015	Pontificia Universidad Católica de Chile
	M.A. in Economics
2012 - 2014	University of Paris Dauphine
	M.Res. in Development Economics
2009 - 2012	University of Paris Dauphine
	B.S. in Applied Economics